Informatics Control Measurement
in Economy and Environment Protection

1/2024

Informatyka Automatyka Pomiary

w gospodarce i ochronie środowiska

www.e-IAPGOS.pl

ISSN 2083-0157

Kwartalnik Naukowo-Techniczny

fot. arch. WEiI PL

60 LAT WEiI

Wydział Elektrotechniki i Informatyki
Politechnika Lubelska

1/2024

styczeń – marzec

Wydanie pod redakcją naukową
prof. dr hab. inż. Waldemara Wójcika

INFORMATYKA AUTOMATYKA POMIARY

W GOSPODARCE i OCHRONIE ŚRODOWISKA

Informatics Control Measurement in Economy and Environment Protection

p-ISSN 2083-0157, e-ISSN 2391-6761, www.e-iapgos.pl

EDITOR STAFF
ZESPÓŁ REDAKCYJNY

Editor-in-Chief
Redaktor naczelny

Paweł KOMADA
Lublin University of Technology, Lublin, Poland
p.komada@pollub.pl

Deputy Editors
Zastępcy redaktora

Jan SIKORA
Research and Development Center Netrix S.A., Lublin, Poland sik59@wp.pl

Dominik SANKOWSKI
Lodz University of Technology, Lodz, Poland
dsan@kis.p.lodz.pl

Pavel FIALA
Brno University of Technology, Brno, Czech Republic fialap@feec.vutbr.cz

Andrzej SMOLARZ
Lublin University of Technology, Lublin, Poland
a.smolarz@pollub.pl

Technical Editor
Redaktor techniczny

Tomasz ŁAWICKI
Lublin University of Technology, Lublin, Poland
t.lawicki@pollub.pl

Statistical Editor
Redaktor statystyczny

Ewa ŁAZUKA
Lublin University of Technology, Lublin, Poland
e.lazuka@pollub.pl

EDITORIAL OFFICE
REDAKCJA

Redakcja czasopisma
Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska
Katedra Elektroniki i Technik Informacyjnych
Politechnika Lubelska
ul. Nadbystrzycka 38A, 20-618 Lublin
tel. +48 81 53 84 309,
iapgos@pollub.pl
www.e-iapgos.pl
iapgos.pollub.pl
ph.pollub.pl/index.php/iapgos

PUBLISHER
WYDAWCA

Politechnika Lubelska
ul. Nadbystrzycka 38D
20-618 Lublin
tel. +48 81 53 84 100
www.pollub.pl
ph.pollub.pl

EDITORIAL BOARD
KOMITET REDAKCYJNY

Editor-in-Chief
Redaktor naczelny

Paweł KOMADA
Lublin University of Technology, Lublin, Poland
p.komada@pollub.pl

Topical Editors
Redaktorzy działowi

Electrical Engineering
Elektrotechnika

Jan SIKORA
Research and Development Center Netrix S.A., Lublin, Poland sik59@wp.pl

Computer Science
Informatyka

Dominik SANKOWSKI
Lodz University of Technology, Lodz, Poland
dsan@kis.p.lodz.pl

Electronics
Elektronika

Pavel FIALA
Brno University of Technology, Brno, Czech Republic fialap@feec.vutbr.cz

Automatic
Automatyka

Waldemar WÓJCIK
Lublin University of Technology, Lublin, Poland
waldemar.wojcik@ pollub.pl

Environmental Engineering
Inżynieria środowiska

Lucjan PAWŁOWSKI
Lublin University of Technology, Lublin, Poland
l.pawlowski@pollub.pl

Mechtronics
Mechatronika

Krzysztof KLUSZCZYŃSKI
Cracow University of Technology, Cracow, Poland krzysztof.kluszczynski@pk.edu.pl

INTERNATIONAL PROGRAMME COMMITTEE
RADA PROGRAMOWO-NAUKOWA

Chairman
Przewodniczący

Waldemar WÓJCIK
Lublin University of Technology, Lublin, Poland

Deputy of Chairman
Zastępca przewodniczącego

Jan SIKORA
Research and Development Center Netrix S.A., Lublin, Poland

Members
Członkowie

Kazimierz ADAMIAK
University of Western Ontario, Ontario, Canada

Darya ALONTSEVA
D.Serikbaev East Kazakhstan State Technical University, Ust-Kamenogorsk, Kazakhstan

Shin-ichi AOQUI
Sojo University, Kumamoto, Japan

Javier BALLESTER
Universidad de Zaragoza, Saragossa, Spain

Yurii BOBALO
Lviv Polytechnic National University, Lviv, Ukraine

Oleksy BORYSENKO
Department of Elektronics and Computer Technics, Sumy, Ukraine

Hartmut BRAUER
Technische Universität Ilmenau, Ilmenau, Germany

Kathleen CURRAN
School of Medicine & Medical Science, Dublin, Ireland

Milan DADO
University of Žilina, Žilina, Slovakia

Jarmila DEDKOVA
Brno University of Technology, Brno, Czech Republic

Andrzej DEMENKO
Poznan University of Technology, Poznań, Poland

Pavel FIALA
Brno University of Technology, Brno, Czech Republic

Vladimir FIRAGO
Belarusian State University, Minsk, Belarus

Ryszard GOLEMAN
Lublin University of Technology, Lublin, Poland

Jan GÓRSKI
AGH University of Science and Technology, Cracow, Poland

Stanisław GRATKOWSKI
West Pomeranian University of Technology Szczecin, Szczecin, Poland

Antoni GRZANKA
Warsaw University of Technology, Warsaw, Poland

Jeni HEINO
Helsinki University of Technology, Helsinki, Finland

Oleksandra HOTRA
Lublin University of Technology, Lublin, Poland

Wojciech JARZYNA
Lublin University of Technology, Lublin, Poland

Mukhtar JUNISBEKOV
M.Kh. Dulaty Taraz State University, Taraz, Kazakhstan

Piotr KACEJKO
Lublin University of Technology, Lublin, Poland

Krzysztof KLUSZCZYŃSKI
Cracow University of Technology, Cracow, Poland

Grzegorz KŁOSOWSKI
Lublin University of Technology, Lublin, Poland

Yurii KRAK
Taras Shevchenko National University of Kyiv, Kiev, Ukraine

Piotr KSIĄŻEK
Medical University of Lublin, Lublin, Poland

Piotr LESIAK
WSEI University, Lublin, Poland

Volodymyr LYTVYNENKO
Kherson National Technical University, Kherson, Ukraine

Artur MEDVIED
Riga Technical University, Riga, Latvia

Paweł MERGO
Maria Curie-Skłodowska University, Lublin, Poland

Zbigniew OMIOTEK
Lublin University of Technology, Lublin, Poland

Andrzej NAFALSKI
University of South Australia, Adelaide, Australia

Il Han PARK
Sungkyunkwan University, Suwon, Korea

Lucjan PAWŁOWSKI
Lublin University of Technology, Lublin, Poland

Sergey PAVLOV
Vinnytsia National Technical University, Vinnytsia, Ukraine

Leonid POLISHCHUK
Vinnytsia National Technical University, Vinnytsia, Ukraine

Denis PREMEL
CEA Saclay, Gif-sur-Yvette, France

Jason RILEY
The Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, USA

Ryszard ROSKOSZ
Gdańsk University of Technology, Gdańsk, Poland

Tomasz RYMARCZYK
Research and Development Center Netrix S.A., Lublin, Poland

Dominik SANKOWSKI
Lodz University of Technology, Lodz, Poland

Stanislav SLOSARCIK
Technical University of Kosice, Kosice, Slovakia

Jan SROKA
Warsaw University of Technology, Warsaw, Poland

Henryka Danuta STRYCZEWSKA
Lublin University of Technology, Lublin, Poland

Batyrbek SULEMENOV
Kazakh National Research Technical University after K.I.Satpayev, Almaty, Kazakhstan

Mirosław ŚWIERCZ
Bialystok University of Technology, Białystok, Poland

Stanisław TARASIEWICZ
Université Laval, Quebec, Canada

Murielle TORREGROSSA
University of Strasbourg, Strasbourg, France

Sławomir TUMAŃSKI
Warsaw University of Technology, Warsaw, Poland

Oleksandr VASILEVSKYI
University of Texas at Austin, Austin, USA

Andrzej WAC-WŁODARCZYK
Lublin University of Technology, Lublin, Poland

Zygmunt WARSZA
Industrial Research Institute for Automation and Measurements, Warsaw, Poland

Sotoshi YAMADA
Kanazawa University, Kanazawa, Japan

Xiaoyi YANG
Beihang University, Beijing, China

Mykola YERMOSHENKO
International Academy of Information Sciences, Kiev, Ukraine

Athanasios ZACHAROPOULOS
University College London, London, United Kingdom

Ivan ZHARSKI
Belarusian National Technical University, Minsk, Belarus

Cao ZHIHONG
Institute of Soil Science Chinese Academy of Sciences, Nanjing, China

Paweł ŻUKOWSKI
Lublin University of Technology, Lublin, Poland

PRINTING HOUSE – DRUKARNIA

PPH Remigraf Sp. z o.o.
ul. Fortuny 5, 01-339 Warszawa
https://remigraf.pl/

nakład: 100 egzemplarzy

OTHER INFORMATION – INNE INFORMACJE

Czasopismo jest indeksowane w bazach:

DOAJ	doaj.org
BazTech	baztech.icm.edu.pl
IC Journals Master List	www.journals.indexcopernicus.com
Google Scholar	scholar.google.pl
POL-index	pbn.nauka.gov.pl
Sherpa RoMEO	www.sherpa.ac.uk
OAJI	oaji.net
SCOPUS	www.scopus.com
EBSCO	www.ebsco.com

Czasopismo *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska* zostało objęte finansowaniem przez Ministerstwo Nauki i Szkolnictwa Wyższego w ramach programu *Wsparcie dla czasopism naukowych* w latach 2019-2020.

Czasopismo znajduje się w wykazie czasopism naukowych opublikowanym w Komunikacie Ministra Nauki z dnia 5 stycznia 2024 r. Unikatowy Identyfikator Czasopisma: 200167 – przypisana liczba punktów przyznawanych za publikację artykułu – 20.

Zasady publikowania artykułów, przygotowania tekstów, zasady etyczne, procedura recenzowania, wykazy recenzentów oraz pełne teksty artykułów dostępne są na stronie internetowej czasopisma:

www.e-iapgos.pl

W celu zwiększenia oddziaływania czasopisma w środowisku naukowym redakcja zaleca:

- w artykułach publikowanych w IAPGOS cytować artykuły z renomowanych czasopism międzynarodowych (szczególnie indeksowanych w bazach Web of Science oraz Scopus) używając oficjalnych skrótów nazw czasopism,
- w artykułach publikowanych w innych czasopismach (zwłaszcza indeksowanych w bazach Web of Science oraz Scopus) cytować prace publikowane w IAPGOS – zwłaszcza posługując się numerami DOI, np.:
  Kluszczyński K. *Modelowanie – umiejętność czy sztuka?* Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska – IAPGOS, 1/2016, 4–15, https://doi.org/10.5604/20830157.1193833.

CONTENTS – SPIS TREŚCI

1. **Tomasz Rymarczyk, Jan Sikora**
Some more on logarithmic singularity integration in boundary element method
Jeszcze o całkowaniu logarytmicznych osobliwości w metodzie elementów brzegowych 5

2. **Viktor Lyshuk, Vasyl Tchaban, Anatolii Tkachuk, Valentyn Zablotskyi, Yosyp Selepyna**
Electromagnetic field equations in nonlinear environment
Równania pola elektromagnetycznego w środowisku nieliniowym 11

3. **Mikhaylo Vasnetsov, Valeriy Voytsekhovich, Vladislav Ponevchinsky, Nataliia Kachalova, Alina Khodko, Oleksandr Mamuta, Volodymyr Pavlov, Vadym Khomenko, Natalia Manicheva**
Optical speckle-field visibility diminishing by reduction of a temporal coherence
Optyczne zmniejszanie widoczności pola plamkowego poprzez redukcję spójności czasowej 17

4. **Dmytro Yevgrafov, Yurii Yaremchuk**
Quality indicators of detection of side radiation signals from monitor screens by a specialized technical means of enemy intelligence
Wskaźniki jakości do wykrywania sygnałów promieniowania bocznego z ekranów monitorów przez wyspecjalizowane środki techniczne wywiadu wroga 21

5. **Samira Boumous, Zouhir Boumous, Yacine Djeghader**
The impact of lightning strike on hybrid high voltage overhead transmission line – insulated gas line
Wpływ uderzenia pioruna na hybrydową linię wysokiego napięcia napowietrzna linia przesyłowa – izolowana linia gazowa 27

6. **Kamil Płachta**
Energy efficiency of photovoltaic panels depending on the step resolution of tracking system
Efektywność energetyczna paneli fotowoltaicznych w zależności od rozdzielczości kroku śledzenia układu nadążnego 32

7. **Shanmukhaprasanthi Tammineni, Swaraiya Madhuri Rayavarapu, Sasibhushana Rao Gottapu, Raj Kumar Goswami**
Digital image restoration using SURF algorithm
Cyfrowa rekonstrukcja obrazów z wykorzystaniem algorytmu SURF 37

8. **Roman Kvyetnyy, Yuriy Bunyak, Olga Sofina, Volodymyr Kotsiubynskyi, Tetiana Piliavoz, Olena Stoliarenko, Saule Kumargazhanova**
Tensor and vector approaches to objects recognition by inverse feature filters
Podejście tensorowe i wektorowe do rozpoznawania obiektów za pomocą filtrów cech odwrotnych 41

9. **Leonid Timchenko, Natalia Kokriatska, Volodymyr Tverdomed, Iryna Yu. Yepifanova, Yurii Didenko, Dmytro Zhuk, Maksym Kozyr, Iryna Shakhina**
Architectural and structural and functional features of the organization of parallel-hierarchical memory
Architektoniczne, strukturalne i funkcjonalne cechy równoległo-hierarchicznej organizacji pamięci 46

10. **Anzhelika Stakhova, Adrián Bekö**
Simulation and computer modeling of bridge structures dynamics using ANSYS
Symulacja i modelowanie komputerowe dynamiki konstrukcji mostów z wykorzystaniem ANSYS 53

11. **Sobhana Mummaneni, Tribhuvana Sree Sappa, Venkata Gayathri Devi Katakam**
Enhancing crop health through digital twin for disease monitoring and nutrient balance
Poprawa zdrowia upraw za pomocą cyfrowego bliźniaka do monitorowania chorób i bilansu składników odżywczych 57

12. **Patryk Mauer**
Review of modelling approaches for website-related predictions
Przegląd podejść do modelowania przewidywań związanych z witrynami internetowymi 63

13. **Andrii Yarovyi, Dmytro Kudriavtsev**
Formation of highly specialized chatbots for advanced search
Tworzenie wysoce wyspecjalizowanych chatbotów do zaawansowanego wyszukiwania 67

14. **Volodymyr Akhramovych, Yuriy Pepa, Anton Zahynei, Vadym Akhramovych, Taras Dzyuba, Ihor Danylov**
Method for calculating the information security indicator in social media with consideration of the path duration between clients
Metoda obliczania wskaźnika bezpieczeństwa informacji w mediach społecznościowych z uwzględnieniem długości ścieżki między klientami 71

15. **Maksym Tymkovych, Oleg Avrunin, Karina Selivanova, Alona Kolomiiets, Taras Bednarchyk, Saule Smailova**
Correspondence matching in 3D models for 3D hand fitting
Dopasowanie zgodności w modelach 3D dla dopasowania dłoni 3D 78

16. **Govada Anuradha, Harini Davu, Muthyalanaidu Karri**
Genetic algorithm-based decision tree optimization for detection of dementia through MRI analysis
Optymalizacja drzewa decyzyjnego oparta na algorytmie genetycznym do wykrywania demencji poprzez analizę MRI 83

17. **Vitaliy Polishchuk, Sergii Pavlov, Sergii Polishchuk, Sergii Shuvalov, Andriy Dalishchuk, Natalia Sachaniuk-Kavets'ka, Kuralay Mukhsina, Abilkaiyr Nazerke**
Medical fuzzy-expert system for prediction of engraftment degree of dental implants in patients with chronic liver disease
Medyczny rozmyty system ekspercki do przewidywania stopnia wszczepienia implantów dentystycznych u pacjentów z przewlekłą chorobą wątroby 90

18. **Les Hotra, Oksana Boyko, Igor Helzhynskyy, Hryhorii Barylo, Pylyp Skoropad, Alla Ivanyshyn, Olena Basalkevych**
Root surface temperature measurement during root canal obturation
Pomiar temperatury powierzchni korzenia podczas obturacji kanałów korzeniowych 95

19. **Zakaryae Khomsi, Mohamed El Fezazi, Achraf Elouerghi, Larbi Bellarbi**
Evaluating the feasibility of thermographic images for predicting breast tumor stage using DCNN
Ocena wydajności obrazów termograficznych do przewidywania stopnia guza piersi przy użyciu DCNN 99

20. Sobhana Mummaneni, Sasi Tilak Ravi, Jashwanth Bodedla, Sree Ram Vemulapalli, Gnana Sri Kowsik Varma Jagathapurao
A comprehensive study: intracranial aneurysm detection via VGG16-Densenet hybrid deep learning on DSA images
Kompleksowe badanie: wykrywanie tętniaka wewnątrzczaszkowego za pomocą hybrydowego głębokiego uczenia się VGG16-Densenet na obrazach DSA 105

21. Mykola Kuzlo, Viktor Moshynskyi, Nataliia Zhukovska, Viktor Zhukovskyy
Deformations of soil masses under the action of human-induced factors
Deformacje masy gleby pod działaniem czynników człowieka..... 111

22. Fedir Smilianets, Oleksii Finogenov
Running a workflow without workflows: a basic algorithm for dynamically constructing and traversing an implied Directed Acyclic Graph in a non-deterministic environment
Obliczanie przepływów pracy bez przepływów pracy: podstawowy algorytm dynamicznego konstruowania i przeszukiwania niejawnego skierowanego grafu acyklicznego w środowisku niedeterministycznym..... 115

23. Dauren Darkenbayev, Arshyn Altybay, Zhaidargul Darkenbayeva, Nurbapa Mekebayev
Intelligent data analysis on an analytical platform
Inteligentna analiza danych na platformie analitycznej 119

http://doi.org/10.35784/iapgos.5864 received: 5.02.2024 | revised: 13.02.2024 | accepted: 21.02.2024 | available online: 31.03.2024

SOME MORE ON LOGARITHMIC SINGULARITY INTEGRATION IN BOUNDARY ELEMENT METHOD

Tomasz Rymarczyk[1,2], Jan Sikora[1,2]
[1]Research & Development Centre Netrix S.A., Lublin, Poland, [2]University of Economics and Innovation in Lublin, Faculty of Transport and Informatics, Lublin, Poland

Abstract. *The accuracy of calculations of integrals with logarithmic singularities for two methods, namely the method of ignoring singularities and the method of subtraction (consisting in separating the singular part from the remaining non-singular), are presented in this paper. Only two-dimensional problems, like Dirichlet's problems, as well as acoustic problems formulated in the frequency domain are considered. Problems related to the accuracy of calculations are discussed and the influence of frequency, as well as the influence of the geometry of the analysed area on the accuracy of calculations, are indicated. When we talk about the influence of geometry, we mean not only discretization, but also the configuration of the area, such as the sharp edges of the boundary line, assuming the use of the classic, without any modifications, Boundary Element Method.*

Keywords: singular integral calculating, ignoring singularity method, singularity subtraction technique, acoustic wave propagation, Boundary Element Method (BEM)

JESZCZE O CAŁKOWANIU LOGARYTMICZNYCH OSOBLIWOŚCI W METODZIE ELEMENTÓW BRZEGOWYCH

Streszczenie. *Dokładność obliczeń całek z osobliwościami logarytmicznymi dla dwóch metod a mianowicie metody ignorowania osobliwości i metody odjęcia (polegającej na wyodrębnieniu części osobliwej od pozostałej nieosobliwej), zostały przedstawione w tym artykule. Rozważono jedynie zagadnienia dwuwymiarowe zagadnienia Dirichleta jak również zagadnienia akustyczne sformułowane w dziedzinie częstotliwości. Omówiono problemy związane z dokładnością obliczeń oraz wskazano na wpływ częstotliwości a także wpływ geometrii analizowanego obszaru na dokładność obliczeń. Mówiąc o wpływie geometrii mamy na myśli nie tylko dyskretyzacje, ale także konfigurację rozpatrywanego obszaru jak na przykład ostre krawędzie linii brzegowej przy założeniu stosowania klasycznej, bez żadnych modyfikacji, Metody Elementów Brzegowych.*

Słowa kluczowe: liczenie całek osobliwych, metoda ignorowania osobliwości, metoda izolacji osobliwości, propagacja fal akustycznych, Metoda Elementów Brzegowych (MEB)

Introduction

The singular integration is a genuine problem for the Boundary Element Method, and it has a considerable influence on the final precision of calculation. The more complicated problem the more sensitive on the singularity Boundary Element Method (BEM) software is.

For the acoustic problems, the extensive review of singular integration is provided by S.M. Kirkup et al in an excellent work [3]. The authors presented several methods like:

1. ignoring the singularity,
2. subtracting out the singularity, (singularity subtraction technique)
3. product integration,
4. substitution or transformation.

In this paper only first two of those method would be considered for two dimensional problems.

1. Fundamental equations and discretization by second order boundary elements

The second order boundary elements have some well-known advantages and disadvantages as well [1, 3–5]. In the subsequent sections we would like to look closely in the problem of singular integrals calculation.

The first of the methods mentioned above seems to be the simplest one. In case of quadratic boundary elements, the numerical integration of the kernels could be done in a similar way as for constant and linear elements. The Jacobian of transformation and the components of unit outward normal vector are calculated according to the following equations:

$$J(\xi) = \frac{d\Gamma}{d\xi} = \sqrt{\left(\frac{dx(\xi)}{d\xi}\right)^2 + \left(\frac{dy(\xi)}{d\xi}\right)^2} \tag{1}$$

The components of unit outward normal are functions of local coordinate ξ:

$$n_x(\xi) = \frac{1}{J(\xi)}\left[\frac{dy(\xi)}{d\xi}\right], \qquad n_y(\xi) = -\frac{1}{J(\xi)}\left[\frac{dx(\xi)}{d\xi}\right] \tag{2}$$

It is important to notice that the choice of local coordinate system from -1 to +1 was not arbitrary, because it happens to be the same as limits used in Gaussian quadrature technique [2].

The boundary curve Γ is now divided into elements Γ_j and the numerical integration performed over each element using the local intrinsic coordinate ξ rather than the boundary segment Γ_j.

$$\int_{\Gamma_j} f(x,y)\, d\Gamma = \int_{-1}^{+1} f\big(x(\xi), y(\xi)\big) J(\xi)\, d\xi \tag{3}$$

where f means any function.

Let's carry out numerical experiment using the Dirichlet problem in the homogeneous region shown in Fig. 1.

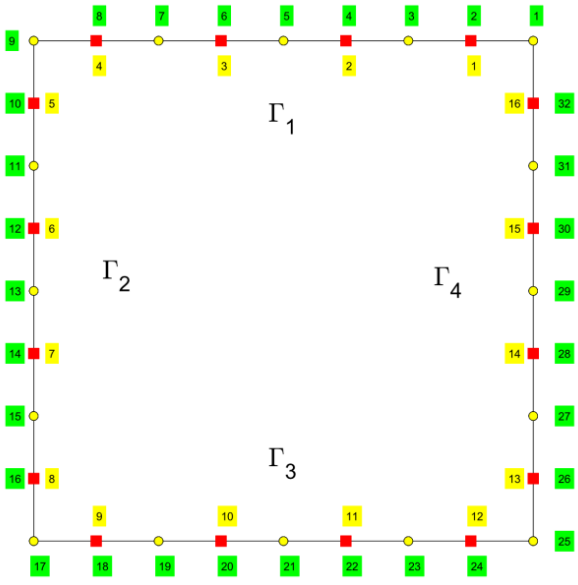


Fig. 1. Example of discretization of the homogeneous region by second order boundary elements with nodes numbering (counterclockwise direction on the green background) and elements numbering (inside subregion with yellow background)

The region consists of the square aera bounded by four lines denoted as Γ (see Fig. 1). In the upper Γ_1 and lower part Γ_3 of the boundary Dirichlet boundary conditions are imposed but on a vertical parts Γ_2 and Γ_4 Neumann homogeneous boundary conditions are defined.

Let us consider Dirichlet Problem presented in Fig. 1. The boundary of the region under consideration will be divided into M quadratic boundary elements as it is shown in Fig. 3. The nodes with the unknown values, in all three nodes of each boundary element are considered. Let consider the following integral equation [5, 7] describing distribution of electric potential inside of the region (see Fig. 1):

$$c(\boldsymbol{r})\varphi(\boldsymbol{r}) + \int_{\Gamma} \frac{\partial G(|\boldsymbol{r}-\boldsymbol{r}'|)}{\partial n}\varphi(\boldsymbol{r}')d\boldsymbol{\Gamma}(\boldsymbol{r}') = \int_{\Gamma} G(|\boldsymbol{r}-\boldsymbol{r}'|)\frac{\partial\varphi(\boldsymbol{r}')}{\partial n}\, d\boldsymbol{\Gamma}(\boldsymbol{r}') \quad (4)$$

where $c(\boldsymbol{r}) = \begin{cases} 1 & \text{for the external problems} \\ +0.5 & \text{for the smooth boundary line} \\ 1-\frac{\gamma}{2\pi} & \text{angle } \gamma \text{ is shown in the Fig. 2} \end{cases}$

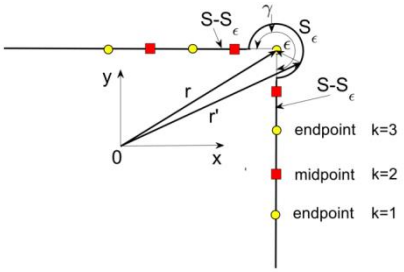


Fig. 2. Boundary point $\boldsymbol{r}$ is in the corner

The numerical integration is performed over each boundary element $\boldsymbol{\Gamma_j}$ using the local intrinsic coordinate ξ, as follows:

$$c(\boldsymbol{r})\varphi_i(\boldsymbol{r}) + \sum_{j=1}^{M}\int_{-1}^{+1}\sum_{k=1}^{3}\frac{\partial G(|\boldsymbol{r}-\boldsymbol{r}'|)}{\partial n}N_k(\xi)\left(\varphi_k(\boldsymbol{r})\right)_j J(\xi)d\xi = \sum_{j=1}^{M}\int_{-1}^{+1}\sum_{k=1}^{3} G(|\boldsymbol{r}-\boldsymbol{r}'|)\, N_k(\xi)\left(q_k(\boldsymbol{r})\right)_j J(\xi)d\xi \quad (5)$$

where M – is the total number of quadratic elements and for simplicity lets denote the normal derivative by $\left(\frac{\partial\varphi_k(\boldsymbol{r})}{\partial n}\right)_j = \left(q_k(\boldsymbol{r})\right)_j$ (see Fig. 4).

We can define a new coordinate system that is local to the element using an intrinsic variable ξ (Eq. 7–9), with its origin at the midpoint of the element and values of -1 and +1 at the end nodes as it is shown in Fig. 3.

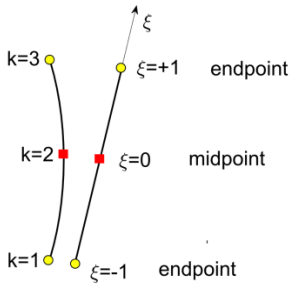


Fig. 3. Transformation from cartesian to the local coordinate system for the isoparametric quadratic element

$$\boldsymbol{r} = \left(x(\xi), y(\xi)\right) \quad (6)$$

where $\boldsymbol{r}$ is the position vector of the nodes.

$$x(\xi) = \sum_{k=1}^{3} N_k(\xi)\, x_k = N_1(\xi)x_1 + N_2(\xi)x_2 + N_3(\xi)x_3 \quad (7)$$

$$y(\xi) = \sum_{k=1}^{3} N_k(\xi)\, y_k = N_1(\xi)y_1 + N_2(\xi)y_2 + N_3(\xi)y_3 \quad (8)$$

where $N_k(\xi)$ are quadratic functions such as: $N_k(\xi) = 1$ at its own node for example the node $k = 1$ and $N_k(\xi) = 0$ at the other two nodes $k = 2$ and $k = 3$, resulting in the following:

$$\begin{aligned} N_1(\xi) &= -\frac{\xi}{2}(1-\xi) = 0.5\xi(1-\xi) \\ N_2(\xi) &= (1+\xi)(1-\xi) = 1-\xi^2 \\ N_3(\xi) &= +\frac{\xi}{2}(1+\xi) = 0.5\xi(1+\xi) \end{aligned} \quad (9)$$

Using the isoparametric elements the same basis functions are used for interpolation both variables $\varphi(\xi)$ and $q = \frac{\partial\varphi(\xi)}{\partial n}$:

$$\varphi(\xi) = \sum_{k=1}^{3} N_k(\xi)\,\varphi_k = N_1(\xi)\varphi_1 + N_2(\xi)\varphi_2 + N_3(\xi)\varphi_3 \quad (10)$$

and

$$\frac{\partial\varphi(\xi)}{\partial n} = \sum_{k=1}^{3} N_k(\xi)\frac{\partial\varphi_k}{\partial n} = = N_1(\xi)\frac{\partial\varphi_1}{\partial n} + N_2(\xi)\frac{\partial\varphi_2}{\partial n} + N_3(\xi)\frac{\partial\varphi_3}{\partial n} \quad (11)$$

The Jacobian of transformation and components of unit outward normal vector are calculated according to Eq. (1) and Eq. (2) respectively.

The second term in Eq. (2) for $n_x(\xi)$ and for $n_y(\xi)$ could be expressed in a following way:

$$\frac{\mathrm{d}x(\xi)}{\mathrm{d}\xi} = \frac{\mathrm{d}N_1(\xi)}{\mathrm{d}\xi}x_1 + \frac{\mathrm{d}N_2(\xi)}{\mathrm{d}\xi}x_2 + \frac{\mathrm{d}N_3(\xi)}{\mathrm{d}\xi}x_3 \quad (12)$$

$$\frac{\mathrm{d}y(\xi)}{\mathrm{d}\xi} = \frac{\mathrm{d}N_1(\xi)}{\mathrm{d}\xi}y_1 + \frac{\mathrm{d}N_2(\xi)}{\mathrm{d}\xi}y_2 + \frac{\mathrm{d}N_3(\xi)}{\mathrm{d}\xi}y_3 \quad (13)$$

and differentials of the shape functions are easily determined basing on Eq. (9):

$$\begin{aligned} \frac{\mathrm{d}N_1(\xi)}{\mathrm{d}\xi} &= \frac{\mathrm{d}}{\mathrm{d}\xi}\left(-\frac{\xi}{2}(1-\xi)\right) = \xi - \frac{1}{2} \\ \frac{\mathrm{d}N_2(\xi)}{\mathrm{d}\xi} &= \frac{\mathrm{d}}{\mathrm{d}\xi}\left((1+\xi)(1-\xi)\right) = -2\xi \\ \frac{\mathrm{d}N_3(\xi)}{\mathrm{d}\xi} &= \frac{\mathrm{d}}{\mathrm{d}\xi}\left(+\frac{\xi}{2}(1+\xi)\right) = \xi + \frac{1}{2} \end{aligned} \quad (14)$$

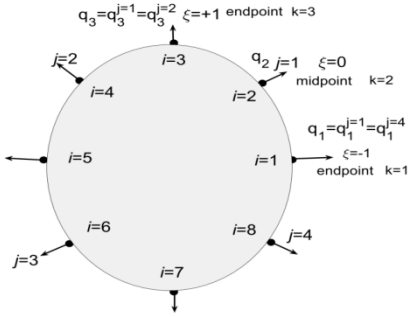


Fig. 4. Smooth boundary discreditation for the quadratic boundary elements

The numerical integration is performed over each boundary element $\boldsymbol{\Gamma_j}$ using the local coordinate ξ (see Fig. 3). The nodal values are constant, so we can finally rewrite Eq. (5) in the following form:

$$c(\boldsymbol{r})\varphi_i(\boldsymbol{r}) + \sum_{j=1}^{M}\sum_{k=1}^{3}\left(\varphi_k(\boldsymbol{r})\right)_j \int_{-1}^{+1} N_k(\xi)\frac{\partial G(|\boldsymbol{r}-\boldsymbol{r}'|)}{\partial n}d\xi = = \sum_{j=1}^{M}\sum_{k=1}^{3}\left(q_k(\boldsymbol{r})\right)_j \int_{-1}^{+1} N_k(\xi)\, G(|\boldsymbol{r}-\boldsymbol{r}'|)\, d\xi \quad (15)$$

If we denote the terms containing the integrals of the kernel's normal derivative $\frac{\partial G(|\boldsymbol{r}-\boldsymbol{r}'|)}{\partial n}$ and $G(|\boldsymbol{r}-\boldsymbol{r}'|)$ as $a_{j,k}^{(j)}$ and $b_{j,k}^{(j)}$ respectively, we will get:

$$a_{i,k}^{(j)}(\boldsymbol{r},\boldsymbol{r}') = \int_{-1}^{+1} N_k(\xi)\frac{\partial G(|\boldsymbol{r}-\boldsymbol{r}'|)}{\partial n}J_j(\xi)d\xi \quad (16)$$

$$b_{i,k}^{(j)}(\boldsymbol{r},\boldsymbol{r}') = \int_{-1}^{+1} N_k(\xi)G(|\boldsymbol{r}-\boldsymbol{r}'|)\, J_j(\xi)d\xi \quad (17)$$

For the smooth boundary, the integral functions $a_{j,k}^{(j)}$, and $b_{j,k}^{(j)}$ can be lumped together in the global functions $A_{i,j}$ and $B_{i,j}$ as follows:

$$c(\boldsymbol{r})\varphi_i(\boldsymbol{r}) + \sum_{j=1}^{M}\sum_{k=1}^{3}\left(\varphi_k(\boldsymbol{r})\right)_j\, a_{i,k}^{(j)}(\boldsymbol{r},\boldsymbol{r}') = = \sum_{j=1}^{M}\sum_{k=1}^{3}\left(q_k(\boldsymbol{r})\right)_j\, b_{j,k}^{(j)}(\boldsymbol{r},\boldsymbol{r}') \quad (18)$$

and:

$$c(\boldsymbol{r})\varphi_i(\boldsymbol{r}) + \sum_{j=1}^{M}\varphi_j(\boldsymbol{r})\, A_{i,j}(\boldsymbol{r},\boldsymbol{r}') = = \sum_{j=1}^{M} q_j(\boldsymbol{r})`\, B_{i,j}(\boldsymbol{r},\boldsymbol{r}') \quad (19)$$

where $\boldsymbol{r}$ depends on index i and $\boldsymbol{r}'$ depends on index j (see Fig. 4).

2. Numerical integration of the Kernel

To form a set of linear algebraic equations, we take each node in turn as a load point $\boldsymbol{r}$ and perform the integrations indicated in Eq. (16) and in Eq. (17) as well as in Eq. (19):

$$[\boldsymbol{A}][\boldsymbol{\varphi}] = [\boldsymbol{B}]\left[\frac{\partial \boldsymbol{\varphi}}{\partial n}\right] \tag{20}$$

where the matrices $[\boldsymbol{A}]$ and $[\boldsymbol{B}]$ contain the integrals of the kernel's normal derivative $\frac{\partial G(|\boldsymbol{r}-\boldsymbol{r}'|)}{\partial n}$ and the kernels $G(|\boldsymbol{r}-\boldsymbol{r}'|)$ respectively, i.e., the functions $A_{i,j}$ and $B_{i,j}$ of Eq. (19).

If R denotes the distance between point $\boldsymbol{r}$ and point, $\boldsymbol{r}'$ then:

$$R = |\boldsymbol{r}-\boldsymbol{r}'| = \sqrt{(x-x')^2+(y-y')^2} \tag{21}$$

The normal derivative of Green's function in case of Laplace equation is:

$$\frac{\partial G(|\boldsymbol{r}-\boldsymbol{r}'|)}{\partial n} = \frac{\partial G}{\partial R}\frac{\partial R}{\partial n} = \frac{\partial G}{\partial R}\left[\frac{\partial R}{\partial x'}\frac{\partial x'}{\partial n} + \frac{\partial R}{\partial y'}\frac{\partial y'}{\partial n}\right] = \\ = -\frac{1}{2\pi R^2}\left[(x'-x)n_{x'} + (y'-y)n_{y'}\right] \tag{22}$$

where:

$$\frac{\partial R}{\partial x'} = \frac{x'-x}{R}; \qquad \frac{\partial R}{\partial y'} = \frac{y'-y}{R} \tag{23}$$

$$\frac{\partial x'}{\partial n} = n_{x'} \qquad \frac{\partial y'}{\partial n} = n_{y'} \tag{24}$$

where $n_{x'}$ and $n_{y'}$ are defined by Eq. (1) and Eq. (2) under one condition that x must be replaced by x' and the same for y coordinate.

In case of non–singular integrals, point $\boldsymbol{r}'$ never meet point $\boldsymbol{r}$ the standard Gauss–Legendre quadrature can be easily applicable to integrals of the general form:

$$\int_{-1}^{+1} f(\xi)\, d\xi = \sum_{i=1}^{ng} w_i\, f(\xi_i), \tag{25}$$

where ng is the total number of Gaussian integration points, and ξ_i is the Gaussian coordinate with an associated weight function w_i. The most frequently used values are listed in the literature under the following address [9].

3. Ignoring the singularity

When the kernels are the regular functions, it is quite easy to integrate them according to Eq. (25). But when the point $\boldsymbol{r}$ lay in the same boundary element as point $\boldsymbol{r}'$ the singularity will become a problem. There are some special methods to deal with them which was mentioned in the introduction.

All methods mentioned in the introduction are very interesting but let us start with the first one, which seems to be the simplest one.

As a test of the proposed approach, let us consider the area of the unit square with mixed Dirichlet-Neumann boundary conditions (Fig. 1). On the edges parallel to the x-axis, on the upper part of edge the potential $\varphi = 10$ and on the lower part of edge $\varphi = -10$. On the vertical edges relative to the x-axis, homogeneous Neumann boundary conditions were imposed. Physically, this corresponds to a flat capacitor but without edge effects. The square area is trivial, but due to the selected boundary element of the second order and the non-smooth shape of the boundary, the task is not trivial at all, particularly if we use the classical boundary element method (Fig. 2). The analytical solution is known, so there is a possibility to calculate errors. The first approach, named as ignoring singularities, is conceptually the simplest one, but as it turns out, it leads to an excessive number of integration points in numerical Gauss quadrature, providing barely satisfactory accuracy of calculations (see Fig. 5).

This approach to singular integrals, unfortunately, as we see in Fig. 5, leads to a vast number of Gaussian integration points. Therefore, this method, which is undoubtedly characterized by simplicity, but at the expense of excessive numerical operations, should be considered as not interesting.

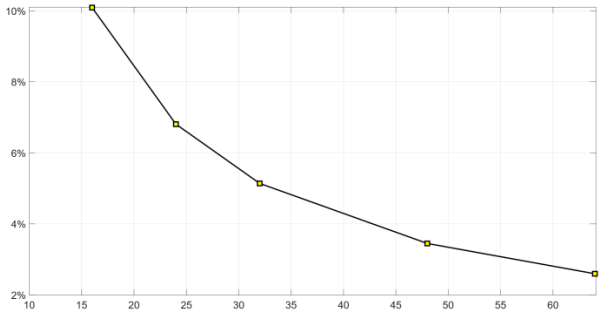


Fig. 5. Accuracy versus number of Gaussian integration points of BEM calculation of the test problem

4. Subtracting out the singularity approach

The case of weakly singular integration occurs when the collocation node lies inside the integration element and if the singularity is of the order $O(\ln(1/R))$ for two-dimensional problem. One of the methods dealing with such singularity is the subtraction method [3]. Integrand first must be separated into singular and non-singular parts. In case of quadratic boundary elements, the numerical integration of the kernels could be done in an analogous way as for constant element [5, 6]. The Jacobian of transformation and the components of unit outward normal are calculated according to Eq. (1) and Eq. (2).

The first derivative of the original coordinates for the second order element are expressed by Eq. (7) and Eq. (8) also by Eq. (12) and Eq. (13).

Dealing first with the second kernel $G(|\boldsymbol{r}-\boldsymbol{r}'|)$ (see Eq. 15) we have to consider two cases. The first case when load and field points are in different elements. Then integrals are not singular. And a bit more difficult problem when those two points are in the same element. In this case the singularity may occur in all three nodes of the quadratic boundary element. Then to calculate the integrals we must consider the following three cases:

1. the point $\boldsymbol{r}$ is the first node (k = 1) of the element,
2. the point $\boldsymbol{r}$ is the second node (k = 2) of the element,
3. the point $\boldsymbol{r}$ is the third node (k = 3) of the element.

The distance between point $\boldsymbol{r}$ and point $\boldsymbol{r}'$ was denoted by R (see Eq. 21). So, for the first node:

$$R^2 = \big(x_1 - x'(\xi)\big)^2 + \big(y_1 - y'(\xi)\big)^2 = \\ = [N_1(\xi)x_1 + N_2(\xi)x_2 + N_3(\xi)x_3 - x_1]^2 + \\ + [N_1(\xi)y_1 + N_2(\xi)y_2 + N_3(\xi)y_3 - y_1]^2 \tag{26}$$

where $N_1(\xi), N_2(\xi)$ and $N_3(\xi)$ are expressed by Eq. (9).
So:

$$R^2 = \\ = \left[-\frac{\xi}{2}(1-\xi)x_1 + (1-\xi)(1+\xi)x_2 + \frac{\xi}{2}(1+\xi)x_3 - x_1\right]^2 + \\ + \left[-\frac{\xi}{2}(1-\xi)y_1 + (1-\xi)(1+\xi)y_2 + \frac{\xi}{2}(1+\xi)y_3 - y_1\right]^2 \tag{27}$$

In an analogous way, we can write the distance R for the next nodes, the second one:

$$R^2 = \\ = \left[-\frac{\xi}{2}(1-\xi)x_1 + (1-\xi)(1+\xi)x_2 + \frac{\xi}{2}(1+\xi)x_3 - x_2\right]^2 + \\ + \left[-\frac{\xi}{2}(1-\xi)y_1 + (1-\xi)(1+\xi)y_2 + \frac{\xi}{2}(1+\xi)y_3 - y_2\right]^2 \tag{28}$$

and the third one:

$$R^2 = \\ = \left[-\frac{\xi}{2}(1-\xi)x_1 + (1-\xi)(1+\xi)x_2 + \frac{\xi}{2}(1+\xi)x_3 - x_3\right]^2 + \\ + \left[-\frac{\xi}{2}(1-\xi)y_1 + (1-\xi)(1+\xi)y_2 + \frac{\xi}{2}(1+\xi)y_3 - y_3\right]^2 \tag{29}$$

Equations (27-29) shows how the distant R could be calculated. However, when the points $\boldsymbol{r}$ and $\boldsymbol{r}'$ are in the same element but $\boldsymbol{r} \neq \boldsymbol{r}'$ the kernels are singular but the shape function $N_k(\xi)$ in the vicinity of $\boldsymbol{r}$ is of the order r. Therefore, the product of the kernels and the shape function is not singular, and the integrals can be evaluated using the standard Gaussian quadrature (see Eq. (25)). So far, all the off–diagonal coefficients of the matrices $[\boldsymbol{A}]$ and $[\boldsymbol{B}]$ (see Eq. (20)) have been calculated.

In case, when the points $\boldsymbol{r}$ and $\boldsymbol{r}'$ are in the same element but $\boldsymbol{r} \to \boldsymbol{r}'$, so distance $R \to 0$ the standard Gaussian quadrature cannot be used, because of the singularity of the kernels. Dealing with the kernel $G_0(\boldsymbol{r}, \boldsymbol{r}') = \frac{1}{2\pi}\ln\frac{1}{R}$, it is clear that as $\boldsymbol{r}$ coincides with $\boldsymbol{r}'$, the singularity is of the form $\ln(1/\eta)$ as $\eta \to 0$. Fortunately, this form of integral can be calculated by using the special logarithmic Gaussian quadrature scheme given bellow:

$$\int_0^1 f(\eta)\ln\frac{1}{\eta}d\eta = \sum_{i=1}^{gl} w_i\ f(\eta_i) \quad (30)$$

where gl is the total number of logarithmic Gaussian integration points used and η_i is the Gaussian coordinate with an associated weight function w_i [5]. Note that the limits of integration are now from 0 to 1 instead of the −1 to +1 range used in the non–singular integrals.

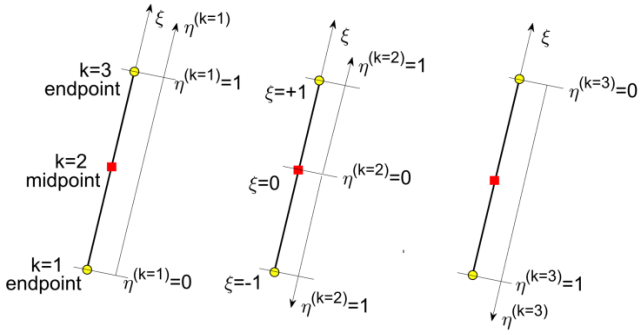


Fig. 6. Coordinates transformation. a) $\boldsymbol{r}$ is in the first node of the element, b) second node and c) third node

A simple linear transformation can be used to transform the integral variable from ξ to η (see Fig. 6) as follows:

1. if the position vector $\boldsymbol{r}$ is the first node (k = 1) of the element: $\eta^{(k=1)} = 0.5(1+\xi)$,

2. if the position vector $\boldsymbol{r}$ is the second node (k = 2) of the element – the element is divided into two sub – elements (see Fig. 6b):

for $-1 < \xi < 0$ $\eta'^{(k=2)} = -\xi$ and for $0 < \xi < +1$ $\eta^{(k=2)} = \xi$,

3. if the position vector $\boldsymbol{r}$ is the third node (k = 3) of the element: $\eta^{(k=3)} = 0.5(1-\xi)$.

For the first node the Eq. (27) can be rearranged in the following way:

$$\begin{aligned} R^2 &= [0.5(1+\xi)]^2 \cdot \\ &\cdot\{[(\xi-2)x_1 + 2(1-\xi)x_2 + \xi x_3]^2 + \\ &+[(\xi-2)y_1 + 2(1-\xi)y_2 + \xi y_3]^2\} = \\ &= \left(\eta^{(1)}\right)^2\left[\left(f_x^{(1)}(\xi)\right)^2 + \left(f_y^{(1)}(\xi)\right)^2\right] \end{aligned} \quad (31)$$

For the second node the Eq. (28) can be rearranged in the following way:

$$\begin{aligned} R^2 &= \\ &= \xi^2\{[0.5(\xi-1)x_1 - \xi x_2 + 0.5(\xi+1)x_3]^2 + \\ &+[0.5(\xi-1)y_1 - \xi y_2 + 0.5(\xi+1)y_3]^2\} = \\ &= \left(\eta^{(2)}\right)^2\left[\left(f_x^{(2)}(\xi)\right)^2 + \left(f_y^{(2)}(\xi)\right)^2\right] \end{aligned} \quad (32)$$

For the third node the Eq. (29) can be rearranged in the following way:

$$\begin{aligned} R^2 &= \\ &= [0.5(1-\xi)]^2\{[-\xi x_1 + 2(1+\xi)x_2 - 2(2+\xi)x_3]^2 + \\ &+[-\xi y_1 + 2(1+\xi)y_2 2(2+\xi)y_3]^2\} = \\ &= \left(\eta^{(3)}\right)^2\left[\left(f_x^{(3)}(\xi)\right)^2 + \left(f_y^{(3)}(\xi)\right)^2\right] \end{aligned} \quad (33)$$

Therefore, a general expression can be written for the logarithmic term as follows:

$$\begin{aligned} G_0(|\boldsymbol{r}-\boldsymbol{r}'|) &= \frac{1}{2\pi}\ln\frac{1}{|\boldsymbol{r}-\boldsymbol{r}'|} = \frac{1}{2\pi}\ln\frac{1}{(\eta^{(k)})^2\left[\left(f_x^{(k)}(\xi)\right)^2+\left(f_y^{(k)}(\xi)\right)^2\right]} = \\ &= \frac{1}{2\pi}\ln\frac{1}{(\eta^{(k)})^2} - \frac{1}{2\pi}\frac{1}{2}\ln\left[\left(f_x^{(k)}(\xi)\right)^2 + \left(f_y^{(k)}(\xi)\right)^2\right] \end{aligned} \quad (34)$$

where: k indicates, the node number and $\eta^{(k)}$ transforms the integration limits from (−1 to +1) to (0 to +1) and changes integrand value according to the position of $\boldsymbol{r}$ in the element.

So, for quadratic boundary element the kernel as $\boldsymbol{r}'$ approaches $\boldsymbol{r}$, can be split into two distinct parts: logarithmic part (with singularity) and a non–logarithmic (regular) one (see first and the second term of Eq. (34)). Dealing now with the logarithmic part of the kernel, it can be shown that it contains terms of order $\frac{1}{\eta}$ as $\eta \to 0$.

Therefore, we can no longer use the Gaussian quadrature technique, even if a large number of Gaussian points are used. Furthermore, we also need to explicitly calculate the parameter $c(\boldsymbol{r})$ (see Eq. 4) because its contribution is added to the diagonal terms of the $[\boldsymbol{A}]$ matrix. This problem is particularly important in cases where the boundary line is not smooth as it is shown for the example in Fig. 1 and Fig. 2. However, because all non–diagonal coefficients of the $[\boldsymbol{A}]$ matrix can be calculated, there is a way to overcome this problem (see Eq. 4) [1].

5. Subtracting out the singularity in case of the constant element discretization for acoustic problems

Above the second approach subtracting out singularity was presented for the second order boundary element in case of Laplace's equation. The singular integral is divided into logarithmic (singular) and regular (non-singular) parts. This approach, as shown above, seems to be laborious and can only be directly applied to the Laplace equation.

So, arise a question how we can treat more complicated problems. We will show subtracting out singularity procedure for acoustic problem, however for constant boundary element for simplicity.

This method is widely applied in BEM code and provide high precision solution [1, 3, 8]. In this case integrand would be split onto two parts the singular part and the regular one in the following way. For example, if $f(x) \sim \psi(x)$ near to the singularity then writing $f(x) = \psi(x) + [f(x) - \psi(x)]$ the first term $\psi(x)$ might be singular but the last one $f(x) - \psi(x)$ may be regular one.

The singularity does not disappear but $\int \psi(x)dx$ in some cases (for example the Laplace's equation) it could be calculated analytically. This is a general approach and could be used not necessarily in acoustic but in many other problems.

$$\begin{aligned} \int_\Gamma G(|\boldsymbol{r}-\boldsymbol{r}'|)\,d\boldsymbol{\Gamma} &= \int_\Gamma [G(|\boldsymbol{r}-\boldsymbol{r}'|) - G_0(|\boldsymbol{r}-\boldsymbol{r}'|)]d\boldsymbol{\Gamma} + \\ &+ \int_\Gamma G_0(|\boldsymbol{r}-\boldsymbol{r}'|)d\boldsymbol{\Gamma} \end{aligned} \quad (35)$$

For 2D system, the fundamental solution for the Laplace's equation is equal:

$$\begin{aligned} \int_\Gamma G_0(|\boldsymbol{r}-\boldsymbol{r}'|)d\boldsymbol{\Gamma} &= \frac{1}{2\pi}\ln\frac{1}{|\boldsymbol{r}-\boldsymbol{r}'|} = \\ &= -\frac{1}{2\pi}\ln\sqrt{(x-x')^2+(y-y')^2} = \\ &= -\frac{1}{2\pi}\frac{1}{2}\ln((x-x')^2+(y-y')^2) \end{aligned} \quad (36)$$

The distance R (see Eq. (21)) between point $\boldsymbol{r}$ and point $\boldsymbol{r}'(\xi)$ depends on local coordinate system as Eq. (21) shows. Consequently, the main diagonal term of the $[\boldsymbol{B}]$ matrix (Eq. (20)) could be expressed in the following way:

$$b_{i,k}^{(j)}(\boldsymbol{r}, \boldsymbol{r}') = \int_{-1}^{+1} N_k(\xi) G_0(|\boldsymbol{r}-\boldsymbol{r}'|)\, J_j(\xi) d\xi \quad (37)$$

To solve more advanced acoustic problems, let us start with the excellent benchmark problem suggested by S. Kirkup in [3].

Let us consider the air as an acoustic environment at 20℃ and the pressure is one atmosphere (one atm is equal to 101.325 kPa). The speed of sound is equal to 344 m/s and the frequency of the test is within the range 300 up to 1200 Hz, hence $k = 5.48 \div 21.92\ m^{-1}$ and the wavelength $\lambda = \frac{c}{f} = 1.15 \div 0.29$ m.

The region of the benchmark is like this one presented in Fig. 1, but this time discretized by constant boundary element for simplicity, just to show how the subtracting method works even for more complicated, from physical point of view, examples.

$$\int_{\Gamma} G(|\boldsymbol{r}-\boldsymbol{r}'|)\,d\boldsymbol{\Gamma} = \int_{\Gamma} G_0(|\boldsymbol{r}-\boldsymbol{r}'|)d\boldsymbol{\Gamma} + \\ + \int_{\Gamma} [G(|\boldsymbol{r}-\boldsymbol{r}'|) - G_0(|\boldsymbol{r}-\boldsymbol{r}'|)]d\boldsymbol{\Gamma} \tag{38}$$

For 2D problem exists the analytical expression for the integral $\int_{\Gamma} G_0(|\boldsymbol{r}-\boldsymbol{r}'|)d\boldsymbol{\Gamma}$, then for acoustic problem described by the Helmholtz equation (38) we will get (consult for example [3, 8]):

$$G(|\boldsymbol{r}-\boldsymbol{r}'|) - G_0(|\boldsymbol{r}-\boldsymbol{r}'|) = \frac{\mathrm{i}}{4}H_0^{(1)}(kR) - \frac{-1}{2\pi}\ln(R) = \\ = \frac{\mathrm{i}}{4}\left(J_0(kR) + iY_0(kR)\right) + \frac{1}{2\pi}\ln(R) = \\ = -\frac{1}{4}Y_0(kR) + \frac{1}{2\pi}\ln(R) + \frac{\mathrm{i}}{4}J_0(kR) \tag{39}$$

where: J_0 and Y_0 are the Spherical Bessel functions of the first and the second kind respectively and R is defined by Eq. (21).

In acoustic problems described in a frequency domain integral formulation is:

$$c(\boldsymbol{r})\varphi(\boldsymbol{r}) + \int_{\Gamma} \frac{\partial G(|\boldsymbol{r}-\boldsymbol{r}'|)}{\partial n}\varphi(\boldsymbol{r})d\boldsymbol{\Gamma} = \\ = \int_{\Gamma} G(|\boldsymbol{r}-\boldsymbol{r}'|)\frac{\partial\varphi(\boldsymbol{r})}{\partial n}d\boldsymbol{\Gamma} \tag{40}$$

where $c(\boldsymbol{r})$ is defined by Eq. (4).

Now the boundary integral equation Eq. (40) for constant boundary elements can be written in terms of local coordinate ξ instead of the boundary line $\boldsymbol{\Gamma}$, as follows:

$$c(\boldsymbol{r})\varphi(\boldsymbol{r}) + \sum_{j=1}^{M}\varphi_j(\boldsymbol{r})\int_{-1}^{+1}\frac{\partial G(|\boldsymbol{r}-\boldsymbol{r}'|)}{\partial n}J(\xi)d\xi = \\ = \sum_{j=1}^{M}\frac{\partial\varphi_j(\boldsymbol{r})}{\partial n}\int_{-1}^{+1}G(|\boldsymbol{r}-\boldsymbol{r}'|)J(\xi)d\xi \tag{41}$$

where M – is the total number of constant elements, and $J(\xi)$ – is the Jacobian of transformation (see Eq. (1)).

The functions under integral sign which contain the kernels can be substituted by the functions $A_{i,j}$ and $B_{i,j}$ as follows:

$$c(\boldsymbol{r})\varphi(\boldsymbol{r}) + \sum_{j=1}^{M}\varphi_j(\boldsymbol{r})\,A_{i,j}(\boldsymbol{r},\boldsymbol{r}') = \\ = \sum_{j=1}^{M}\frac{\partial\varphi_j(\boldsymbol{r})}{\partial n}B_{i,j}(\boldsymbol{r},\boldsymbol{r}') \tag{42}$$

To form a set of linear algebraic equations, we take each node in turn as a load point $\boldsymbol{r}$ and perform the integrations indicated in Eq. (42). This will result in the following matrices:

$$[\boldsymbol{A}][\boldsymbol{\varphi}] = [\boldsymbol{B}]\left[\frac{\partial\boldsymbol{\varphi}}{\partial n}\right], \tag{43}$$

where the matrices $[\boldsymbol{A}]$ and $[\boldsymbol{B}]$ contain the integrals of the kernel's normal derivative $\frac{\partial G(|\boldsymbol{r}-\boldsymbol{r}'|)}{\partial n}$ and the kernels $G(|\boldsymbol{r}-\boldsymbol{r}'|)$ respectively, i.e., the functions $A_{i,j}$ and $B_{i,j}$ of Eq. (42).

Therefore, the kernel derivative with respect the normal direction in the collocation point can be expressed:

$$\frac{\partial G(|\boldsymbol{r}-\boldsymbol{r}'|)}{\partial n} = -\frac{\mathrm{i}}{4}kH_1^{(1)}(k|\boldsymbol{r}-\boldsymbol{r}'|)\left(\frac{x-x'}{r}n_x + \frac{y-y'}{r}n_y\right) \tag{44}$$

where $H_1^{(1)}$ is the spherical Hankel function of the first kind and of order one.

The integrand after the subtracting out procedure is regular with the separated singular term which could be calculated analytically (see for example [1, 3]).

Inside the domain which is the interior of a square (like the one presented in Fig. 1) but of side equal to 1 m and origin of cartesian system is in the left lower corner. Then distribution of the velocity potential φ as analytical solution is equal to:

$$\varphi(\mathbf{p}) = \sin\left(\frac{k}{\sqrt{2}}x\right)\sin\left(\frac{k}{\sqrt{2}}y\right) \tag{45}$$

This equation (45) is the solution of the Helmholtz equation (see Eq. (40)) with the following boundary condition. On the part of the boundary the Dirichlet boundary conditions are imposed:

$$\varphi(\mathbf{p}) = \begin{cases} 0 \text{ on } \boldsymbol{\Gamma_2} & \text{when } x = 0 \\ 0 \text{ on } \boldsymbol{\Gamma_3} & \text{when } y = 0 \end{cases} \tag{46}$$

and on the other part:

$$\varphi(\mathbf{p}) = \begin{cases} \sin\left(\frac{k}{\sqrt{2}}a\right)\sin\left(\frac{k}{\sqrt{2}}y\right) \text{ on } \boldsymbol{\Gamma_4} & \text{when } x = 1 \\ \sin\left(\frac{k}{\sqrt{2}}x\right)\sin\left(\frac{k}{\sqrt{2}}a\right) \text{ on } \boldsymbol{\Gamma_1} & \text{when } y = 1 \end{cases} \tag{47}$$

where $a = 1$m is the size of the square region.

For above Dirichlet boundary conditions solution and the equipotential lines are presented in the below figures:

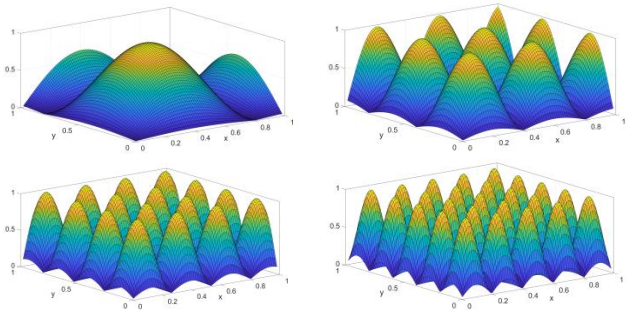


Fig. 7. Potential surface over the region of interest for different frequency of excitation

Drawing of equipotential lines corresponding to the velocity potential surface (above) for different frequencies with marked points for which the mean relative error was calculated (see table 1 and 2) and the equipotential lines are presented below (Fig. 8). The choice of these checkpoints is a compromise between points that define equipotential lines (too many) and a small number of points that could reliably reflect the mean relative error of the calculations. Their location is shown in the Fig. 8.

As the analytical solution is known (see Eq. 45) then the measure of error could be calculated only in special points shown in Fig. 8.

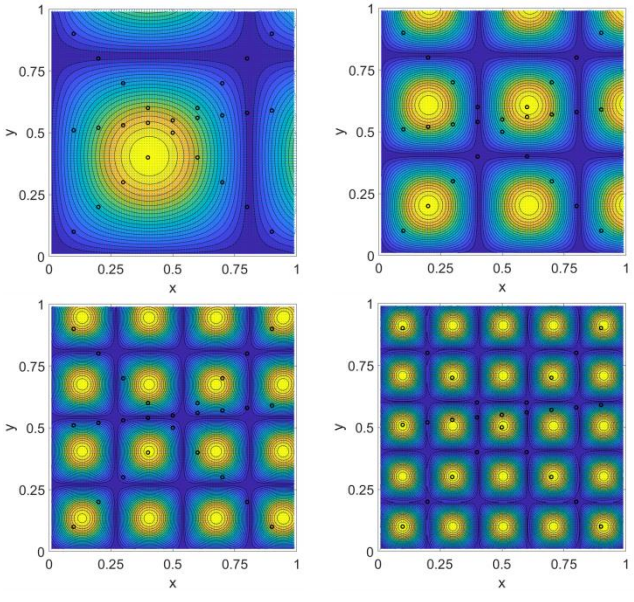


Fig. 8. Region of interest and equipotential lines distribution for different frequency excitation

The points are selected in such a way as to provide a comparison of the analytical solution with the numerical one for the entire range from minimum to maximum absolute values of the velocity potential. Therefore, the above method can be taken as an estimation of the accuracy of the calculations.

The measure of error used to estimate the accuracy of the subtracting out method is the mean relative error (*MRE*):

$$MRE = \frac{1}{N}\sum_{j=1}^{N}\frac{|\widehat{\varphi_j}-\varphi_j|}{|\varphi_j|} \tag{48}$$

where $\widehat{\varphi_j}$ and φ_j are the numerically computed and analytical values of φ at the specially selected points presented in Fig. 8.

The basic acoustic data for benchmark calculation are presented in table 1. In Fig. 9 there is a graphical representation of the mean relative error versus frequency when the discretization of the area remains constant. It is well known that the number of boundary element per wavelength must be between 10 to 8 or 6. The lower limit strongly depend on the problem.

Table 1. Basic data for acoustic benchmark calculation

f [Hz]	300	600	900	1200
MRE [%]	0.66	5.35	6.97	32.55
$\boldsymbol{k}$ [1/m]	5.47	10.96	16.44	21.92
$\boldsymbol{\lambda}$ [m]	1.15	0.57	0.38	0.29
element length [m]	0.0833	0.0833	0.0833	0.0833
max arg. of Henkel function	0.1966	0.3932	0.5898	1.977
no of BE per $\boldsymbol{\lambda}$	13.8	6.8	4.6	3.5

As we can see the subtract out procedure of singular integrals calculations provide reliable results. If all standards of acoustic calculations are preserved than the mean relative error is extremely low (see the first column of the table 1).

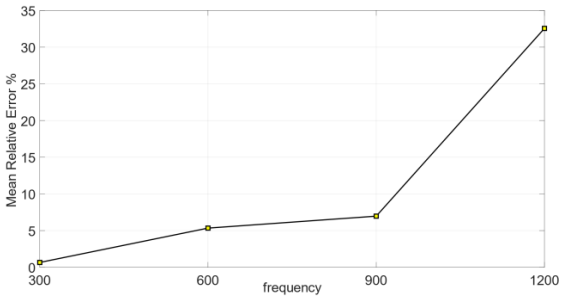


Fig. 9. Mean Relative Error versus frequency for constant discretization

To check the influence of subtracting out procedure on the calculations, the number of the elements per wavelength was kept constant on the level of a little bit more seven. Then the mean relative error for different frequency oscillated between 1 and 4.7% (see the table 2). It is worth to notice that for the highest frequency equal to 1200 Hz the error is only 1.15%.

Table 2. Basic data for acoustic benchmark calculation

f [Hz]	300	600	900	1200
MRE [%]	2.00	4.72	0.97	1.15
$\boldsymbol{k}$ [1/m]	5.48	10.96	16.44	21.92
$\boldsymbol{\lambda}$ [m]	1.15	0.57	0.38	0.29
element length [m]	0.1667	0.0833	0.0556	0.0417
max arg. of Henkel function	0.45	0.45	0.393	0.393
no of BE per $\boldsymbol{\lambda}$	6.88	6.88	6.88	6.88

The results are quite satisfactory, and the maximal error is less than 5%.

6. Conclusions

In this paper the methods of logarithmic singularity in 2D space were considered.

Of the many different methods, only two have been considered, namely the method of ignoring singularities and the method of subtracting out singularity. The first method is presented on the example of a boundary element of the second order. As a result of the numerical experiment, this method was assessed as ineffective.

The second method is similarly presented on the example of a three-node boundary element of the second order for the Laplace equation. This case is a bit more complicated, because in one element in each of the three nodes we have singularities. It shows how a singular part can be extracted from an integrand function. The procedure is tedious and such a separation is not always possible due to the arithmetic intricacies.

Therefore, another way was indicated, illustrating it with an acoustic example. The kernels for acoustic problem are so complicated that, to simplify this problem, it was decided to use a constant element where the singularity occurs only in one node. The numerical experiment confirmed the efficiency and accuracy of this method.

References

[1] Aliabadi M. H.: The Boundary Element Method, Volume 2, Applications in Solids and Structures. John Wiley & Sons, LTD, 2002.
[2] Harwood A. R. G.: Numerical Evaluation of Acoustic Green's Functions. PhD School of Mechanical, Aerospace & Civil Engineering, University of Manchester, 2014.
[3] Kirkup S.: The Boundary Element Method in Acoustics: A Survey. Appl. Sci. 9, 2019, 1642 [https://doi.org/10.3390/app9081642].
[4] Rymarczyk T.: Tomographic Imaging in Environmental, Industrial and Medical Applications. Innovatio Press Publishing Hause 2019.
[5] Sikora J.: Boundary Element Method for Impedance and Optical Tomography. Warsaw University of Technology Publishing Hause, Warsaw 2007.
[6] Smith R. N. L.: Direct Gauss quadrature formulae for logarithmic singularities on isoparametric elements. Engineering Analysis with Boundary Elements 24, 2000, 161–167.
[7] [https://www.comsol.com/acoustics-module] (available: 10.01.2024).
[8] [https://reference.wolfram.com/language/PDEModels/tutorial/Acoustics/AcoustiscsFrequencyDomain.html] (available: 10.01.2024).
[9] Gaussian Quadrature Weights and Abscissae [po-max.github.io] (available: 10.01.2024).

Prof. Tomasz Rymarczyk
e-mail: tomasz@rymarczyk.com

He is the director in Research and Development Centre in Netrix S.A. and the director of the Institute of Computer Science and Innovative Technologies in the University of Economics and Innovation, Lublin, Poland. He worked in many companies and institutes developing innovative projects and managing teams of employees. His research area focuses on the application of non-invasive imaging techniques, electrical tomography, image reconstruction, numerical modelling, image processing and analysis, process tomography, software engineering, knowledge engineering, artificial intelligence, and computer measurement systems.

https://orcid.org/0000-0002-3524-9151

Prof. Jan Sikora
e-mail: sik59@wp.pl

Prof. Jan Sikora (Ph.D. D.Sc. Eng.) graduated from Warsaw University of Technology Faculty of Electrical Engineering. During 50 years of professional work, he has obtained all grades, including the position of full professor at his alma mater. Since 1998 he has also worked for the Institute of Electrical Engineering in Warsaw. In 2008, he has joined Electrical Engineering and Computer Science Faculty in Lublin University of Technology. During 2001–2004 he has worked as a Senior Research Fellow at University College London in the prof. S. Arridge's Group of Optical Tomography.

https://orcid.org/0000-0002-9492-5818

http://doi.org/10.35784/iapgos.5533 *received: 31.10.2023 | revised: 08.03.2024 | accepted: 18.03.2024 | available online: 31.03.2024*

ELECTROMAGNETIC FIELD EQUATIONS IN NONLINEAR ENVIRONMENT

Viktor Lyshuk[1], Vasyl Tchaban[2], Anatolii Tkachuk[1], Valentyn Zablotskyi[1], Yosyp Selepyna[1]
[1]Lutsk National Technical University, Faculty of Computer and Information Technologies, Department of Electronics and Telecommunications, Lutsk, Ukraine,
[2]Lviv Polytechnic National University, Institute of Energy and Control Systems, Department of Theoretical and General Electrical Engineering, Lviv, Ukraine.

Abstract. *The paper proposes electromagnetic field equations from the point of view of their adaptation to numerical methods. Maxwell's equations with partial derivatives are used, written concerning field vectors, which most fully reproduce the picture of physical processes in electrical engineering devices. The values of these vectors provide comprehensive information about the field at any spatio-temporal point. The concept of creating mathematical models of electrical devices adequate to physical processes has been developed. Mathematical transformations are carried out according to the rules of differential calculus. Dynamic processes in the elements of electrotechnical devices were analyzed using the apparatus of mathematical modeling. An algorithm for implementing differential equations with partial derivative numerical methods using computer simulation was implemented. The obtained results made it possible to understand the nature of electromagnetic phenomena in nonlinear media. The paper provides calculations of the field parameters in a flat ferromagnetic plate and the groove of the rotor of an electric machine.*

Keywords: electromagnetic field differential equations, mathematical model, numerical integration

RÓWNANIA POLA ELEKTROMAGNETYCZNEGO W ŚRODOWISKU NIELINIOWYM

Streszczenie. *W artykule zaproponowano równania pola elektromagnetycznego pod kątem ich adaptacji do metod numerycznych. Wykorzystano równania Maxwella z pochodnymi cząstkowymi, zapisywane względem wektorów pola, które najpełniej odtwarzają obraz procesów fizycznych w urządzeniach elektrycznych. Wartości tych wektorów dostarczają wyczerpujących informacji o polu w dowolnym punkcie czasoprzestrzennym. Opracowano koncepcję tworzenia modeli matematycznych urządzeń elektrycznych adekwatnych do procesów fizycznych. Przekształcenia matematyczne przeprowadzono zgodnie z zasadami rachunku różniczkowego. Do analizy procesów dynamicznych zachodzących w elementach urządzeń elektrotechnicznych wykorzystano modelowanie matematyczne. Zaimplementowano algorytm realizacji równań różniczkowych metodami numerycznymi pochodnych cząstkowych z wykorzystaniem symulacji komputerowej. Uzyskane wyniki pozwoliły zrozumieć naturę zjawisk elektromagnetycznych w ośrodkach nieliniowych. W artykule przeprowadzono obliczenia parametrów pola w płaskiej płycie ferromagnetycznej oraz w rowku wirnika maszyny elektrycznej.*

Słowa kluczowe: równania różniczkowe pola elektromagnetycznego, model matematyczny, całkowanie numeryczne

Introduction

Today, scientists are paying more and more attention to the methods of nonlinear issues in engineering calculations. The need to calculate spatial electromagnetic fields arises when solving a wide range of problems in electrical engineering, electronics, and telecommunications [1, 2]. These include magnetohydrodynamic energy generators, astrophysical objects, electromagnetic pulse propagation, plasma accelerators, mobile communications, etc.

Knowledge of research methods and description of electromagnetic fields and waves will make it possible to build field mathematical models of electrical and electronic devices. The simulation itself is a powerful means of researching the above devices. Thus, the construction of a mathematical model based on the equations of the electromagnetic field of a real physical object is realized by a skillful combination of the laws of electrical engineering with differential equations [5, 8].

Depending on the conditions of the problem and the use of the mathematical apparatus (ordinary differential equations or equations with partial derivatives), mathematical models can be considered as a Cauchy problem or a mixed problem.

1. Literature review

The analysis of recent studies shows that it is possible to unify equations and models by focusing them on the use of powerful numerical methods, in particular, explicit integration. This is realized by abandoning the traditional methods of the theory of electric circuits in favor of electromagnetic circuit methods and electromagnetic field methods in their close combination [1, 4].

The effectiveness of mathematical modeling in the study of transient processes in electrical devices directly depends on methodological (methods of computational mathematics) and technical factors in the presence of high-performance computers [5, 9].

The paper proposes a theoretically grounded mathematical apparatus, focused on the construction of optimal computing algorithms. The real designs of electrotechnical devices are quite complex in terms of geometry, which is related to the optimal use of conductor, structural, and insulating materials. When building a computational algorithm, the differential equations in the mathematical model must be approximated by difference schemes based on algebraic equations [13].

An essential feature of our class of problems from a computational point of view is the sharp heterogeneity of electrophysical properties, which gives rise to nonlinearity. Thus, the electrical conductivity of the construction materials of electrical engineering devices can change along spatial coordinates with a jump from zero to finite values, and in some cases to infinity. The equations of electrodynamics form the basis of mathematical models for this class of problems [12].

In this work, a mathematical apparatus is used for modeling [3], which is based on the theory of nonlinear differential equations, the solution of which is possible with the correct application of numerical methods oriented to computer technology. As is known, electrotechnical materials are characterized by isotropic and anisotropic properties. Let's first consider the principles of forming equations for an anisotropic medium, that is, one whose physical properties depend on the direction. In a nonlinear anisotropic medium, the values of electrical conductivities, dielectric, and magnetic permillivity are functions of the electromagnetic state and are described by diagonal matrices.

At the beginning of its development, the research and analysis methods of electrical devices were developed as methods in the timeless domain. Electrical devices were treated as ideal devices with linearized electromagnetic couplings, which resulted in incorrect results. That is, the mathematical apparatus inadequately described the physical processes in these devices.

The maximum use of steel in the magnetic conductors of electrical equipment led to the fact that the electromagnetic connections in the nominal states differed significantly from the linear ones. At the same time, calculation methods were based on approximate consideration of nonlinearities, which did not meet the requirements of practice. The application of nonlinear

differential equations significantly complicated their integration. However, only such an approach for the analysis of processes in electrotechnical devices would ensure high accuracy of the calculation.

During the analysis of electrical devices by electromagnetic field methods, it is necessary to integrate differential equations with ordinary and partial derivatives in a single time-space. For this, equations with partial derivatives must be discretized using finite difference or finite element methods [14].

As you know, the consideration of the skin effect in the grooves of electric machines is based on the theory of electric circuits. Here, a multi-loop substitute scheme was used, determining whose parameters are complicated due to the increase in the number of links. This increased the order of the differential equations and their stiffness, which led to a loss of accuracy of the results. Summarizing the review of the literature available to us, we can state that the most promising direction of analysis and research of electromagnetic processes in electrotechnical devices is the application of electromagnetic field theory methods based on nonlinear differential equations. These are time-domain methods. Only in this way is it possible to carry out an adequate analysis with the help of mathematical models that describe physical processes as accurately as possible.

2. Researches methodology

Maxwell's equations for such a case take the form [3]. The function rot is the rotor of a three-dimensional vector field the coordinates of which are determined by the determinant of the third order. The first row is the coordinates of the coordinate axes *x*, *y*, *z*, the second corresponds to partial differentiation operators, and the third corresponds to the coordinate of the vector field [11, 12].

$$\begin{aligned} rot\vec{H} &= \vec{\delta} \\ rot\vec{E} &= -\frac{\partial \vec{B}}{\partial t} \end{aligned} \tag{1}$$

$$\vec{\delta} = G\vec{E} + \varepsilon'\frac{\partial \vec{E}}{\partial t} \tag{2}$$

where $\vec{H}$ is the magnetic field intensity vector, A/m; $\vec{\delta}$ – vector of current density, G – matrix of electric conductivities the diagonal elements of which are determined from equation (3); $\vec{D}$ – electric induction vector; $\vec{E}$ – vector of electric field intensity, ε' – matrix of differential electric permillivity electric induction vector $D = f(E)$.

$$G = \begin{bmatrix} \gamma_x & & \\ & \gamma_y & \\ & & \gamma_z \end{bmatrix} \tag{3}$$

The elements of the matrix G are determined by the characteristics of the conductor in the direction of the anisotropy *x*, *y*, *z* main axes by formula (4):

$$\gamma_i(E) = \delta_i / E, \quad i = x, y, z \tag{4}$$

We find the electric field strength from the concepts of trigonometry.

The advantage of writing equation (1) in vector form is that it does not depend on the choice of spatial coordinate system. However, the expressions of the rotor components differ in different coordinate systems.

According to figure 1, the sum of products of magnetic voltages *Hdl* on all sides of the contour *abcda* is: $+H_y dy$ along side *ab*, $\left(H_z + \frac{\partial H_z}{\partial y}dy\right)dz$ along *bc*, $-\left(H_y + \frac{\partial H_y}{\partial z}dz\right)dy$ along *cd*, $-H_z dz$ along *da*.

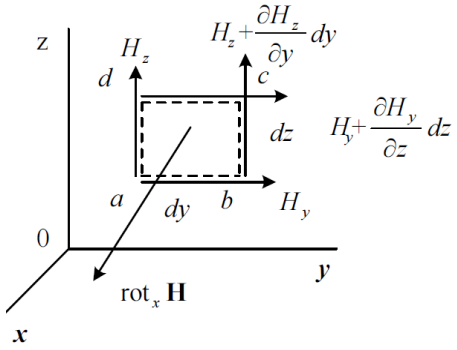


Fig. 1. To the explanation of the definition of rotH components in Cartesian coordinates

Regarding the *dydz* plane according to [3], we obtain the following equation (5):

$$\begin{aligned} &\left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z}\right)\mathbf{x}_0 + \left(\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x}\right)\mathbf{y}_0 + \\ &+\left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y}\right)\mathbf{z}_0 = \delta_x \mathbf{x}_0 + \delta_y \mathbf{y}_0 + \delta_z \mathbf{z}_0 \end{aligned} \tag{5}$$

where x_0, y_0, z_0 – orts.

The electric field strength is found from the concepts of trigonometry, which is shown by formula (6):

$$E = \sqrt{E_x^2 + E_y^2 + E_z^2} \tag{6}$$

The electric induction vector is determined from the characteristics of the dielectric according to equation (7):

$$\vec{D} = \varepsilon\vec{E} \tag{7}$$

where ε is the matrix of static electrical permillivity similar to expression (2) is determined from expression (8):

$$\varepsilon_i(E) = D_i / E, \quad i = x, y, z \tag{8}$$

The elements of the matrix ε' are determined by the characteristic of the dielectric $D = f(E)$.

Let's rewrite the magnetization characteristic $B=f(H)$ in another form [10] equation (9):

$$\vec{H} = N\vec{B} \tag{9}$$

where N is the diagonal matrix of inverse magnetic reluctance, defined similarly to (3).

The elements of the N matrix are determined from the magnetization curve in the directions of the main anisotropy axes.

Magnetic induction is found similarly to (5) and this is shown by formula (10):

$$B = \sqrt{B_x^2 + B_y^2 + B_z^2} \tag{10}$$

Then the magnetic reluctance are equal (11):

$$\nu_i(B) = H_i / B, \quad i = x, y, z \tag{11}$$

For software implementation of field calculation, it is necessary to approximate curves (4), (8), (11) and apply equations (1), (2). For this, it is necessary to know the initial conditions (the value of the field vectors at t=0) and the boundary conditions on the surface of the closed space throughout time.

Another case of nonlinearity is the electromagnetic field in a lattice ferromagnet. Such material is widely used for the construction of electrical devices as a magnetic conductor. Here we have a case of launching layers of ferromagnets and dielectrics. The calculation of the electromagnetic field in such an environment, taking into account the boundary conditions within homogeneous environments, is practically difficult to implement due to the large volume of the calculations. Therefore, in practical calculations, the flat-band environment is replaced by some anisotropic homogeneous environment, which is equivalent in electromagnetic terms to the first one.

In the transverse direction, we carry out the replacement, taking into account the serial connection of the ferromagnetic and non-magnetic gap magnetic resistances, and in the longitudinal direction – a parallel connection.

We apply the Cartesian coordinate system so that the *x*-axis passes across, and the *y, z* axis passes along the layers of the charged magnetic conductor.

Let's write the boundary conditions for the vectors $\vec{B}, \vec{H}$ at the interface of two homogeneous environments with the normal vector $\vec{n}_0$, respectively [12] equation (12), (13):

$$\left(\vec{B}_2 - \vec{B}_1\right)\cdot\vec{n}_0 = 0 \tag{12}$$

$$\vec{n}_0 \times \left(\vec{E}_2 - \vec{E}_1\right) = 0 \tag{13}$$

then

$$B_x = B_{fx} = B_{fd};\ H_y = H_{fy} = H_{dy};\ H_z = H_{fz} = H_{dz} \tag{14}$$

Here, the indices *f* and *d* indicate belonging to ferro- and diamagnets, and its absence – to an equivalent environment.

The equality of magnetic voltages in the transverse direction of the real and equivalent environment taking into account (14) can be expressed as dependence (15):

$$\left(\nu_f h_f + \nu_d h_d\right)B_x = \left(h_f + h_d\right)H_x \tag{15}$$

where ν are inverse magnetic reluctance, *h* are layer thicknesses.

From equation (15), we obtain the inverse static permillivity along the *x* axis. This is shown by equation (16):

$$\nu'_x = \xi\nu_f \tag{16}$$

where ξ is the anisotropy coefficient in the transverse direction.

$$\xi = \frac{h_f + \frac{\nu_d h_d}{\nu_f}}{h_f + h_d} \tag{17}$$

In the longitudinal direction, we obtain equation (18), which shows the equality of the magnetic fluxes:

$$\left(\frac{h_f}{\nu_f} + \frac{h_d}{\nu_d}\right)H_k = \left(h_f + h_d\right)B_k,\ k = y,z \tag{18}$$

From (19) here we find the coefficients ν'_x, ν'_y.

$$\nu'_k = \chi\nu_f,\ k = y,z \tag{19}$$

Where χ is the anisotropy coefficient of the later direction

$$\chi = \frac{h_f + h_d}{h_f + \frac{h_d\nu_f}{\nu_d}} \tag{20}$$

The value ν'_f is determined by the real induction field B_f in the ferromagnet according to equations (10), (11). The ratio of inductions in the transverse and tension in the longitudinal directions in ferromagnets and equivalent environment is determined in accordance with (14). The ratio (21) between the remaining projections is obtained by equations (9), (16), (19):

$$H_x = \xi H_{fx};\quad B_y = B_{fy}/\chi;\quad B_z = B_{fz}/\chi \tag{21}$$

According to the first expression (13) and the last two expressions (20), we get (22):

$$\nu_f = \sqrt{B_x^2 + \chi\left(B_y^2 + B_z^2\right)} \tag{22}$$

The values of B_x, B_y, B_z at each integration step are obtained from the equations of the electromagnetic field. Solving (21), for example, by Newton's method, we find according to (15)–(19) the magnetic reluctance of the matrix *N* in equation (8).

If we add electrical conductivities in the longitudinal direction, and supports in the transverse direction, then we get the expressions of the matrix elements (3). The electrical conductivity of the dielectric is zero

$$\gamma_x = 0;\quad \gamma_k = \frac{h_f\gamma_f}{h_f + h_d},\ k = x,z \tag{23}$$

If the external electromagnetic field changes so slowly that the polarization process is proportional to its changes, then the ratio between the vectors does not depend on the time derivatives of these vectors, and such a field is called quasi-stationary. The magnetic field due to bias currents can be neglected, but the change of the magnetic field over time should be taken into account. As a rule, the conditions of quasi-stationarity are always fulfilled. In a quasi-stationary field in a conductive environment in the absence of a extraneous currents field, the current density vector is determined only by conduction currents [10, 12].

Maxwell's second equation is another form of Faraday's law and has the form of equation (24):

$$\nabla\times\vec{E} = -\frac{\partial\vec{B}}{\partial t} \tag{24}$$

According to (24), the magnetic induction changes over time, i.e. it generates an electric field.

The equation of the continuity of the magnetic field according to [3, 4] has the form (25):

$$\nabla\cdot\vec{B} = 0 \tag{25}$$

Equations (1), (24), (25) are the basic equations of the quasi-stationary electromagnetic field.

The relationship between the magnetic induction vector $\vec{B}$ and the magnetic field strength $\vec{H}$ is expressed through the medium parameters according to (11).

According to equations (5), the system of electromagnetic field equations in Cartesian coordinates consists of scalar equations (26) written in the projections of the electric and magnetic field intensity vectors:

$$\left.\begin{aligned}
&\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = \gamma E_x; \quad \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = -\frac{1}{\nu}\frac{\partial H_x}{\partial t};\\
&\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} = \gamma E_y; \quad \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = -\frac{1}{\nu}\frac{\partial H_y}{\partial t};\\
&\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = \gamma E_z; \quad \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -\frac{1}{\nu}\frac{\partial H_z}{\partial t}.
\end{aligned}\right\} \tag{26}$$

According to (1), (5), the differential equation (26) takes the form (2):

$$\frac{\partial H_x}{\partial x} + \frac{\partial H_y}{\partial y} + \frac{\partial H_z}{\partial z} = 0 \tag{27}$$

The complete system of electromagnetic field equations consists of the system of equations (23), (24), and (27) only imposes restrictions on the distribution of field vector projections in space.

3. Results

For a complete picture of the field, the basic equations (1), (24), (25) should be supplemented with initial and boundary conditions. The initial conditions are unknown at $t=+0$ and represent the spatial position of the field vectors $H=H(x,y,z)$ and $E=E(x,y,z)$ in time. The equations relating to the boundary conditions establish the dependence between the sought-after functions on both sides of the integration boundary.

We will exclude the electric components in the equations of the electromagnetic field, since the boundary conditions are easier to determine relative to the magnetic field vectors. Taking into account this factor, we will make the operation $\nabla\times$ from the left and right parts (1), (2). Then we get the equation:

$$\nabla\times\nabla\times\vec{H} = \gamma\nabla\times\vec{E} \tag{28}$$

We take the vector $\nabla\times\vec{E}$ in the right part of (28) from Maxwell's second equation (24), replacing *B* with *H* according to (11), then we get equation (29):

$$\nabla\times\nabla\times\vec{H} = -\frac{\gamma}{\nu}\frac{\partial\vec{H}}{\partial t} \tag{29}$$

According to [3], expression (29) takes the form

$$\frac{\gamma}{\nu}\frac{\partial \vec{H}}{\partial t} = \nabla^2 \vec{H} - \nabla\left(\nabla \cdot \vec{H}\right) \tag{30}$$

Taking into account (24), (28), equation (30) takes the form (31):

$$\frac{\partial \vec{B}}{\partial t} = \frac{1}{\gamma}\nabla^2 \vec{H} \tag{31}$$

According to (31), the electric and magnetic field strengths are mutually perpendicular [7].

Figure 2 shows the results of calculating the electromagnetic field in a ferromagnetic plate made of E4A steel with a thickness of β = 2.4 mm. Curve 1 corresponds to a saturated magnetic system, curve 2 to an unsaturated magnetic system. Calculations were carried out in the MAPLE program.

Boundary and initial conditions have the form of equations:

$$H(0, t) = H(\beta, t) = 750 sin 314t; \quad H(z, 0) = 0 \tag{32}$$

Dependence (9) is determined by the main magnetization curve for this grade of steel.

The scalar form of this equation consists of equations written for the projections of the vector $\vec{H}$ on the coordinate axis (33):

$$\frac{\partial \vec{B}}{\partial t} = \frac{1}{\gamma}\frac{\partial^2 H}{\partial z^2} \tag{33}$$

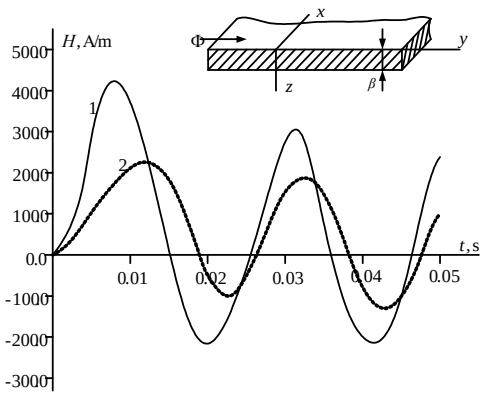


Fig. 2. Rated stress curves magnetic field inside the sheet

Consider another nonlinear problem in the field of electrodynamics. We are talking about the phenomenon of current displacement (skin effect) in rotating electric machines with deep rotor grooves. In such machines, eddy currents in the body of current conductors perform operational functions, and therefore must be taken into account with high accuracy.

Taking into account that the current in the conductor flows only along its length, and the vector of magnetic field tension, by the accepted assumption $\mu_{Fe} = \infty$, is perpendicular to the walls of the groove, we get a clear example of a plane electromagnetic wave.

By choosing a rectangular coordinate system, you can write down the conditions (34):

$$\vec{E} = \vec{x}_0 E; \qquad \vec{H} = \vec{y}_0 H \tag{34}$$

Then we come to equations (23), (26), which, taking into account the dimensions of the groove, take the form (35):

$$-\frac{\partial H}{\partial z} = \frac{b}{a}\gamma E; \qquad -\frac{\partial E}{\partial z} = \frac{1}{\nu}\frac{\partial H}{\partial t} \tag{35}$$

where *a*, *b* are respectively the width of the groove and the current conductor.

Excluding the electric field intensity from system (35), we obtain the equation (36):

$$\frac{\partial H}{\partial t} = c^2 \frac{\partial^2 H}{\partial z^2} \tag{36}$$

where

$$c^2 = \frac{a\nu}{b\gamma} \tag{37}$$

The derivative $\frac{\partial \Phi}{\partial t}$ can be represented as (38):

$$\frac{\partial \Phi}{\partial t} = \frac{\partial}{\partial t}\int_s \vec{B} dS \tag{38}$$

Where *S* is the surface through which the induction flux determines the flux Φ.

Taking into account formula (11) and (35), expression (38) takes the form (39):

$$\frac{\partial \Phi}{\partial t} = l\int_0^{\xi} \frac{1}{n}\frac{\partial H}{\partial t} dz = l\int_{\xi}^{0} \frac{\partial E}{\partial z} dz = \left[E(0) - E(\xi)\right] l \tag{39}$$

Hence, the tension of the groove part is equal to (40):

$$U = E(0) \cdot l \tag{40}$$

This voltage reflects surface phenomena throughout the grooved part of the conductor. Its calculation is related to the calculation of the electromagnetic field in the groove.

Consider the method of determining the boundary conditions for the electromagnetic field equation (29). According to Biot-Savar law, the strength of the magnetic field *H* depends on the current flowing in the conductor, as well as on the distance from it, so the boundary conditions must be sought on the basis of this law. For practical use, another form of its recording is more convenient – the law of full current. So, knowing the strength of the current in the groove, we determine the boundary conditions for the magnetic field tension vector using the full current law [6, 11]

$$\int \vec{H} d\vec{l} = I \tag{41}$$

Choosing the path of integration along the line that passes along the surface of the groove and in the body of the iron, we replace the integral (41) by the sum

$$\int \vec{H}\ d\vec{l} = \int_a \vec{H}_a d\vec{l} + \int_c \vec{H}_c d\vec{l} = 0 = I \tag{42}$$

where $\vec{H}_a, \vec{H}_c$ are the magnetic field intensity vectors on the groove surface and in the steel; *a* – the width of the groove opening zone; *c* is the length of the trajectory in the ferromagnet; *I* – conductor current.

We write the integral along the path that passes along the bottom of the groove (42):

$$\int_a \vec{H}_a d\vec{l} + \int_c \vec{H}_c d\vec{l} = 0 \tag{43}$$

The magnetic voltage drop in steel is close to zero. Therefore, it can be assumed with sufficient accuracy that the integral in steel is zero. In the groove opening zone, vector *H* has only a tangential component ($\vec{H} = \vec{y}_0 \cdot H_y$ – in Cartesian coordinates), and H_y=*const.*

Then we determine the boundary conditions from (42), (43). The initial conditions of the integrating functions are assumed to be zero.

In the upper and lower parts of the groove according to Fig. 2

$$H_y = H(0) = \frac{I}{a}; \quad H_c = H(10) = 0 \tag{44}$$

We find the electric field strength from equation (35). The equation has the form (45):

$$E = -\frac{a}{\gamma b}\frac{\partial H}{\partial z} \tag{45}$$

Having the value *E*(0), the voltage *U* of the conductor is found according to (40). Equations (36) together with the boundary conditions (44) constitute the boundary value problem in the time domain for the differential equations of the electromagnetic field. Therefore, the current line voltage is determined by the value of the electric field intensity vector on the surface of the groove (z=0).

The described method makes it possible to solve the problem of displacement of the current into the groove in the general case. This problem is solved on the basis of numerical methods of analysis using computer technology [14]. At the same time, the differential equations must be replaced by finite-difference equations. Let's draw up an explicit differential scheme for determining the field strength at the nodes of the spatial grid at different moments of time.

$$H(z,t+\Delta t)=H(z,t)+ \\ +c\left[H(z-\Delta z,t)-2H(z,t)+H(z+\Delta z,t)\right] \tag{46}$$

where c is found according to (37).

In the case of rigid differential equations, we use an implicit difference scheme, which for equation (36) has the form

$$dH(z,t)=aH(z-\Delta z,t+\Delta t)+ \\ +bH(z,t+\Delta t)+cH(z+\Delta z,t+\Delta t) \tag{47}$$

where

$$a=c=1;\quad d=-\frac{\gamma}{\nu_0}\frac{(\Delta z)^2}{\Delta t};\quad b=d-2 \tag{48}$$

Here $\nu_0=\frac{1}{\mu_0}=\frac{1}{4\pi}\cdot 10^7\ \frac{m}{H}$ is the magnetic constant.

In equation (47), the known values of H and E at point z at time t are related to three unknown values of the same functions at points $z-\Delta z$, z, $z+\Delta z$ at time $t+\Delta t$.

Figure 3 shows the time dependence of the voltage per unit length of the groove section of the current conductor in the state of the given current, calculated according to formulas (45), (46).

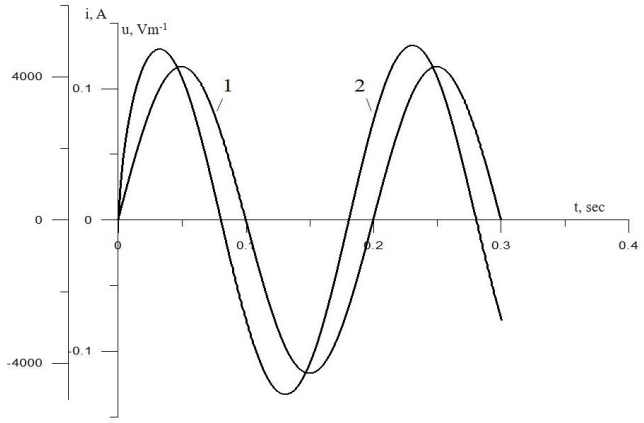


Fig. 3. Calculated curves of current (1) and voltage per unit length (2) of the slot current conductor in the state of sudden switching on the current source I = 4282sin(31.4t). Groove dimensions: h×l×a=0.038 m×0.23 m×0.005 m

Figure 4 shows the spatial discretization of the groove and the distribution of the field strength in it. The segment *ab* corresponds to the voltage, and *bc* is the internal electromotive force.

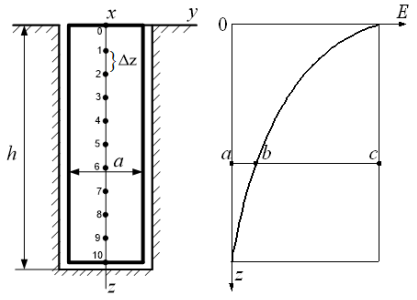


Fig. 4. Spatial discretization of the groove and distribution field strength in it

Figure 3 shows that the effect of current displacement is observed in the deep groove of the rotor of the electric machine. On the surface of the groove, the voltage is the greatest, which has a positive effect on the starting characteristics of electric asynchronous motors.

4. Conclusions

The proposed equations and methods of the theory of electromagnetic circuits and the electromagnetic field take into account such complex phenomena as the saturation of the steel of magnetic conductors, the surface effect in current conductors, the asymmetry of electrical circuits, the mutual rotational movement of electrical circuits, etc. The differential equations are written in the normal Cauchy form and are non-rigid and can be integrated by explicit methods that are simple in computer implementation. The equations appearing in the models make it possible to calculate with sufficient accuracy the quantitative characteristics of the object under study, to predict its behavior in various nominal and emergency modes, as well as to carry out optimization.

The calculation process is carried out based on the solution of the two-point boundary value problem of the differential equations of the electromechanical state. The differential equations of the electromechanical state are integrated based not on the given initial conditions, but on those that exclude the transient reaction and make it possible to enter directly into the steady process, bypassing the transient one.

The use of mathematical modeling and computer simulation methods makes it possible to abandon field experiments, which in many cases are difficult and expensive. Practical tasks of calculation and analysis of transient processes in electro-technical devices should be carried out only by mathematical modeling methods, which will allow them to be correctly designed and operated. The proposed method can be adapted to construct the equations of the electromagnetic field in moving media.

References

[1] Basu P. K., Dhasmana H.: Electromagnetic Theory Fundamentals. Electromagnetic Theory. Springer, Cham, 2023 [https://doi.org/10.1007/978-3-031-12318-4_1].
[2] Chew W. C.: Lectures on Electromagnetic Field Theory. Purdue University, 2020.
[3] Davis B. S.: Understanding the Electromagnetic Field. World Scientific Publishing Company, Singapore 2023.
[4] Janaswamy R.: Engineering Electrodynamics: A collection of theorems, principles and field representations. IOP Publishing, Bristol, 2020.
[5] Kostiuchko S., Polishchuk M., Zabolotnyi O., Tkachuk A., Twarog B.: The Auxiliary Parametric Sensitivity Method as a Means of Improving Project Management Analysis and Synthesis of Executive Elements. Miraz M. H. et al. (eds): Emerging Technologies in Computing. iCETiC 2021. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 395, Springer, Cham 2021, 174–184 [https://doi.org/10.1007/978-3-030-90016-8_12].
[6] Müller C.: Foundations of the Mathematical Theory of Electromagnetic Waves. Springer, Berlin, 2013 [https://doi.org/10.1007/978-3-662-11773-6].
[7] Rauf B.: Electrical Engineering for Non-Electrical Engineers. 3rd Edition. River Publishers, 2022.
[8] Raychaudhuri A. K.: Classical Theory of Electricity and Magnetism: A Course of Lectures. Springer, 2022.
[9] Riad S. M., Salama I. M.: Electromagnetic Fields and Waves: Fundamentals of Engineering. McGraw-Hill Education, 2020.
[10] Rizzoni G., Kearns J.: Fundamentals of Electrical Engineering. 2nd Edition. McGraw-Hill Education, 2022.
[11] Shadid W. G.: Electric Model for Electromagnetic Wave Fields. IEEE Access 9, 2021, 88782–88804 [https://doi.org/10.1109/ACCESS.2021.3090862].
[12] Tchaban V., Kostiuchko S., Krokhmalny B.: Equations of State Variables of Electromagnetic Circuits in Engineering Education of MEMS-Specialists. IEEE XVIIth International Conference on the Perspective Technologies and Methods in MEMS Design – MEMSTECH, 2021, 78-81 [https://doi.org/10.1109/MEMSTECH53091.2021.9468082].
[13] Tchaban V.: Nova elektrotekhnika. Prostir M, L'viv, 2019.
[14] Zhang X.-Z.: Flow Measurement by Electromagnetic Induction: Theory and numerical methods. IOP Publishing, Bristol, 2020.

Ph.D. Viktor Lyshuk
e-mail: v.lyshuk@lntu.edu.ua

Associate professor, Faculty of Computer and Information Technologies, Department of Electronics and Telecommunications, Lutsk National Technical University, Ukraine.

https://orcid.org/0000-0003-4049-8467

Prof. Vasyl Tchaban
e-mail: vasyl.y.chaban@lpnu.ua

Professor, Institute of Energy and Control Systems, Department of Theoretical and General Electrical Engineering, Lviv Polytechnic National University, Lviv, Ukraine.

https://orcid.org/0000-0002-0697-0337

Ph.D. Anatolii Tkachuk
e-mail: a.tkachuk@lntu.edu.ua

Scientific direction: development and implementation of new electronic and SMART machine systems & devices, innovative IT developments for the implementation of Industry 4.0 and 5.0 technologies, design and implementation of innovative IT systems of artificial intelligence, robotics, Internet of Things (IoT) technologies, using advanced cloud technologies and the latest software.

https://orcid.org/0000-0001-9085-7777

Ph.D. Valentyn Zablotskyi
e-mail: v.zablotsky@lntu.edu.ua

Research interests: Technological support of wear resistance conjugate parts machines and devices working surfaces. Research physical quantities of sensors functional features. Features of the optical communication lines organization and operation.

https://orcid.org/0000-0002-2921-0031

Ph.D. Yosyp Selepyna
e-mail: y.selepyna@lntu.edu.ua

Associate professor, Faculty of Computer and Information Technologies, Department of Electronics and Telecommunications, Lutsk National Technical University.
Research interests: Modeling of electronic devices and systems. Digital signal processing and coding in telecommunication systems and networks.

https://orcid.org/0000-0002-2421-1844

http://doi.org/10.35784/iapgos.5495 received: 22.10.2023 | revised: 18.02.2024 | accepted: 03.03.2024 | available online: 31.03.2024

OPTICAL SPECKLE-FIELD VISIBILITY DIMINISHING BY REDUCTION OF A TEMPORAL COHERENCE

Mikhaylo Vasnetsov[1], Valeriy Voytsekhovich[1], Vladislav Ponevchinsky[1], Nataliia Kachalova[1,2], Alina Khodko[1], Oleksandr Mamuta[1], Volodymyr Pavlov[3], Vadym Khomenko[1], Natalia Manicheva[4]

[1]Institute of Physics of the National Academy of Sciences of Ukraine, Kyiv, Ukraine, [2]L.M. Litvinenko Institute of Physical and Organic Chemistry and Carbon Chemistry of the National Academy of Sciences of Ukraine, Kyiv, Ukraine, [3]Vinnitsia National Technical University, Vinnitsia, Ukraine, [4]Odessa Polytechnic National University, Odessa, Ukraine

Abstract. *The paper is a report of an experimental study to suppress the speckle structure of a coherent optical field. The technique proposed is based on the reduction of the temporal coherence utilizing enriching the output spectrum of Nd:YVO_4 laser with intra-cavity second harmonic generation by additional emission lines. Temperature-controlled simultaneous emission of two components at 1.063 μm and 1.066 μm with nearly equal intensities in IR is achieved. In the second-harmonic output the emission lines 531.7 nm, 532.3 nm and 532.8 nm were recorded. The influence of the spectrum variation on the formation of a speckle field was checked. Successfully removed intensity zeros and reduced contrast (visibility) from 0.92 to 0.65 in a light scattered by a ground glass diffuser at the angle 35°. A simple consideration of the speckle field dumping mechanism is presented.*

Keywords: speckle field, coherence, visibility

OPTYCZNE ZMNIEJSZANIE WIDOCZNOŚCI POLA PLAMKOWEGO POPRZEZ REDUKCJĘ SPÓJNOŚCI CZASOWEJ

Streszczenie. *Praca jest raportem z badań eksperymentalnych mających na celu stłumienie struktury plamkowej spójnego pola optycznego. Proponowana technika polega na redukcji spójności czasowej poprzez wzbogacenie widma wyjściowego lasera Nd:YVO_4 o wewnątrzwnękową generację drugiej harmonicznej przez dodatkowe linie emisyjne. Osiągnięto kontrolowaną temperaturowo jednoczesną emisję dwóch składników przy 1,063 μm i 1,066 μm przy prawie równych intensywnościach w podczerwieni. Na wyjściu drugiej harmonicznej zarejestrowano linie emisyjne 531,7 nm, 532,3 nm i 532,8 nm. Sprawdzono wpływ zmiany widma na powstawanie pola plamkowego. Udało się usunąć zera intensywności i zmniejszyć kontrast (widoczność) z 0,92 do 0,65 w świetle rozproszonym przez dyfuzor ze szkła szlifowanego pod kątem 35°. Przedstawiono proste rozważenie mechanizmu zmniejszenia pola plamkowego.*

Słowa kluczowe: pole plamkowe, spójność, widzialność

Introduction

The interference in wave optics manifests itself as observable intensity variation over space and/or time, usually represented in the form of periodic distribution (fringes pattern) [3]. However, the interference in a general sense is a summation of participating field vectors which cannot be resolved in time when the oscillating frequencies differ substantially, or the mutual coherence is low. In the opposite case, under the circumstances of nearly plane monochromatic waves with a little frequency shift, the interferometric fringes can be observed with the naked eye. The measure of the visibility is a contrast of the pattern, i.e. the ratio of the residual of the intensity maximum and minimum, to their sum. In the case of several interfering waves (or fields with a wide spatial spectrum), the sum of their intensities equals the average intensity over a given time or space domain, but the spatial pattern is a matter of their temporal coherence. It could be expected that the field distribution will exhibit a smooth profile, but this is not true in general. For an optical vector field with irregularly varying polarization, the intensity distribution does not possess special peculiarities but the polarization state obeys some topological rules [9, 13]. Below we restrict ourselves by the situation of a scalar field, i.e. with the same polarization state across the observation region.

An example of an irregular optical field is a speckle structure usually observed in a scattered laser (monochromatic) radiation [5]. This unavoidable feature accompanies laser beam reflection by rough surface or propagation through inhomogeneous media like turbulent atmosphere, liquid flow or another diffuser. In any case, interference of monochromatic wavelets results in a modulated intensity distribution. More correct is to explain the result as a random phase-amplitude field formation, where the constructive summation generates bright speckles surrounded by dark borders of destructive interference containing zero amplitude points [2]. The presence of exact intensity zeroes makes the visibility of a speckle field equal to unity. We note that the number of zero-intensity minima coincides with the number of the bright maxima, i.e. there is one amplitude zero (phase vortex) between neighbor speckles.

The presence of speckle structure seriously degrades the quality of the image, even where the scattering is relatively small. For instance, in optical tomography coherent speckle noise reduces the contrast and makes difficult to resolve boundaries between highly scattering structures in tissue [11]. In this view, the problem of speckle noise suppression is of importance in many optical applications [10, 14].

The basic principle of speckle suppression is an averaging of different speckle patterns that is their overlapping in space or time resulting in the reduced speckle contrast. These speckle patterns could be obtained using a moving diffuser, polarization diversity, wavelength diversity, and/or angle diversity [6]. The averaging of the speckle structure is usually achieved using a rapidly moving (for example, rapidly rotating) light-scattering diffuser [4]. A current review of methods for speckle-noise suppression can be found in [7, 8].

1. Reduction of laser emission coherence

The approach accepted to the problem solution in this study consists in the diminishing of the temporal coherence of the optical field in order to damp the speckle structure. For this research, we used a commercial cw Nd:YVO_4 laser with intra-cavity second harmonic generation using BiB_3O_6 crystal (BIBO). The laser output consists normally of IR emission (1064 nm) and double-frequency radiation (532 nm) which makes it a convenient tool for our goal. Light emission power (in green light) was measured 7 mWt.

The peculiarity of this laser is sporadically appearing additional oscillating wavelengths. These oscillating spectral components were detected in a first diffraction order of a grating 1400 mm^{-1} as schematically shown in Fig. 1. An explanation of the origin of additional components lies in the possibility of an emission line at 1083 nm to appear, apart main emission at 1064 nm (we refer to the details in the report [15]). As a result, three wavelengths of laser emission (532 nm, 537 nm, 542 nm) were observed. The output at these lines oscillating simultaneously originates owing to the second-harmonic generation of 1064 nm and 1083 nm, and sum-frequency generation. All these spectral

components satisfy the longitudinal synchronism conditions for frequency doubling and are not separated within the output laser beam. In this view, the emission spectrum can be controlled without variation a beam direction and divergence. As it is seen, the emission at 532 nm is noticeably stronger, due to the relative power difference in the oscillations at 1064 nm and 1083 nm. The measured fundamental emission spectrum is shown in the inset of Fig. 1.

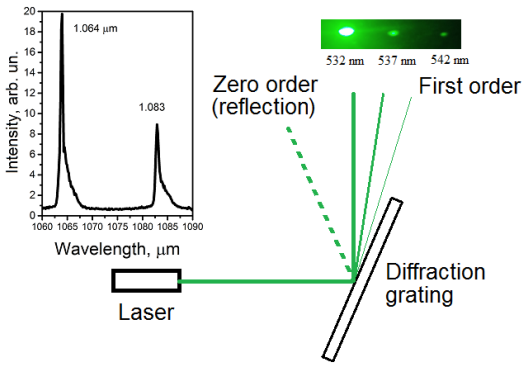


Fig. 1. Elementary scheme for observation of spectral components in the emission of Nd:YVO$_4$/BIBO laser (second harmonic generation). The laser output beam is directed to a grating 1400 mm^{-1}. In the first order of diffraction, apart of the strongest diffracted beam with the wavelength 532 nm two satellites are seen with the wavelengths 537 and 542 nm. Insert shows the measured fundamental frequency spectrum with an additional emission line at 1.083 μm

The outcoming problem that had to be solved was stabilization and switching the laser output regimes (single line operation at 532 nm or three-line operation at 532 nm, 537 nm and 542 nm). However, this regime was found less convenient for the experiment. The reason was a necessity to cool the laser to a temperature lower than 15° C for a steady operation. Instead, we detected a fine splitting of the main oscillation line under temperature variation of the environment by heating the laser. (For the details we refer to the spectroscopic measurements of temperature-dependent stimulated emission in Nd^{3+}:YVO_4 [12]). Therefore, this tool can be better used for our goal.

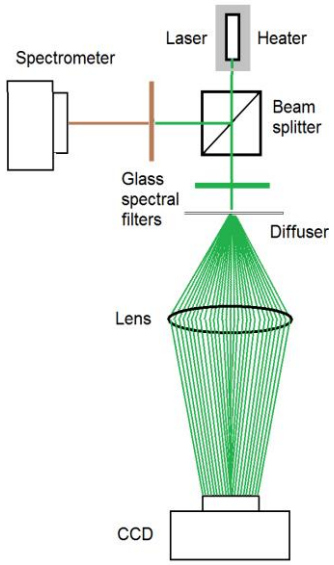


Fig. 2. General scheme for the observation of the laser output spectrum variation and the speckle-field transformations with temperature. Laser output contains both fundamental frequency emission in IR region and second-harmonic green emission. The heater system is used to control the laser temperature. Glass spectral filters are used to cut off the IR radiation in the field scattered by the ground glass diffuser and separate IR or double-frequency components detected by the spectrometer. The CCD camera is placed at the focal plane of 80cm focal distance converging lens

Below we report on the results obtained with laser emission controlled by the temperature. The experimental scheme is presented in Fig. 2. A laser head temperature can vary from 15°C to 50°C. The output beam containing the fundamental-frequency IR component and the doubled-frequency radiation was split by a beamsplitter to measure simultaneously the spectral content of the fundamental frequency emission by the spectrometer Ocean Optics USB 4000 and record the speckle-field pattern. A ground glass diffuser scattered green light isolated from IR component and CCD camera (Lumenera LM 135 M) was placed in the focal plane of the lens.

We observed a gradual spectral transformation in a temperature region from 15°C to room temperature and then from room temperature to 45°C. First, the emission at 1083 nm line was detected at 15°C and disappeared at about below 25°C. This regime was not stable enough for precise temperature control. Another extra emission line at 1066 nm was detected in the temperature interval after 40°C. With better stability and reproducibility we performed the experiments in this interval. The corresponding spectra of single-line emission at 27.5°C and double-line emission at 45°C are shown in Fig. 3.

Again, the emission in this bichromatic regime (IR) results in the three-chromatic output in the second harmonic radiation, as seen in Fig. 4. The measured spectrum contains the components at 531.7 nm, 532.3 nm and 532.8 nm. Fine tuning of the relative intensities of the emission lines was achieved by the pump LED current control.

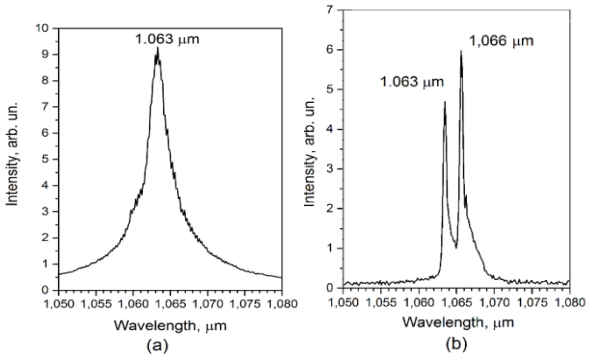


Fig. 3. Laser output spectrum transformation with temperature: (a) at the room temperature 27.5°C, (b) 45°C

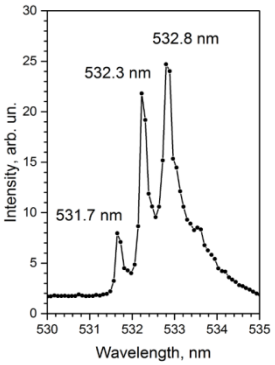


Fig. 4. Measured spectrum of second-harmonic output with three maxima at the temperature 45°C

2. Visibility measurements

To start the visibility measurements, we elaborated a laboratory set up with a test object (light diffuser) as a ground glass plate (Fig. 2). Under the temperature control we monitored the scattered field with a CCD camera, and a processing of the digital image (intensity cross-section) gives the visibility quantitative measure.

First, the spectrum transformation from single oscillation frequency to the three-chromatic regime, as described above, did not produce noticeable changes in the field scattered at small angles, i. e. along the incident beam direction. To detect the scattered pattern variations, we have modified the general scheme with a circular aperture separating the light along a definite direction This choice ensures an isotropic scattering to the angles exceeding 45° without polarization variation.

The modified experimental scheme is shown in Fig. 5 with an example of the speckle field observed in the far-field zone (focal plane of the lens).

With the angle of scattering chosen 35° by the aperture with the diameter 10 mm placed on a distance 10 cm from the illuminated spot on the diffuser we definitely observed radical transformation of the pattern by eye. Instead of highly contrast speckle field, a nearly uniform intensity distribution appears at the observation plane. Figure 6a,b illustrates the pattern variation which occurs with the three-chromatic emission.

In a quantitative measure the visibility reduction was detected from 0.92 to 0.65, as calculated from the measured cross-sections in Fig. 6 (c) and (d), recorded correspondingly for the single IR emission line and the bichromatic regime.

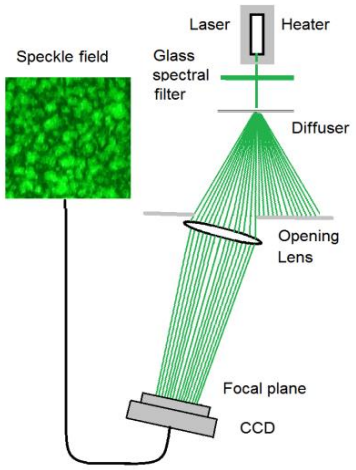


Fig. 5. A principal scheme of a speckle field generation and processing. Thin ground glass plate serves as a diffuser. The glass filter cuts off the IR radiation. A CCD camera is placed at the focal plane of 80cm focal distance converging lens

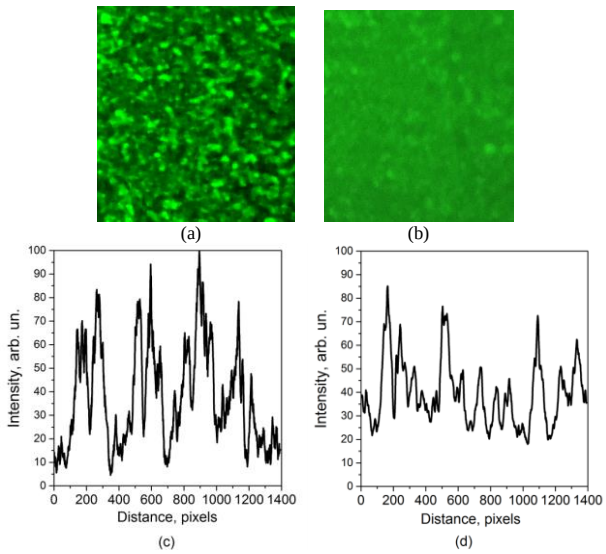


Fig. 6. A comparison of the obtained speckle patterns in a scattered field with a single radiating line (a) and three lines (b), and the corresponding intensity cross-sections (c), (d). The scale givet in pixels avjunts to 7 mm window of CCD camera

3. Discussion of the results

To give a simple explanation of the speckle-field smoothing by used three-chromatic laser emission and its dependence on the angle of scattering, we can model a diffuser as a thin plate with continuous superposition of periodic phase gratings of spatial frequencies $\Omega = 2\pi/\Lambda$, where Λ is a spatial period of a corresponding grating. We assume an optical beam illuminating the diffuser to incident normally to the plate and have a dimension d in the cross section, as depicted schematically in Fig. 7.

In the far field, scattering to the angle θ is produced by the spatial components which diffract the incident light into this direction, according to the diffraction law written for the first diffracted order and normal incidence as

$$\Lambda \sin\theta = \lambda \tag{1}$$

where λ is the wavelength. Then, variation of the coming light wavelength to a small value $\Delta\lambda$ will not destroy the speckle pattern, but slightly move it along the spatial coordinate θ:

$$\Delta\theta = \Delta\lambda/\Lambda \cos\theta \tag{2}$$

or

$$\Delta\theta = (\Delta\lambda/\lambda)\tan\theta \tag{3}$$

Evidently, the pattern observed at small scattering angles ($\theta \ll 1$) will remain nearly the same with the wavelength variation. That is, in the case of bichromatic incident field, i. e. when both wavelengths λ and $\lambda + \Delta\lambda$ are present in the illuminating beam (for simplicity we consider only two spectral components) the generated by them speckles will overlap without spatial shift and add incoherently owing to the frequency difference.

To damp the speckle pattern, namely remove the intensity zeroes and thus diminish the visibility, we must achieve the shift of the constituent patterns at least to the mean spatial angle of an individual speckle. At the far field, the angular speckle dimension is estimated as λ, thus we have

$$(\Delta\lambda/\lambda)\tan\theta > \lambda/d \tag{4}$$

and therefore

$$\tan\theta > \lambda^2/d\Delta\lambda \tag{5}$$

In our experimental situation, an estimation $\Delta\lambda$ = 1 nm λ = 500 nm and d = 1 mm gives $\tan\theta$ = 0.25 or the corresponding angle θ = 14°, in a good agreement with the experiment.

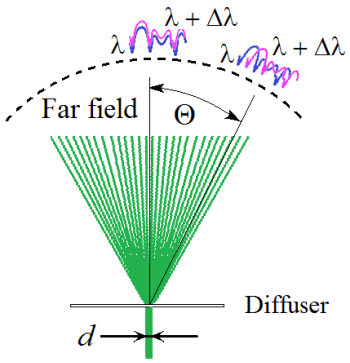


Fig. 7. Schematic explanation of the summation of speckle fields generated separately by spectral components λ and $\lambda + \Delta\lambda$. In the small-angle scattering, the speckle patterns overlap without angular shift. At the angle Θ the fragments of the speckle pattern are angularly shifted

4. Conclusions

In the summary of the research performed we can report on "proof in principle" result of the influence of a laser emission spectrum on the speckles visibility. In general, the main conclusion of this study is formula (5) which connects the spectral width with the angle of the speckles overlapping. We note this result is in a good consistence with the reported for natural light speckles observation [1].

The theoretical part of the work was accompanied with experimental verification. We elaborated a laboratory set up with temperature-controlled bichromatic emission in IR of Nd:YVO_4 laser at the wavelengths 1.063 μm and 1.066 μm. This regime permits to reduce the temporal coherence of the output double-frequency radiation with keeping the quality of a laser beam spatial divergence. Then, we checked the influence of a spectrum transformation on the observed speckle pattern in the far field. In the small-angle scattering we did not detect noticeable variation of the pattern. At higher angles (35°) we fairly observed the damping of the speckle-field pattern.

Acknowledgements

The research was supported by the grant of the National Research Foundation of Ukraine 2022.01/0135.

References

[1] Arkhelyuk A. D. et al.: Characteristics of polychromatic speckle fields and temporal coherence. Proc. SPIE 8338, 2011, 833805.
[2] Baranova N. B. et al.: Wave-front dislocations: topological limitations for adaptive systems with phase conjugation. J. Opt. Soc. Am. 73, 1983, 525–528.
[3] Born M., Wolf E.: Principles of Optics. Cambridge University Press, Cambridge 1999.
[4] Dayan Li et al.: Speckle suppression by doubly scattering systems. Applied Optics 52, 2013, 8617–8626.
[5] Goodman J.: Some fundamental properties of speckle. J. Opt. Soc. Am. 66, 1976, 1145–1150.
[6] Goodman J. W., Speckle Phenomena in Optics: Theory and Applications, 1st ed. Roberts & Company, 2006, 125–149.
[7] Kompanets I., Zalyapin N.: Methods and Devices of Speckle – Noise Suppression. Optics and Photonics Journal, 10, 2020, 219–250.
[8] Kukharchuk V. et al.: Features of the angular speed dynamic measurements with the use of an encoder. Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska – IAPGOS 12(3), 2022, 20–26.
[9] Nye J., Wright F. J.: Natural Focusing and Fine Structure of Light: Caustics and Wave Dislocations. Am. J. Phys. 68(8), 2000, 776 [https://doi.org/10.1119/1.19543].
[10] Pavlov S. et al.: Electro-optical system for the automated selection of dental implants according to their colour matching. Przegląd elektrotechniczny 93(3), 2017, 121–124.
[11] Schmitt J. M., Xiang S. H., Yung K. M.: Speckle in Optical Coherence Tomography. Journal of Biomedical Optics 4, 1999, 95–105.
[12] Turri G. et al.: Temperature-dependent stimulated emission crosssection in Nd3+:YVO4 crystals. J. Opt. Soc. Am. 26(11), 2009.
[13] Vasnetsov M., Soskin M., Pasko V.: Topological configurations of cross-coupled polarization singularities in a space-variant vector field. Optics Communications 63, 2016, 181–187.
[14] Wójcik W., Pavlov S., Kalimoldayev M.: Information Technology in Medical Diagnostics II. London: Taylor & Francis Group, CRC Press, Balkema book, 2019 [https://doi.org/10.1201/ 9780429057618].
[15] Zhang Z. et al.: Intra-cavity second harmonic generation with Nd:YVO4/BIBO laser at 542 nm. Optics Communications 267, 2006, 487–490.

Prof. Mikhaylo Vasnetsov
e-mail: mikhail.v.vasnetsov@gmail.com

Leading researcher, Head of the Department of Optical Quantum Electronics at the Institute of Physics of the National Academy of Sciences of Ukraine. Laureate of the State Prize of Ukraine in the field of science and technology (2021). Professor of the Department of Kyiv Academic University.
The main direction of scientific activity is the optics and laser physics, holography, liquid crystals, vortex, photonics, optoelectronics, nonlinear optics, applied optics.

https://orcid.org/0000-0001-9095-5334

Ph.D. Oleksandr Mamuta
e-mail: mamuta.aleksandr@gmail.com

Researcher of the Department of Coherent and Optics at the Institute of Physics of the National Academy of Sciences of Ukraine.
the main direction of scientific activity is the optics and laser physics, applied optics, ultrafast laser spectroscopy, processes of physical and technical processing.

https://orcid.org/0000-0002-6404-5879

Ph.D. Valeriy Voytsekhovich
e-mail: valvvs55@gmail.com

Senior researcher of the Department of Coherent and Optics at the Institute of Physics of the National Academy of Sciences of Ukraine.
The main direction of scientific activity is the physical electronics, optics and laser physics, biophysics, physics of atoms and molecules, laser spectroscopy, nonlinear optics, applied optics.

https://orcid.org/0000-0002-7196-729X

M.Sc. Volodymyr Pavlov
e-mail: machinehead6926@gmail.com

M.Sc., scientific researcher Vinnytsia National Technical University. Scientific direction – biomedical information optoelectronic and laser technologies for diagnostics and physiotherapy influence. President VNTU student chapter

https://orcid.org/0000-0002-0717-7082

M.Sc. Vladislav Ponevchinsky
e-mail: ponevchinsky@nas.gov.ua

Researcher of the Department of Optical Quantum Electronics at the Institute of Physics of the National Academy of Sciences of Ukraine.
The main direction of scientific activity is the laser optics, computers and electronics, with experience in image processing, speckles and optical vortices.

https://orcid.org/0009-0003-7250-3619

M.Sc. Vadym Khomenko
e-mail: khomenko.vadim@gmail.com

Junior researcher of the Department of Coherent and Optics at the Institute of Physics of the National Academy of Sciences of Ukraine.
The main direction of scientific activity is the optics and laser physics, ultrafast laser spectroscopy, software development, data analysis the development and implementation of: optical schemes of various complexity, as well as automated systems of information collection and processing.

https://orcid.org/0000-0002-7926-9277

Ph.D. Nataliia M. Kachalova
e-mail: kachalova.nataliya@gmail.com

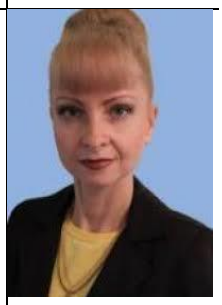
Senior researcher of the Department of Coherent and Optics at the Institute of Physics of the National Academy of Sciences of Ukraine.
The main direction of scientific activity is the Optics and laser physics, laser spectroscopy, femtochemistry, applied optics.

https://orcid.org/0000-0002-9956-5380

Ph.D. Natalia Manicheva
e-mail: n.v.manichev@op.edu.ua

Department of Biomedical Engineering Odessa Polytechnic National University, Odessa, Ukraine. Scientific direction – biomedical information optoelectronic and laser technologies for diagnostics and physiotherapy influence.

https://orcid.org/0000-0002-3043-5342



Ph.D. Alina Khodko
e-mail: khodkoalina@gmail.com

Researcher of the Department of Coherent and Optics at the Institute of Physics of the National Academy of Sciences of Ukraine.
the main direction of scientific activity is the laser physics, ultrafast laser spectroscopy, time-resolved pump-probe experiments, advanced methods of data processing and analysis.

https://orcid.org/0000-0001-5685-1013

http://doi.org/10.35784/iapgos.5610 *received: 24.11.2023 | revised: 13.02.2024 | accepted: 10.03.2024 | available online: 31.03.2024*

QUALITY INDICATORS OF DETECTION OF SIDE RADIATION SIGNALS FROM MONITOR SCREENS BY A SPECIALIZED TECHNICAL MEANS OF ENEMY INTELLIGENCE

Dmytro Yevgrafov, Yurii Yaremchuk
Vinnytsia National Technical University, Vinnytsia, Ukraine

Abstract. *The problem of calculating the quality indicators of detection of side radiation signals from static images on the monitor screen by a specialized technical means of enemy intelligence is solved, which implements an asymptotically optimal compatible algorithm for detecting side radiation signals and estimating the duration of image immutability on the monitor screen. The reliability of the calculation results was confirmed by modeling detection processes in the Mathcad and Excel software environments.*

Keywords: side electromagnetic radiation and targeting, Neumann-Pearson criterion, probability of false alarm, probability of signal miss

WSKAŹNIKI JAKOŚCI DO WYKRYWANIA SYGNAŁÓW PROMIENIOWANIA BOCZNEGO Z EKRANÓW MONITORÓW PRZEZ WYSPECJALIZOWANE ŚRODKI TECHNICZNE WYWIADU WROGA

Streszczenie*. W artykule rozwiązano problem obliczania wskaźników jakości wykrywania sygnałów promieniowania bocznego ze statycznych obrazów na ekranie monitora przez wyspecjalizowany techniczny środek rozpoznania wroga, który implementuje asymptotycznie optymalny wspólny algorytm wykrywania sygnałów promieniowania bocznego i szacowania czasu trwania stabilności obrazu na ekranie monitora. Wiarygodność wyników obliczeń została potwierdzona poprzez modelowanie procesów detekcji w środowisku oprogramowania Mathcad i Excel.*

Słowa kluczowe: zakłócenia elektromagnetyczne i przesłuchy, kryterium Neumanna-Pearsona, prawdopodobieństwo fałszywego alarmu, prawdopodobieństwo braku sygnału

Introduction

In [9], the asymptotically optimal algorithm of side radiation signals (SR) from monitor screens using liquid-crystal structures (LCS) is synthesized, which is used in modern specialized technical means of enemy intelligence (STMEI). In the STMEI, the functionality of the likelihood ratio (FLR) is calculated, which for $T_a >> T_f$ amounts to:

$$L(T_a) = \frac{4}{N_0 T_a}\sum_{k=1}^{K}\left[\left(\int_0^{T_a} x(t)\cos\left(\frac{2\pi kt}{T_f}\right)\mathrm{d}t\right)^2 + \left(\int_0^{T_a} x(t)\sin\left(\frac{2\pi kt}{T_f}\right)\mathrm{d}t\right)^2\right] \quad (1)$$

where T_f – the period of following the frame scan of the monitor screen with the LCS, T_a – analysis time $T_a \in [T_{a\,\min}, T_{a\,\max}]$, $T_{a\,\min}$, $T_{a\,\max}$ – lower and upper analysis time limits, k – harmonic number with information harmonic frequencies k/T_f, k = 1, 2, …, K, ($K \approx 3.3\cdot10^7$, for T_f =1/60 Hz). To calculate (1), the signal implementation is analyzed

$$x(t) = \begin{cases} n(t), & \text{when the signal is not present} \\ n(t) + s(t, \mathbf{v}_0), & \text{when the signal is present} \end{cases} \quad (2)$$

where $n(t)$ – White Gaussian noise (WGN) with one-way spectral power density N_0 and the correlation function:

$$B(t_1, t_2) = \frac{N_0}{2}\delta(t_2 - t_1) \quad (3)$$

$s(t, \mathbf{v}_0)$ – SR signal, continuous in time t, $\mathbf{v}_0$ – vector of actual values of unknown parameters of the SR signal, $\delta(x)$ – Delta function.

A compatible algorithm for detecting SR and estimating the time interval at which they do not change consists in comparing the absolute maximum of (1) with the detection threshold h, at which a decision is made to detect SR from the monitor screen when

$$\sup_{T_{a\,\min} \le T_a \le T_{a\,\max}} L(T_a) > h \quad (4)$$

and the decision that there is no SR signal – when $\sup_{T_{a\,\min} \le T_a \le T_{a\,\max}} L(T_a) \le h$.

The scheme of the asymptotically Bayesian compatible algorithm for detecting the SR signal and estimating the duration used in modern STMEI is shown in Fig. 1. It consists of K – energy storage channels k – x harmonics in quadratures, K – adders in processing channels from $T_{a\,\min}$ to $T_{a\,\max}$, maximum selection devices (MSD) and a comparison scheme with the detection threshold h.

1. Problem statement

To find out the effectiveness of modern STMEI, it is necessary to answer a simple question: at what distances from the monitor screens with the LCS, SR will be detected. It will not go into the specifics of the propagation of radio waves in the near zone and will assume that the electric and magnetic components of the SE near the STMEI are formed the same as for a wave in the far zone, that is, at distances of tens and hundreds of meters from monitor screens.

Let us find the quality indicators of signal detection by the asymptotically optimal (asymptotically Bayesian) compatible algorithm for detecting SEMR&T and estimating the duration of a static image on the monitor screen (Fig. 1) at certain distances of the STMEI from the monitor screens with LCS: probability of false alarm

$$\alpha = 1 - F_0(h, T_a/0) \quad (5)$$

and the probability of signal miss:

$$\beta = F_0(h, T_a/1) \quad (6)$$

where $F_0(x, T_a/0)$, $F_0(x, T_a/1)$ – distribution of absolute maxima (DAM) of process (2) when in the implementation $x(t)$ of the signal there is no SEMR&T and when it is present, respectively. Approaches to the analysis of the signal processing algorithm (Fig. 1) using the apparatus of one-component Markov processes is highlighted in [1–3, 5, 7], and using two-component Markov processes – in [4, 6].

Since the solution to the problem of analyzing quality indicators of STMEI will depend on a specific "picture" on the monitor screen and a specific vector of parameters $\mathbf{v}_0 = (a_{1c0}, a_{1s0}, a_{2c0}, a_{2s0}, .., a_{Kc0}, a_{Ks0})$, for any SR signal with a frame tracking period T_f, which do not change on the monitor screen over time T_{a0}, the energy of the SR signal will amount to:

$$E = \int_0^{T_{a0}} s^2(t, \mathbf{v}_0)\,\mathrm{d}t \tag{7}$$

and will manifest itself, according to (2), through the deterministic component of the random process (1):

$$\det\left[L(T_{a0}, K)\right] = \frac{4}{N_0 T_{a0}} \sum_{k=1}^{K} a_{k0}^2 \times$$
$$\times \left[\left(\int_0^{T_{a0}} \cos\left(\frac{2\pi k t}{T_f} - \phi_k \right) \cos\left(\frac{2\pi k t}{T_f} \right) \mathrm{d}t \right)^2 + \right. \tag{8}$$
$$\left. + \left(\int_0^{T_{a0}} \cos\left(\frac{2\pi k t}{T_f} - \phi_k \right) \sin\left(\frac{2\pi k t}{T_f} \right) \mathrm{d}t \right)^2 \right]$$

a_{k0}^2 – squares of amplitudes k – x harmonics, k = 1, 2, …, K which are sent to the STMEI with initial phases ϕ_k, and are associated with quadratures $a_{kc0} = a_{k0}\cos(\phi_k)$, $a_{ks0} = a_{k0}\sin(\phi_k)$, estimates of which can be obtained by processing the input implementation of the signal in the STMEI:

$$\hat{a}_{kc0} = \frac{2}{T_{a0}} \int_0^{T_{a0}} x(t) \cos\left(\frac{2\pi k t}{T_f} \right) \mathrm{d}t \tag{9a}$$

$$\hat{a}_{ks0} = \frac{2}{T_{a0}} \int_0^{T_{a0}} x(t) \sin\left(\frac{2\pi k t}{T_f} \right) \mathrm{d}t \tag{9b}$$

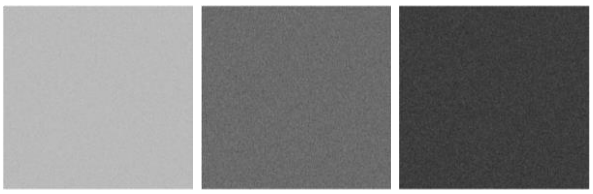
Estimates of (9) are more accurate than of (7) and T_{a0}/T_f is higher.

Let us simplify the problem by considering the SR signal for which $\forall k$ = 1, 2, …, K,

$$a_{kc0}^2 + a_{ks0}^2 = a_{k0}^2 = a^2 = \mathrm{const} \tag{10}$$

a^2 – SR signal strength at any k-th harmonic. Signal (10) is similar in properties to WGN, which has a constant spectral power density N_0. But unlike WGN, which will look like "snow" on a black-and-white monitor screen, this signal looks like a chaotic in terms of brightness, but static (not variable) "picture", the view of which is modeled in the MathCad environment, for the monitor of 800×768 pixels with T_f = 1/60 Hz. The generated signal was normalized to a maximum value of 255 and converted using the WRITEBMP("data.bmp") to the graphic image file shown in Fig. 2, for different values a^2.

The selected test signal obviously has maximum entropy. Therefore, if it is detected by the STMEI, any image with the same energy on the monitor screen, the leak from which is being investigated, it will be undoubtedly "intercepted" by the STMEI with indicators better than (6) for fixed values (5) (Neumann-Pearson criterion).



a) b) c)

Fig. 2. View of the test signal on the monitor screen: a) for $a^2 = a_1^2$, b) for $a^2 = a_2^2$, c) for $a^2 = a_3^2$, $a_1^2 < a_2^2 < a_3^2$

Taking into account (10), for $T_{a0}/T_f = T_{a0}^*$, expression (8) can be represented as follows:

$$\det\left[L(T_{a0}, K)\right] = q^2 + \frac{q^2}{16\pi^2 T_{a0}^{*2} K} \sum_{k=1}^{K} \frac{1}{k^2} \sin\left(4\pi k T_{a0}^*\right) -$$
$$- \frac{q^2}{\pi T_{a0}^* K} \sum_{k=1}^{K} \frac{1}{k} \sin(2\phi_k) \left[1 - \cos\left(4\pi k T_{a0}^*\right)\right] \tag{11}$$

which is simplified when $T_{a0} >> T_f$ ($T_{a0}^* \to \infty$) or $K \to \infty$:

$$\det\left[L(T_{a0}, K)\right] = \frac{a^2 K T_{a0}}{N_0} = q^2 K \tag{12}$$

q^2 – output signal-to-noise ratio T_a – the STMEI receiver channel (Fig. 1), if there is an SR signal from the monitor screen, to which the rating $\hat{T}_{a0}$ corresponds.

Objective: find quality indicators of signal detection using the algorithm (4): the probability of false alarm (5) and signal miss (6), depending on the energy indicators of the SR signal (7).

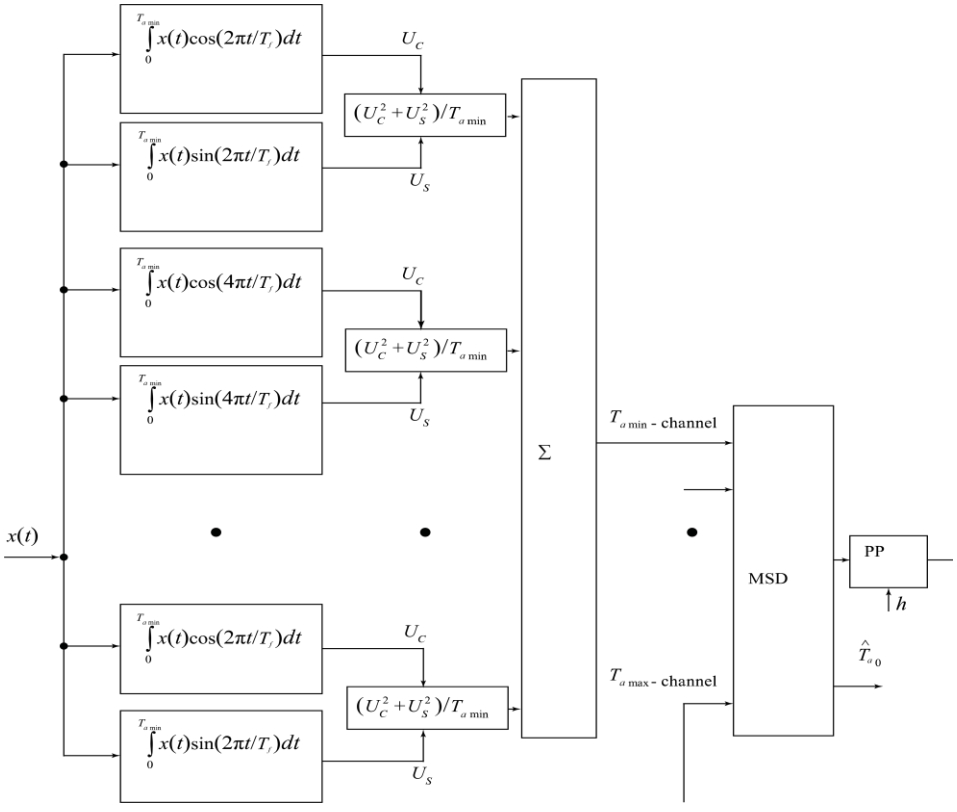


Fig. 1. Asymptotically Bayesian compatible algorithm for detecting IEMR&I and estimating the duration of a static image on a monitor screen

2. Solution of the problem of analysing the quality of side radiation detection

Whatever the signal structure $s(t,\mathbf{v}_0)$ is, its energy (7) is included in quadratures of amplitudes a_{kc} and a_{ks}, $k = 1, 2, \ldots, K$, k-th amplitude of the SR signal, and the task of analysing the quality of detection of SR signals by the STMEI is focused on the search for DAMs $F_0(x,T_a/0)$, $F_0(x,T_a/1)$ of random process (1):

$$L(T_a)=\frac{4}{N_0T_a}\xi(T_a) \tag{13}$$

in which [8]:

$$\xi(T_a)= =\sum_{k=1}^{K}\left[\left(\int_0^{T_a}x(t)\cos\left(\frac{2\pi kt}{T_f}\right)\mathrm{d}t\right)^2+\left(\int_0^{T_a}x(t)\sin\left(\frac{2\pi kt}{T_f}\right)\mathrm{d}t\right)^2\right]= \tag{14}$$

$$=\int_0^{T_a}x(t_1)\int_0^{T_a}x(t_2)\,f(t_2-t_1,K)\,\mathrm{d}t_1\mathrm{d}t_2$$

where

$$f(x,K)=\frac{\sin\left[(K+1)\pi x/T_f\right]\cos\left[K\pi x/T_f\right]}{\sin\left[\pi x/T_f\right]}-1 \tag{15}$$

When there is no SR signal at the input of the STMEI, in accordance with (2): $x(t)=n(t)$, and the random process (14) is distributed according to the law with mathematical expectation:

$$m_\xi(T_a)=\frac{N_0}{2}\int_0^{T_a}\int_0^{T_a}\delta(t_2-t_1)\,f(t_2-t_1,K)\,\mathrm{d}t_1\mathrm{d}t_2=\frac{N_0KT_a}{2} \tag{16}$$

The correlation function of process (14) can be represented as follows:

$$m_{\xi1\xi2}(T_{a1},T_{a2})=\int_0^{T_{a1}}\mathrm{d}t_1\int_0^{T_{a1}}\mathrm{d}t_2\int_0^{T_{a2}}\mathrm{d}t_3 \int_0^{T_{a2}}\langle n(t_1)n(t_2)n(t_3)n(t_4)\rangle\,f(t_2-t_1,K)\,f(t_4-t_3,K)\,\mathrm{d}t_4 \tag{17}$$

where $f(x,K)$ is set in (15), and

$$\langle n(t_1)n(t_2)n(t_3)n(t_4)\rangle= =\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1x_2x_3x_4w_4(x_1,x_2,x_3,x_4)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3\mathrm{d}x_4 \tag{18}$$

$w_4(x_1,x_2,x_3,x_4)$ – four-dimensional probability density of a Gaussian stationary process $n(t)$ at points in time t_1, t_2, t_3, t_4, which, taking into account (3), is represented by the expression (19):

$$w_4(x_1,x_2,x_3,x_4)=\frac{\sigma^2}{4\pi^2\sqrt{\left(\sigma^4-N_0^2\delta^2(t_2-t_1)/4\right)\left(\sigma^4-N_0^2\delta^2(t_3-t_2)/4\right)\left(\sigma^4-N_0^2\delta^2(t_4-t_3)/4\right)}}\times$$
$$\times\exp\left[-\left(\frac{N_0^2\delta^2(t_2-t_1)}{8\sigma^2\left(\sigma^4-N_0^2\delta^2(t_2-t_1)/4\right)}+\frac{1}{2\sigma^2}\right)x_1^2-\frac{\sigma^2}{2\left(\sigma^4-N_0^2\delta^2(t_4-t_3)/4\right)}x_4^2\right]\times$$
$$\times\exp\left[-\left(\frac{N_0^2\delta^2(t_3-t_2)}{8\sigma^2\left(\sigma^4-N_0^2\delta^2(t_3-t_2)/4\right)}+\frac{\sigma^2}{2\left(\sigma^4-N_0^2\delta^2(t_2-t_1)/4\right)}\right)x_2^2\right]\times$$
$$\times\exp\left[-\left(\frac{N_0^2\delta^2(t_4-t_3)}{8\sigma^2\left(\sigma^4-N_0^2\delta^2(t_4-t_3)/4\right)}+\frac{\sigma^2}{2\left(\sigma^4-N_0^2\delta^2(t_3-t_2)/4\right)}\right)x_3^2\right]\times \tag{19}$$
$$\times\exp\left[\frac{N_0\delta(t_2-t_1)x_1x_2}{2\left(\sigma^4-N_0^2\delta^2(t_2-t_1)/4\right)}+\frac{N_0\delta(t_3-t_2)x_2x_3}{2\left(\sigma^4-N_0^2\delta^2(t_3-t_2)/4\right)}+\frac{N_0\delta(t_4-t_3)x_3x_4}{2\left(\sigma^4-N_0^2\delta^2(t_4-t_3)/4\right)}\right]$$

A substitution of (19) into (18) and integrating it by $\mathrm{d}x_4$, $\mathrm{d}x_2$, $\mathrm{d}x_1$, $\mathrm{d}x_3$ leads to the expression:

$$\langle n(t_1)n(t_2)n(t_3)n(t_4)\rangle= =\frac{N_0^4\delta(t_4-t_3)\delta^2(t_3-t_2)\delta(t_2-t_1)}{4\sigma^4}+ +\frac{N_0^2\delta(t_4-t_3)\delta(t_2-t_1)}{2} \tag{20}$$

while a substitution of (20) to (17) and integrating it by $\mathrm{d}t_4$, $\mathrm{d}t_1$, $\mathrm{d}t_2$, $\mathrm{d}t_3$ – to the expression:

$$B_\xi(T_{a1},T_{a2})= =\langle\xi(T_{a1})\xi(T_{a2})\rangle=\frac{N_0^2K^2T_{a1}T_{a2}}{2} \tag{21}$$

In addition, taking into account (20), we can find the correlation function for $k \neq l$ of spectral components of SR:

$$B_{k,l}(T_a)= =\int_0^{T_a}\mathrm{d}t_1\int_0^{T_a}\mathrm{d}t_2\int_0^{T_a}\mathrm{d}t_3\int_0^{T_a}\langle n(t_1)n(t_2)n(t_3)n(t_4)\rangle\times \times\cos\left[\frac{2\pi k(t_2-t_1)}{T_f}\right]\cos\left[\frac{2\pi l(t_4-t_3)}{T_f}\right]\mathrm{d}t_4=\frac{N_0^2T_a^2}{2}, \quad \forall k,l=1,2,..,K$$

which amounts to (21) when the sum of the series in (14) consists of only one term $K = 1$, for $T_{a1}=T_{a2}=T_a$.

This means that all random spectral components of the SR signal are correlated with each other with the unit correlation coefficient, and since $K \neq \infty$, the distribution of the sum of the spectral components in (14) is not normalized, and the variance of the random variable $L(T_a)=Kn_0^2$ (n_0 – normally distributed random variable with zero mean and unit variance) – amounts to $2K^2$. This variance has the square of a centered Gaussian random variable with the variance K. Therefore, the distribution of the random variable (13), when there is no SR, is represented as follows [8, p. 221]:

$$F(h,T_a/0)=2F\left(\sqrt{h^*}\right)-1 \tag{22}$$

$h^*=\sqrt{h/K}$ – given detection threshold,
$F(x)$ – integral normal distribution function.

Therefore, if there is no SR signal at the input of the STMEI, the process (13) is distributed according to the law (22), while the correlation function of the Gaussian process forming (13) will be equal to:

$$B_L\left(T_{a1},T_{a2}\right)=\frac{KT_{a1}T_{a2}}{T_a^2} \tag{23}$$

and the desired probability of false alarm (5) can be calculated in accordance with [8, p. 170]:

$$\alpha=1-F\left(h,T_{a1}/0\right)\exp\left(-\int_{T_{a1}}^{T_{a2}}\frac{\pi\left(h,t/0\right)}{F\left(h,t/0\right)}\,\mathrm{d}t\right) \tag{24}$$

where $F\left(h,t/0\right)$ is set (22),

$$\begin{aligned}\pi\left(h,t/0\right)&=\int_0^\infty y w_2\left(h,y,t/0\right)\,\mathrm{d}y=\\&=\sqrt{\frac{2h}{\pi}}\frac{B_{LL'}}{\sigma_L^3}\exp\left[-\frac{h}{2\sigma_L^2}\right]\end{aligned} \tag{25}$$

$w_2\left(h,y,T_a/0\right)$ – two-dimensional probability density of the process distribution $L\left(T_a\right)$ and its derivative $L'\left(T_a\right)$ at identical points in time, having the parameters:

$$\begin{aligned}&\sigma_L^2=\sigma_L^2\left(T_a\right)=B_L\left(T_a,T_a\right)=K;\\&B_{LL'}=B_{LL'}\left(T_a\right)=\left.\frac{\partial B_\xi\left(T_{a1},T_{a2}\right)}{\partial T_{a2}}\right|_{T_{a1}=T_{a2}=T_a}=\frac{K}{T_a};\\&\sigma_{L'}^2=\sigma_{L'}^2\left(T_a\right)=\left.\frac{\partial^2 B_L\left(T_{a1},T_{a2}\right)}{\partial T_{a1}\partial T_{a2}}\right|_{T_{a1}=T_{a2}=T_a}=\frac{K}{T_a^2}\end{aligned} \tag{26}$$

where $B_L\left(T_{a1},T_{a2}\right)$ – is defined (23), σ_L^2 – variance of the centered Gaussian process, $\sigma_{L'}^2$ – variance of the derivative of the centered Gaussian process; $B_{LL'}$ – mutual correlation function of the centered Gaussian process and its derivative at identical points in time.

Integrating the two-dimensional probability density in expression (25), taking into account (26), allows us to present the dependence of the probability of false alarm (24) on the, represented as follows:

$$\alpha=1-\left(2F\left(\sqrt{h^*}\right)-1\right)\cdot\left(\frac{T_{a0\max}}{T_{a0\min}}\right)^{-\sqrt{\frac{2h^*}{\pi}}\frac{\exp\left(-\frac{h^*}{2}\right)}{\left(2F\left(\sqrt{h^*}\right)-1\right)}} \tag{27}$$

Dependence of the probability of false alarm (27) on the given detection threshold of the STMEI receiver, for $T_{a0\max}/T_{a0\min}$ = 1; 2; 10, is shown in Fig. 3. As we can see, a fixed false alarm is achieved at higher h^*, for longer intervals $T_{a0\max}/T_{a0\min}$ in the uncertainty of values T_{a0}, $T_{a0\min}\le T_{a0}\le T_{a0\max}$.

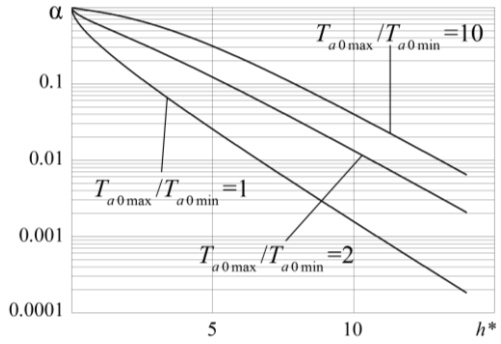


Fig. 3. Dependence of the probability of false alarm on the given detection threshold of the receiver of a specialized technical means of enemy intelligence

According to the Neumann-Pearson criterion, fixing the false alarm level (27) makes it possible to find the given detection threshold h^*, and calculate the probability of signal miss (6) when there is an SR signal at the input of the STMEI receiver. Random variable $L\left(T_a\right)$, taking into account (12), is the square of a Gaussian random variable with variance $\sigma_L^2=K$ and mathematical expectations $q\sqrt{K}$, which is distributed according to the law:

$$F\left(h,T_a/1\right)=F\left(\sqrt{h^*}-q\right)+F\left(\sqrt{h^*}+q\right)-1 \tag{28}$$

If after the time of the SR signal T_{a0} the rest of the time $T_{a0}\le T_a<T_{a0\max}$ there was no SR signal, a random variable $L\left(T_a\right)$ for $T_{a0\min}\le T_a<T_{a0}$ can be considered a non-stationary Gaussian process with mathematical expectation $q\sqrt{K}$, for $T_{a0}\le T_a<T_{a0\max}$ – with zero mathematical expectation, and the probability of signal miss (6) is calculated according to the expression

$$\beta=F\left(h,T_{a1}/1\right)\exp\left(-\int_{T_{a1}}^{T_{a2}}\frac{\pi\left(h,t/1\right)}{F\left(h,t/1\right)}\,\mathrm{d}t\right) \tag{29}$$

where $F\left(h,t/1\right)$ is set (28),

$$\begin{aligned}&\pi\left(h,t/1\right)=\int_0^\infty y w_2\left(h,y,t/1\right)\,\mathrm{d}y=\\&=\frac{B_{LL'}\left(\sqrt{h}-q\sqrt{K}\right)\cdot\mathbf{1}\left(\sqrt{h}-q\sqrt{K}\right)}{2\sqrt{2\pi h}\sigma_L^2}\times\exp\left[-\frac{\left(\sqrt{h}-q\sqrt{K}\right)^2}{2\sigma_L^2}\right]+\\&+\frac{B_{LL'}\left(\sqrt{h}+q\sqrt{K}\right)}{\sqrt{2\pi}\sigma_L^3}\exp\left[-\frac{\left(\sqrt{h}+q\sqrt{K}\right)^2}{2\sigma_L^2}\right]\end{aligned} \tag{30}$$

$\mathbf{1}(x)$ – single function.

At the same time, for any T_{a0} process parameters $L\left(T_a\right)$ as before, will be calculated taking into account (26), and the desired probability of signal miss (29), taking into account (30), will be equal to:

$$\begin{aligned}\beta=&\left[F\left(\sqrt{h^*}-q\right)+F\left(\sqrt{h^*}+q\right)-1\right]\times\\&\times\left(\frac{T_{a0}}{T_{a0\min}}\right)^{-\frac{\left(\sqrt{h^*}-q\right)\exp\left(-\frac{\left(\sqrt{h^*}-q\right)^2}{2}\right)\cdot\mathbf{1}\left(\sqrt{h^*}-q\right)+\left(\sqrt{h^*}+q\right)\exp\left(-\frac{\left(\sqrt{h^*}+q\right)^2}{2}\right)}{\sqrt{2\pi}\left[F\left(\sqrt{h^*}-q\right)+F\left(\sqrt{h^*}+q\right)-1\right]}}\times\\&\times\left(\frac{T_{a0\max}}{T_{a0}}\right)^{-\sqrt{\frac{2h^*}{\pi}}\frac{\exp\left(-\frac{h^*}{2}\right)}{\left(2F\left(\sqrt{h^*}\right)-1\right)}}\end{aligned} \tag{31}$$

in which h^* is found according to (27), for a fixed α, while q – defined by (12).

In Fig. 4 the dependences of the probabilities of correct signal detection $1-\beta$ from the parameter q, for α = 0.01, $T_{a0\max}/T_{a0\min}$ = 1; 2; 10 and $T_{a0}/T_{a0\min}$ = 1. As expected, with increasing uncertainty $T_{a0\max}/T_{a0\min}$ the detection curves are dislocated to the right.

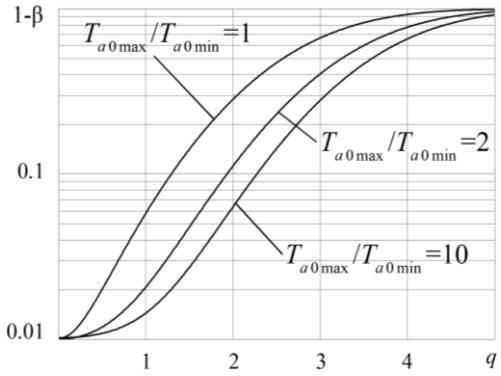


Fig. 4. Probability of correct detection of information leakage signals from the monitor screen, for $T_{a0}/T_{a0\min}=1$

3. Modeling in MathCAD and Excel environments

Let us investigate the correctness of the calculation using expressions (27) and (31). For this purpose, we will simulate the process of exceeding the detection threshold h in the tool (Fig. 1), when the input of the STMEI, in accordance with (2), receives the implementation of WGN or the sum of SR with the characteristics of (10) and WGN. At the same time, WGN is given close to δ – correlated normally distributed array n_i, $i = 1, 2, \ldots, N$, with zero mean and variance σ^2, N – array volume. The array with these characteristics is generated in the Mathcad software environment, and its further processing in the Excel software environment allows to conveniently controlling all the statistical characteristics of intermediate digital arrays and making sure that the simulation results obtained are reliable.

For the statistical validity of the estimates generated when creating the model, it is necessary to operate n independent tests on arrays with N/n elements, while to estimate, for example, the probability of a false alarm $\alpha = 0.01$, it is necessary to have more than $n = 100$ tests. If $N = 15000$ arrays that can be used for statistical processing will have volumes less than 150 samples. That is, during the simulation of the process at time intervals $t \in [0, T_a]$ integrals in expression (14) are replaced by sums:

$$\Delta t \sum_{i=1}^{\mathrm{ceil}(T_a/\Delta t)} n_i \cos\left(\frac{2\pi k \Delta t i}{T_f}\right),\ \Delta t \sum_{i=1}^{\mathrm{ceil}(T_a/\Delta t)} n_i \sin\left(\frac{2\pi k \Delta t i}{T_f}\right)$$

where $\mathrm{ceil}(x)$ – the whole part of x, Δt – time interval for sampling the process.

In Fig. 5 solid lines represent the dependences of probabilities of a false alarm on the detection threshold, for small $K = 20; 30; 50$ and $T_{a0\max}/T_{a0\min} = 1$, calculated according to (27). The triangles in the figure illustrate the results of modelling the process of false signal detection for If $N = 15000$ and $n = 1000$ ($T_a = 15\Delta t$).

As you can see, the results of calculations are quite accurately confirmed by practice, especially for small K. Increasing discrepancies between theory and practice, with increasing K, is associated with an increase in variance $\sigma_L^2 = K$ and the need to increase the number of independent software tests to create a more adequate model of false signal detection.

In the case when SR is observed by STMEI, for $T_{a0\max} = T_{a0\min} = T_{a0}$ simulation of the correct detection process using a noise array with $N = 15000$ samples is possible only for small ratios T_{a0}/T_f, since $T_{a0} = 15\Delta t$. In this case, the interval Δt, in fact, determines the time when one pixel of information is emitted from the monitor screen. Therefore, even without taking into account the time for transmitting string synchronization pulses [10] for a monitor with a scan of 800×768 pixels $T_f > 614400\Delta t$, while $T_{a0}/T_f < 2.4\cdot10^{-5}$.

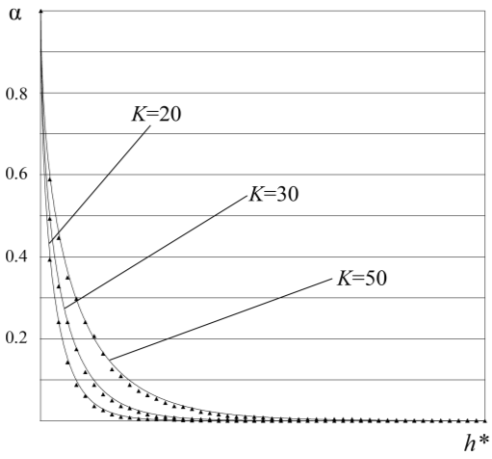


Fig. 5. Calculated and modeled dependences of probabilities of a false alarm on the reduced detection threshold of the STMEI receiver

In Fig. 6–8 solid lines illustrate the dependences calculated using (31) of the probabilities of correct detection of the SR signal $1-\beta$ from the given threshold h^*, for small $K = 20; 30; 50$, $T_{a0\max}/T_{a0\min} = 1$, and $q = 1$ – Fig. 6, $q = 2$ – Fig. 7, $q = 3$ – Fig. 8. For comparison, triangles indicate the results of modeling the process of passing the SR signal. As you can see, the results of calculations according to (28) are almost perfectly confirmed by modeling the process in the Mathcad and Excel software environment, at least for $T_{a0\max} = T_{a0\min} = T_{a0}$.

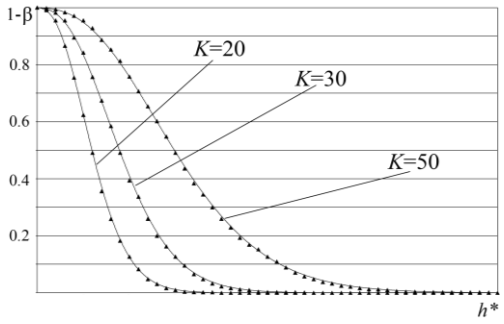


Fig. 6. Calculated and modeled the dependences of the probabilities of correct detection of SR on the given detection threshold of the STMEI receiver, for q = 1

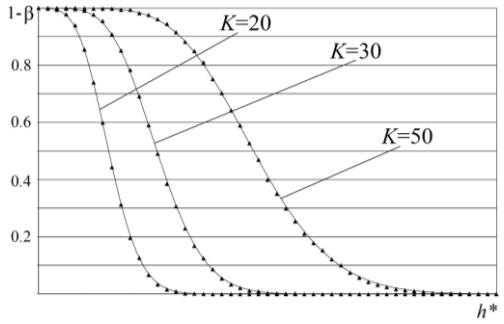


Fig. 7. Calculated and modeled the dependences of the probabilities of correct detection of SR on the given detection threshold of the STMEI receiver, for q = 2

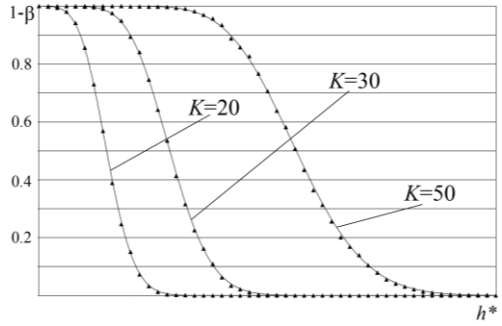


Fig. 8. Calculated and modeled the dependences of the probabilities of correct detection of SR on the given detection threshold of the STMEI receiver, for q = 3

In order to use the same software to create an adequate model for studying the correct detection of the SR signal, for $T_{a0\max} / T_f = T^*_{a0\max}$, it is necessary to increase the noise signal array to $N = 6.25 \cdot 10^8 T^*_{a0\max}$, which is unrealistic neither in the Mathcad environment nor in the Excel environment. In addition, for processing $K \approx 3.3 \cdot 10^7$, for T_f = 1/60 Hz (and not as it was investigated for $K = 50$) it is necessary to multiply $6.6 \cdot 10^5$ times the size of the substitution tables in the Excel software environment, which is also impossible to do. In addition, the study of expression (25) for $T_{a0\max} \neq T_{a0\min}$ will require approximately $T_{a0\max} / T_{a0\min}$ times larger amounts of calculations, which is possible only with the use of professional programs for processing information arrays in C++ or C sharp software environments.

Conclusions

1. The expressions found in this paper for calculating the probability of false alarm and signal miss of information leakage from monitor screens for a specialised enemy reconnaissance vehicle are confirmed by modelling the processes of false and correct detection of side radiation signals in the Mathcad and Excel software environments, at least when there is no a priori uncertainty about the time of change of the static image on the monitor screen.
2. Finally, the reliability of the obtained expressions for calculating the quality indicators of detecting signals of side radiation from monitor screens should be confirmed by additional modelling of the process of detecting information leakage by processing significant data sets in the C++ or C sharp software environments.

References

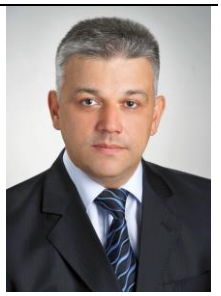
[1] Adler R. J.: Random field and geometry. Springer, New York 2007.
[2] Adler R. J.: Random field and their geometry. Springer, New York 2003.
[3] Behl D., Handa S., Arora A.: A bug Mining tool to identify and analyze security bugs using Naive Bayes and TF-IDF. International Conference on Reliability Optimization and Information Technology – ICROIT, IEEE 2014, 294–299.
[4] Chmielowski L., Kucharzak M.: Impact of Software Bug Report Preprocessing and Vectorization on Bug Assignment Accuracy. Progress in Image Processing, Pattern Recognition and Communication Systems. Springer, Cham 2022.
[5] Gnedenko B. V.: Sur la distribution limite du terme maximum d'une serie aleatore. Annals of Mathematics. 44(3), 1943, 423–453.
[6] Leadbetter V. R.: Extreme value theory for continuous parameter stationary processes. Technical rept. ADA083802, 1980-03-01.
[7] Leadbetter V. R., Lindgren G., Rootzen H.: Extreme and Related Properties of Random Sequences and Processes. Springer, New York 1983.
[8] Marple S. L.: Digital spectral analysis: with application. Prentice Hall, New Jersey 1987.
[9] Yevgrafov D. V., Yaremchuk Y. Y.: Asymptotically optimal algorithm for processing side radiation signals from monitor screens on liquid crystal structures. Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska – IAPGOS 13(4), 2023, 99–102.
[10] Yevgrafov D. V., Yaremchuk Y. Y.: Calculated spectra of information leakage signals from monitor screens with liquid crystal structures. Registration, storage and processing of data 23(2), Kyiv 2021, 3–12.

Ph.D. Dmytro V. Yevgrafov
e-mail: yevg.dmytro@gmail.com

The Ph.D. (Engineering), and the senior researcher. He is the author of more than 50 scientific academic papers, including 2 monographs, 45 articles in scientific specialized publications, 4 training manuals, materials and abstracts of reports at academic conferences.
The author's scientific interests are in statistical radio engineering, radio electronic intelligence and physical protection of information.

https://orcid.org/0000-0001-9651-1558

D.Sc. Eng. Yurii Ye. Yaremchuk
e-mail: yurevyar@vntu.edu.ua



The Doctor of Science in Engineering, and the professor. He is the author of more than 300 publications, including 2 monographs, 140 articles in scientific specialized publications, 20 textbooks and training manuals, 20 utility model patents and 25 certificates of copyright registration for work.
The author's scientific interests are in cryptographic and steganographic protection of information, technical protection of information and security of information systems.

https://orcid.org/0000-0002-6303-7703

http://doi.org/10.35784/iapgos.5445 *received: 09.10.2023 | revised: 26.02.2024 | accepted: 09.03.2024 | available online: 31.03.2024*

THE IMPACT OF LIGHTNING STRIKE ON HYBRID HIGH VOLTAGE OVERHEAD TRANSMISSION LINE – INSULATED GAS LINE

Samira Boumous, Zouhir Boumous, Yacine Djeghader
Mohamed Cherif Messaidia University, Electrical Engineering Department, Laboratory of Electrical Engineering and Renewable Energy, Souk Ahras, Algeria

Abstract. *The electrical network is the set of elements where loads are connected to the generation plants by transmission lines. They can be either overhead or underground cables. A new technology has been introduced to replace these transmission lines with underground cables gas insulated line "GIL". The latest has many advantages over underground cables and overhead transmission lines, such as low transmission losses, less capacitive load, reliability, personal safety, same operation as overhead lines and negligible electrical aging. GIL can handle much more power than overhead lines due to its large conductive area. GIL is the best for high voltage. In this paper, the simulation of lightning strike effects on a 400 kV hybrid transmission line located in the Wilaya of Setif in northern Algeria is presented in the absence and presence of line arresters and GIL arresters. The results of this paper can provide a rich and valuable theoretical reference for GIL simulation modeling and evaluation of lightning strike impact on hybrid overhead – GIL lines.*

Keywords: overhead transmission line, gas insulated line, lightning strike, surge arrester

WPŁYW UDERZENIA PIORUNA NA HYBRYDOWĄ LINIĘ WYSOKIEGO NAPIĘCIA NAPOWIETRZNA LINIA PRZESYŁOWA – IZOLOWANA LINIA GAZOWA

Streszczenie. *Sieć elektryczna to zbiór elementów, w których obciążenia są połączone z elektrowniami za pomocą linii przesyłowych. Mogą to być linie napowietrzne lub podziemne. Nowa technologia została wprowadzona w celu zastąpienia tych linii przesyłowych podziemnymi kablami w izolacji gazowej "GIL". Najnowsza technologia ma wiele zalet w porównaniu z kablami podziemnymi i napowietrznymi liniami przesyłowymi, takich jak niskie straty przesyłowe, mniejsze obciążenie pojemnościowe, niezawodność, bezpieczeństwo obsługi, takie samo działanie jak w przypadku linii napowietrznych i znikome starzenie elektryczne. GIL może obsługiwać znacznie większą moc niż linie napowietrzne ze względu na dużą powierzchnię przewodzącą. GIL jest najlepszy dla wysokich napięć. W niniejszym artykule przedstawiono symulację skutków uderzenia pioruna w hybrydową linię przesyłową 400 kV zlokalizowaną w Wilaya, Setif w północnej Algierii w przypadku braku i obecności ograniczników liniowych i ograniczników GIL. Wyniki tego artykułu mogą stanowić bogate i cenne teoretyczne odniesienie do modelowania symulacji GIL i oceny wpływu uderzenia pioruna na hybrydowe linie napowietrzne – GIL.*

Słowa kluczowe: napowietrzna linia przesyłowa, linia w izolacji gazowej, uderzenie pioruna, ogranicznik przepięć

Introduction

Compared to traditional bare wire and cable, the cost of GIL is relatively high. However, the cost of geothermal cables is decreasing as the technology of production processes improves. In the context of today's ubiquitous smart grid and the global energy Internet of energy, the requirements for transmission and transmission of various links are becoming increasingly high, and the application of GIL is becoming increasingly broad. The global transmission and distribution giants have invested in GIL research [18]. Due to its special structure and excellent performance, GIL is used in some power plants or large hub power plants, where the installation method is special and the safety requirements are high. In the future, with the development of technology, it will be applied to long distance transmission. However, due to its cost and other reasons, GIL has not been widely promoted. Currently, the GIL lines used in the world have reached thousands of kilometers [11]. Several recent researches have invested in the study and simulation of different combinations of hybrid lines, in [3] the authors have studied the scheme design of integrated grounding system for a 550 kV gas insulated transmission line, researchers in [12] have presented a short circuit current calculation of EHV overhead Line-GIL hybrid line, in [25] authors proposed a novel light weight 2D multi-slice electromagnetic field-circuit coupled method for computing transient electromagnetic force. In reference [4], researchers studied The induced voltage and current for hybrid transmission system.

In high voltage systems, the protection and reliability of these elements plays an important role in the analysis and simulation of these systems because of the danger that can be caused by the shock wave of a direct lightning strike on the line or a short circuit, an overload, etc.

Researchers in [22], have studied the holistic modelling of 500 kV GIS step-up substation equipment with a double circuit and long GIS busbar in a coastal power plant is established using Electromagnetic Transient Program. In [6], the FEA method is used for insulation design of multi-conductor transmission lines. In the design of -800 kV transmission line, coaxial and non-coaxial structures are mainly considered. The purpose of our article is to investigate the influence of lightning arresters on the evolution of voltages and currents on a three-phase hybrid line impacted by a direct lightning strike on one phase of a tower located in the middle of the line.

1. Components of the system to be studied

1.1. Surge arrester

A surge arrester is a protective device used to protect electrical equipment from damage due to lightning surges or other transients in electrical power systems. In the case of lightning surges, a surge arrester provides a low-impedance path for the surge current to flow to ground, thus limiting the voltage applied to the protected equipment and preventing damage [14].

Figure 1, illustrate some examples of line arrester hung on tower arms.

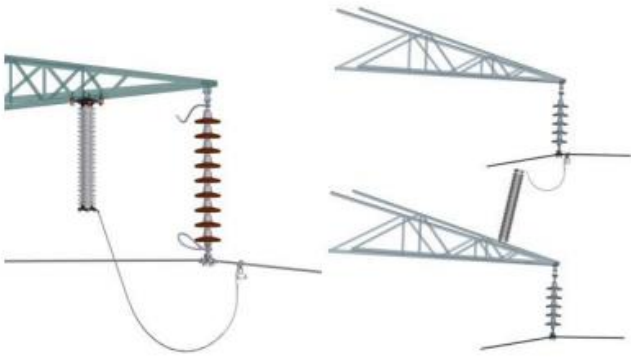


Fig. 1. OHL surge arrester [8]

Surge arresters typically consist of a nonlinear resistor (such as a metal oxide varistor) connected in parallel with the protected equipment, with a low impedance path to ground. When a lightning surge or other transient occurs, the voltage across the surge arrester exceeds its breakover voltage, causing the nonlinear resistor to rapidly conduct the surge current

to ground. There are different types of surge arresters available for different applications, including overhead line arresters, distribution arresters, and low-voltage arresters [8, 9].

1.2. Lightning strike

A lightning strike is a sudden and powerful discharge of electricity from the atmosphere to the ground or another object, caused by the buildup of electrical charge in thunderstorms. Lightning strikes can cause significant damage to electrical equipment and infrastructure, and can also be a significant hazard to people and animals. Lightning is a complex phenomenon that is still not fully understood, but it is thought to be caused by the buildup of electrical charge in thunderstorms. When the electric field within a thunderstorm becomes strong enough, it can ionize the air, creating a path for the electrical discharge to flow. The resulting lightning strike can transfer tens of thousands of amperes of current to the ground in a fraction of a second, creating a high-voltage, high-frequency electrical surge [17, 24].

Lightning strikes can cause a variety of effects, depending on the magnitude of the strike and the proximity to the target. For example, a direct lightning strike can cause fire, explosion, or structural damage, while a nearby lightning strike can cause electromagnetic interference (EMI) or electrical surges that can damage sensitive electronic equipment [17].

Due to the dangerous effects of lightning strikes on property and people, a recent study is published which gives the detection of lightning has arrived thanks to the space-based operational lightning imagers [1].

1.2.1. Overhead line

An overhead line, also known as an overhead power line or overhead transmission line, is a system of electrical conductors used to transmit high-voltage electricity from power plants to distribution substations or directly to large industrial customers. The conductors are typically supported by tall steel or concrete poles, or towers, and are usually made of aluminium or copper. Overhead lines are widely used in power transmission and distribution because they are relatively inexpensive and have lower electrical losses compared to underground cables. However, they can be vulnerable to weather-related disruptions such as high winds, lightning, and ice accumulation [11–13, 15–20].

1.2.2. Gas insulated line GIL

Gas Insulated Line (GIL) is a type of high-voltage power transmission system that uses a gas-insulated conductor to transmit electricity over long distances. The GIL consists of a metal housing or tube that contains one or more conductors, which are insulated by a compressed gas, typically sulphur hexafluoride (SF6). The gas provides excellent electrical insulation and allows the GIL to operate at high voltages without the risk of electric breakdown. GIL is typically used for underground and underwater transmission lines, where it can offer higher power density, greater reliability, and lower environmental impact compared to traditional overhead power lines or underground cables. However, GIL systems can be expensive to install and maintain, and there are concerns about the environmental impact of SF6 gas, which is a potent greenhouse gas [4, 9, 20–23]. The basic structure of 500 kV GIL is illustrated in figure 2, it is mainly composed of metal enclosure, inner conductor, post insulator, basin insulator and particle trap [16].

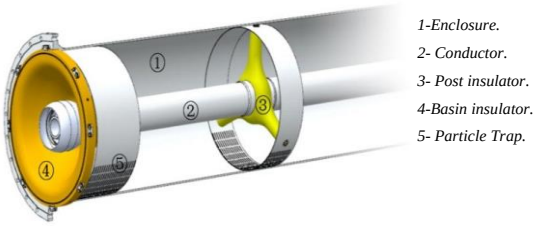


Fig. 2. Basic structure of gas insulated transmission line (GIL) [16]

2. Simulation models

EMTP-ATP software is used to modelling and simulation, Electromagnetic Transients Program is a software tool used by power systems engineers to analyse electromagnetic transients and associated insulation issues, it is developed by Dr. Scott Meyer and co. in USA. The used models of the OHL line and these components have been developed in [5, 21]. The surge arresters were represented by the Pinceti and Giannettoni frequency-dependent model [1, 2, 5, 7, 10, 14] as shown in Fig. 3.

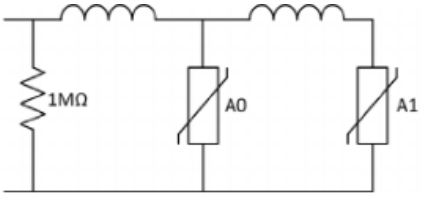


Fig. 3. Frequency-dependent surge arrester model, with: $R = 1\ M\Omega$, $L_0 = 4.5\ \mu H$, $L_1 = 13.6\ \mu H$

3. Power system description

The OAT-El Hassi line (400 kV) located in the north east of Algeria in the Wilaya of Setif. The line is divided into a number of identical sections as shown in figure 4 for the 400 kV line. Authors chose a length of portion equal to 2.4 km divided into 10 spans (Fig. 4). GIL is located in the middle of the OHL line, it is of length of 2 km, two identical surge arresters have been placed upstream and downstream of GIL in order to compare the simulation results obtained in presence and absence of the GIL arresters, a direct lightning strike is introduced on the ground wire of the first tower (Tower 1) of power system presented in figure 4.

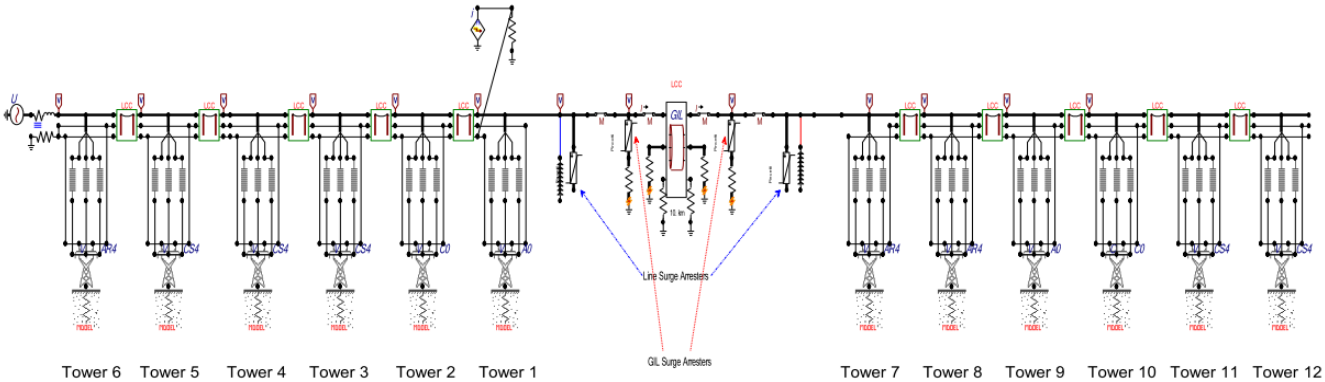


Fig. 4. Study power system

In this study, a positive polarity of lightning current is used to perform shielding failure pattern analysis on the modelled circuit. Simulations are thus conducted with a lightning-strike, the characteristics given on the table 1. Authors in the reference [2] have detailed the calculation of the GIS simulation parameters in the EMTP-ATP software such as the grounding grid resistance, grounding strip inductance and grounding strip resistance.

Table 1. Characteristics of lightning strike

Front time (µs)	2	2	2	2
Tail time (µs)	30	70	180	260
Lightning current amplitude, kA	120	120	120	120

4. Simulation results

If the grounding wire is properly installed and connected, it can provide a path for the lightning to discharge safely into the ground. In this case, the lightning strike may cause a surge in the electrical system, but the surge protectors and other safety devices in the electrical system should be able to handle the surge and protect the appliances and other devices connected to the system.

In this context, our research is based on the role of lightning arrestors in hybrid high voltage transmission lines; The impact of a series of lightning strikes of amplitude 120 kA with different tail times is simulated. In figure 5, the magnitude of lightning current for a time simulation 50 µs is plotted.

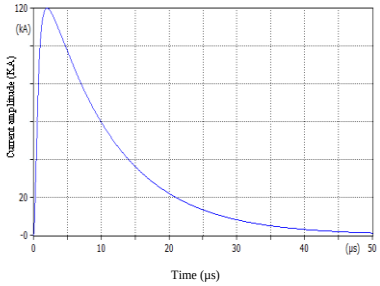


Fig. 5. Lightning strike current amplitude

The induced voltage in the A phase of the transmission line is simulated in five different cases configuration with different combinations of line and GILs arresters.

Case 1: OHL with surge arrester line, without GIL.

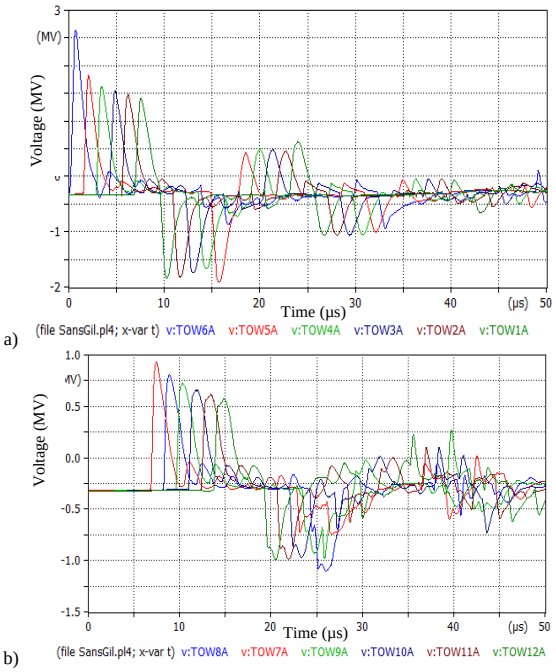


Fig. 6. Induced voltage across the phase conductor "A" (case 1): a) voltage evolution in tower 6, 5, 4, 3, 2, and 1, b) voltage evolution in tower 7, 8, 9, 10, 11, and 12

Case 2: Hybrid OHL with surge arrester line, without GIL surge arrester.

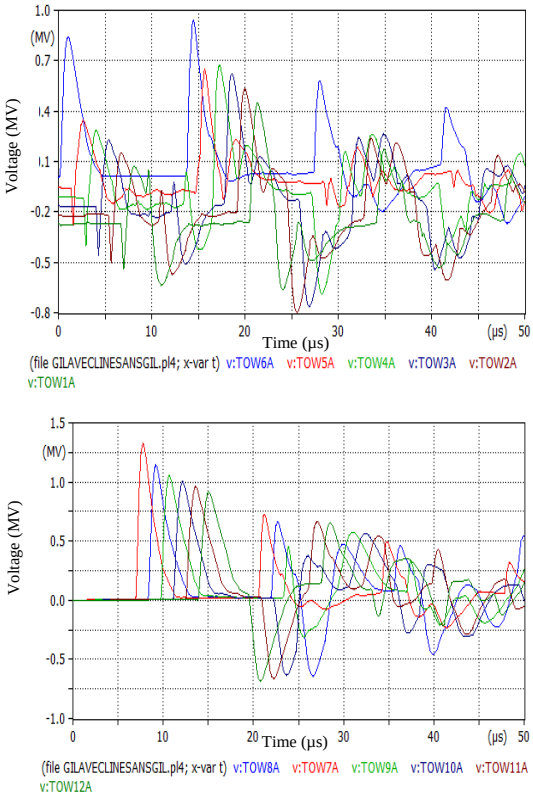


Fig. 7. Induced voltage across the phase conductor "A" (case 2): a) voltage evolution in tower 6, 5, 4, 3, 2, and 1, b) voltage evolution in tower 7, 8, 9, 10, 11, and 12

Case 3: Hybrid OHL without surge arrester line, without GIL surge arrester.

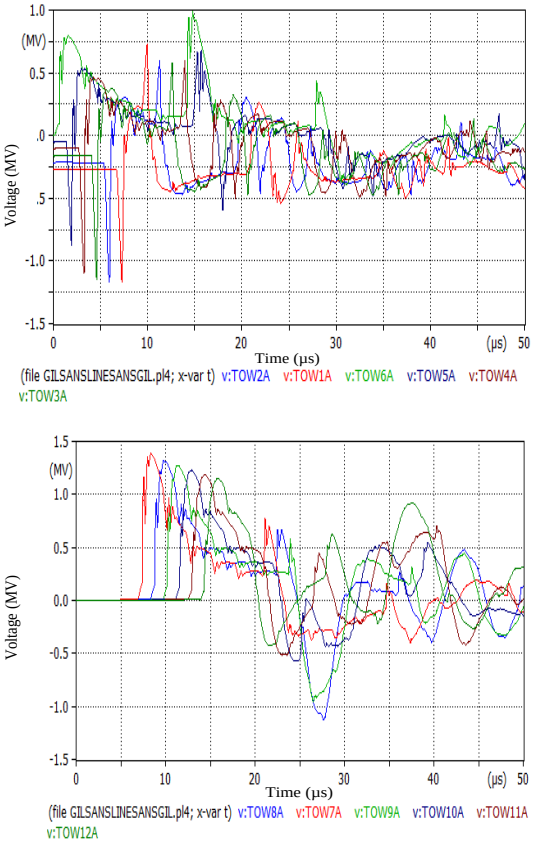


Fig. 8. Induced voltage across the phase conductor "A" (case 3): a) voltage evolution in tower 6, 5, 4, 3, 2, and 1, b) voltage evolution in tower 7, 8, 9, 10, 11, and 12

Case 4: Hybrid OHL with surge arrester line, with surge GIL surge arrester.

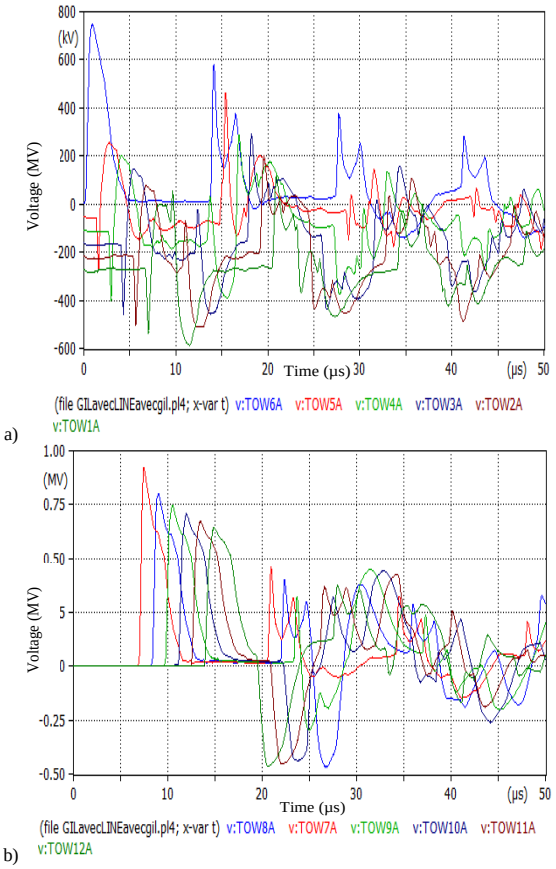


Fig. 9. Induced voltage across the phase conductor "A" (case 4): a) voltage evolution in tower 6, 5, 4, 3, 2, and 1, b) voltage evolution in tower 7, 8, 9, 10, 11, and 12

Case 5: Hybrid OHL without surge arrester line, with GIL surge arrester.

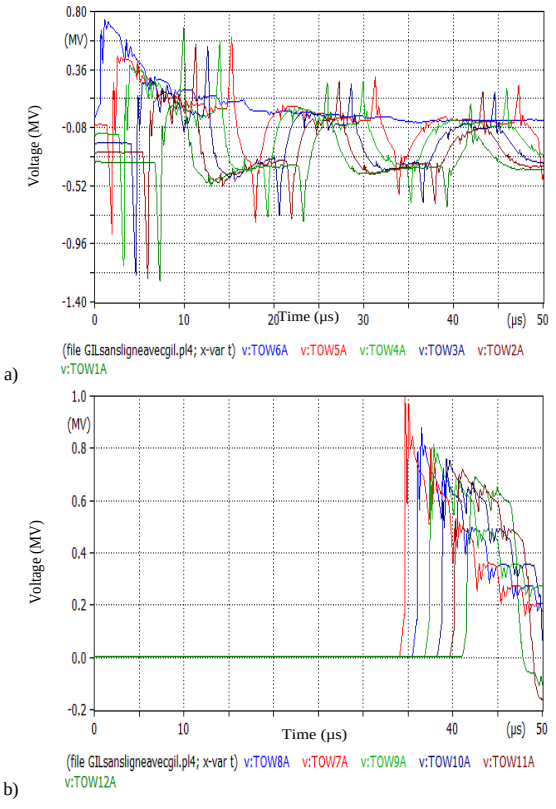


Fig. 10. Induced voltage across the phase conductor "A" (case 5): a) voltage evolution in tower 6, 5, 4, 3, 2, and 1, b) voltage evolution in tower 7, 8, 9, 10, 11, and 12

5. Discussion

As illustrated in figures (6, 7, 8, 9, 10), the induced voltage across the phase conductor is plotted on a time range from 0 to 50 µs. by changing the configuration and combination of protection systems with and without arresters, the induced voltage patterns change in waveform and magnitude.

We have chosen to present separately the voltages before and after the GIL section, and this in the presence and absence of the overhead transmission line arresters and GIL. the maximum value of the induced voltage is around 2.6 MV, it is reached in the first study case (simple OHL with surge arrester installed) with an almost sinusoidal waveform as shown in figure 6. While it is around 730 kV in case No 4 (OHL and GIL with surge arresters installed) (Fig. 9).

The results obtained in the 3rd case (OHL and GIL without arresters) are visibly disturbed and distorted due to the propagation of the lightning wave on the line without any discharge by arresters (Fig. 8).

The analysis of the results obtained in the cases 4 and 5 shows that the installation of the surge arrester in the gas insulated line (GIL) presented on the figures 9 and 10 has strongly influenced the magnitude and the waveform of the voltages induced on the left and on the right of the GIL, which is in good agreement with the results obtained by the authors of the reference [12].

6. Conclusion

In this paper, the effect of lightning strike on hybrid overhead transmission line is studied.

The use of arresters in Gas Insulated Lines (GIL) is important for several reasons: Overvoltage protection: Arresters protect the GIL from overvoltage surges that can occur due to lightning strikes or switching operations. These surges can damage the insulation of the GIL, leading to costly repairs and downtime. Arresters divert the surge current away from the GIL, preventing damage to the insulation.

Safety: Arresters protect personnel and equipment from the effects of high voltage surges that can occur in the GIL. Without arresters, a surge could cause an electrical fault or explosion, posing a danger to personnel and equipment.

Reliability: Arresters improve the reliability of the GIL system by preventing damage to the insulation and reducing downtime. By protecting against overvoltage surges, arresters help to ensure that the GIL operates as intended, with fewer interruptions or faults.

Cost-effectiveness: The use of arresters in GIL can be a cost-effective solution to protect against overvoltage surges. Without arresters, the cost of repairing or replacing damaged GIL insulation can be high. Additionally, arresters are relatively low cost compared to the potential costs of downtime, equipment damage, or personal injury.

Overall, the use of arresters in GIL is an important part of protecting the GIL system and ensuring its safe and reliable operation.

Neuro-Fuzzy Inference System (ANFIS) controller for Maximum Power Point The MPPT controller plays a crucial role in extracting the maximum available power from renewable energy sources, such as solar panels or wind turbines, in a microgrid system. GSA is used to optimize the parameters of the MPPT controller. The GSA algorithm can explore the parameter space of the controller, such as the gains, setpoints, or control rules, to find the optimal values that maximize the power extraction from the renewable energy sources. GSA's ability to balance exploration and exploitation can be leveraged to fine-tune the controller's parameters and improve its performance.

The ANFIS controller can be employed as the MPPT controller in the microgrid system. ANFIS combines the advantages of fuzzy logic and neural networks to create an intelligent control system capable of capturing and utilizing

expert knowledge in the form of fuzzy rules. ANFIS can model the nonlinear characteristics and dynamics of the renewable energy sources and adjust its parameters based on the feedback signals to track the maximum power point.

The efficiency, stability and reliability of a photovoltaic energy are considered major factors for establishing this energy resource on the market. In this research, common maximum power point tracking techniques, using Gravitational Search Algorithm and ANFIS Controller. Adaptive neural fuzzy inference system and artificial neural network have been proposed for a grid-connected PV system to maximise the output power of a PV array. The aim has also been improving the stability and reliability of a PV power conversion, especially in the context of a rapid change in atmospheric conditions.

References

[1] Badjor M., Semenova E., Kulikov A.: Measures to protect overhead lines from ac contact network. Energy Systems 7, 2022, 38–45 [https://doi.org/10.34031/es.2022.1.004].

[2] Chandrakar K., Gorayan R.: Analysis of transient enclosure voltages in GIS (EMTP simulation studies). International Journal of Research in Engineering and Technology 2, 2013, 120–125 [https://doi.org/10.15623/ijret.2013.0202006].

[3] Chen G. et al.: Environment-friendly insulating gases for HVDC gas-insulated transmission lines. CSEE Journal of Power and Energy Systems 7(3), 2021, 510–529.

[4] Cheng S., Zhao Y., Xie K., Hu B.: A novel multi-slice electromagnetic field-circuit coupling method for transient computation of long-distance gas-insulated transmission lines. High Voltage, 2024, 1–13.

[5] Colqui J. S. L. et al.: Implementation of Modal Domain Transmission Line Models in the ATP Software. IEEE Access 10, 2022, 15924–15934 [https://doi.org/10.1109/ACCESS.2022.3146880].

[6] Gao K. et al.: Progress in Environment-friendly Gas-insulated Transmission Line (GIL). High Voltage Engineering 44, 2018, 3105–3113 [https://doi.org/10.13336/j.1003-6520.hve.20180925001].

[7] Gatta F. M. et al.: Single-Pole Autoreclosure in uncompensated EHV AC mixed overhead-cable lines: A parametric time-domain analysis. Electric Power Systems Research 210, 2022, 108055.

[8] Giraudet F.: Line surge arresters: applications, designs, trends, monitoring and recommendations. Conference EARTHING AFRICA 2017, South Africa.

[9] Grebović S. et al.: The principles of a new line surge arrester's transient current measurement system. Electric Power Systems Research 223, 2023, 109633.

[10] Ioannidis A. I., Datsios Z. G., Tsovilis T. E.: Estimating the shielding failure flashover rate of single-circuit overhead lines with horizontal phase configuration via stochastic lightning attachment simulations. Electric Power Systems Research 223, 2023, 109620.

[11] Koch H.: Gas Insulated Lines (GIL). Krieg T., Finn J. (eds): Substations. CIGRE Green Books. Springer, Cham. 2019 [https://doi.org/10.1007/978-3-319-49574-3_27].

[12] Li B., Gu T., Li Z., Li B.: Fault section identification method for the UHV GIL-overhead hybrid line. The Journal of Engineering 2019(16), 2019 [https://doi.org/10.1049/joe.2018.8693].

[13] Lin W. et al.: Evaluating the Lightning Strike Damage Tolerance for CFRP Composite Laminates Containing Conductive Nanofillers. Applied Composite Materials 29, 2022, 1537–1554 [https://doi.org/10.1007/s10443-022-10028-1].

[14] Liu B. et al.: Insulation design of -800 kV gas insulation transmission line for negative ion based neutral beam injector. Fusion Engineering and Design 196, 2023, 114027.

[15] Montanyà J. et al.: Potential use of space-based lightning detection in electric power systems. Electric Power Systems Research 213, 2022, 108730 [https://doi.org/10.1016/j.epsr.2022.108730].

[16] Niu H. et al.: Multi-Physical Coupling Field Study of 500 kV GIL: Simulation, Characteristics, and Analysis. IEEE Access 8, 2020, 131439–131448 [https://doi.org/10.1109/ACCESS.2020.3009694].

[17] Qiuqin S. et al.: Surge analysis for lightning strike on overhead lines of wind farm. Electric Power Systems Research 194, 2021, 107066 [https://doi.org/10.1016/j.epsr.2021.107066].

[18] Rui Q. et al.: Methods for alleviation of impacts of axial diffusion on decomposition products monitoring in gas-insulated transmission lines. High Voltage 7, 2022, 41–51.

[19] Runyu F. et al.: Very fast transient overvoltage calculation and evaluation for 500-kV gas insulated substation power substation with double circuit and long gas insulated substation busbar. IET Gener. Transm. Distrib. 17, 2023, 252–262.

[20] Sadovic S., Sadovic T.: Line Surge Arresters Applications On The Multi Circuit Overhead Lines. Journal of Energy – Energija 60, 2011, 75–80 [https://doi.org/10.37798/2011601-4265].

[21] Samira B., Boumous Z., Anane Z., Nouri H.: Comparative study of 220 kV overhead transmission lines models subjected to lightning strike simulation by using electromagnetic and alternative transients program. Electrical Engineering & Electromechanics 4, 2022, 68–74 [https://doi.org/10.20998/2074-272X.2022.4.10].

[22] Shakeel A., Park K., Shin K.-Y., Lee B.-W.: A Study of Fast Front Transients of an HVDC Mixed Transmission Line Exposed to Bipolar Lightning Strokes. Energies 14, 2021, 2896 [https://doi.org/10.3390/en14102896].

[23] Sieminski A., Donovan C.: Forecasting overhead distribution line failures using weather data and gradient-boosted location, scale, and shape models. 2022 [https://doi.org/10.48550/arXiv.2209.03495].

[24] Vendin S., Solov'ev S., Kilin S., Yakovlev A.: Modeling and Analysis of Lightning Protection in an Emergency Situation of a Lightning Strike. Elektrotekhnologii i elektrooborudovanie v APK 3, 2021, 37–47 [https://doi.org/10.22314/2658-4859-2021-68-3-37-47].

[25] Wenjia X., Xiang Z., Qiyan M.: Research on Induced Voltage and Current for Hybrid Transmission System Composed of GIL and Overhead Line. International Journal of Emerging Electric Power Systems 19(6), 2018, 20180108 [https://doi.org/10.1515/ijeeps-2018-0108].

M.Sc. Samira Boumous
e-mail: samira.boumous@univ-soukahras.dz

She is born in 1975, Algeria. Presently she is an associate professor in electrical engineering, she is a senior lecturer in Souk Ahras University, member in "LEER" Laboratory of Electrical Engineering and Renewable Energy.
She is currently working on smart grid modeling, optimization techniques and renewable energies.

https://orcid.org/0000-0003-2213-6542

M.Sc. Zouhir Boumous
e-mail: zohir.boumous@univ-soukahras.dz

He is born in 1974, Algeria. Presently he is an associate professor in electrical engineering, he is a senior lecturer in Souk Ahras university, member in "LEER" Laboratory of Electrical Engineering and Renewable Energy. He focuses his research on fault detection in electrical systems, tolerant control, electric vehicles and renewable energies.

https://orcid.org/0000-0003-4972-1805

M.Sc. Yacine Djeghader
e-mail: yacine.djeghader@univ-soukahras.dz

Presently he is an associate professor in the Department of Electrical Engineering, University of Mohamed-Cherif Messaadia of Souk Ahras, Algeria; Laboratory of Electrical Engineering and Renewable Energy (LEER).
His fields of interest are power quality, power systems, FACTS, and power electronic.

https://orcid.org/0000-0002-7097-7808

http://doi.org/10.35784/iapgos.5777 *received: 11.01.2024 | revised: 17.01.2024 | accepted: 11.03.2024 | available online: 31.03.2024*

ENERGY EFFICIENCY OF PHOTOVOLTAIC PANELS DEPENDING ON THE STEP RESOLUTION OF TRACKING SYSTEM

Kamil Płachta
Wroclaw University of Science and Technology, Faculty of Electronics, Photonics and Microsystems, Wroclaw, Poland

Abstract. *The article presents an energy analysis of a 3.5 kWp photovoltaic installation placed on a two-axis tracking system, depending on resolution of step tracking system, that tracks apparent position of the Sun on the celestial sphere. Measurements were taken during July and August, months with similar solar radiation intensity. During the first month, the tracking system changed the spatial orientation of the photovoltaic panels with a frequency of 20 minutes, while in the second month the resolution of the tracking step was 120 minutes. The total energy production by the photovoltaic installation cooperating with the tracking system was 589.5 kWh and 579.85 kWh, for a tracking step resolution of 20 and 120 minutes, respectively. The monthly difference between the two analysed periods does not exceed 1.7%. However, when analysing the days with the highest energy production – exceeding 28 kWh/day, the photovoltaic installation which changed its spatial orientation with greater frequency produced 309.83 kWh, and with a smaller one 259.88 kWh. In the case of sunny, cloudless days, the difference in the efficiency of both solutions is equal to 19%. During days with lower solar radiation, the efficiency of the photovoltaic installations was similar. It can be concluded that increasing the step resolution of the tracking system increases energy production on sunny, cloudless days. It should be taken into account that increasing the frequency of changing the position of photovoltaic panels increased energy consumption by tracker motors from 2.48 kWh to 3.75 kWh, which constitutes 13.2% of the energy gain obtained over the entire tested period, but less than 1% during days with the highest amount of solar radiation.*

Keywords: photovoltaic panels, solar tracker, solar map, efficiency of energy production

EFEKTYWNOŚĆ ENERGETYCZNA PANELI FOTOWOLTAICZNYCH W ZALEŻNOŚCI OD ROZDZIELCZOŚCI KROKU ŚLEDZENIA UKŁADU NADĄŻNEGO

Streszczenie. *W artykule przedstawiono analizę energetyczną instalacji fotowoltaicznej o mocy 3,5 kWp umieszczonej na dwuosiowym układzie nadążnym, w zależności od rozdzielczości kroku śledzenia pozornej pozycji Słońca na sferze niebiskiej. Pomiary wykonano w trakcie lipca i sierpnia, miesięcy charakteryzujących się zbliżoną wartością natężenia promieniowania słonecznego. W trakcie pierwszego miesiąca, układ nadążny zmieniał orientację przestrzenną paneli fotowoltaicznych z częstotliwością równą 20 minut, natomiast w drugim miesiącu rozdzielczość kroku śledzenia wyniosła 120 minut. Całkowita produkcja energii elektrycznej przez instalację fotowoltaiczną współpracującą z układem nadążnym była równa 589,5 kWh oraz 579,85 kWh, odpowiednio dla rozdzielczości kroku śledzenia równego 20 oraz 120 minut. Miesięczna różnica między dwoma badanymi okresami nie przekroczyła 1.7%. Natomiast analizując dni o największej produkcji energii elektrycznej – powyżej 28 kWh, instalacja fotowoltaiczna zmieniająca swoją orientacje przestrzenną z większą rozdzielczością kroku śledzenia wyprodukoała 309,83 kWh, natomiast z mniejszą 259,88 kWh. W przypadku słonecznych, bezchmurnych dni, różnica w efektywności obu rozwiązań wynosi 19%. W trakcie dni charakteryzujących się mniejszą wartością nasłonecznienia, efektywność instalacji była do siebie zbliżona. Podsumowując, zwiększenie rozdzielczości kroku układu nadążnego powoduje wzrost produkcji energii elektrycznej w słoneczne, bezchmurne dni. Natomiast, zwiększenie częstotliwości zmiany położenia paneli fotowoltaicznych zwiększa zużycie energii elektrycznej z 2,48 kWh do 3,75 kWh, co stanowi 13,2% uzyskanego zysku energetycznego w całym badanym okresie, ale niespełna 1% w trakcie dni o największej wartości nasłonecznienia.*

Słowa kluczowe: panele fotowoltaiczne, system śledzący, mapa słoneczna, efektywność produkcji energii

Introduction

The first works on a system tracking apparent position of the Sun date back to the 1970s and 1980s. In 1979, Jerome H. Weslow and James A. Rodrian submitted a patent to the American Patent Office, which described a solution allowing to determine the apparent position of the Sun on the celestial sphere, the so-called tracker. The description concerned the method of moving and positioning the tracking system. Tracking systems are divided according to the number of rotation axes: single- and dual-axis solutions. The single-axis systems are divided into vertical (north-south) and horizontal (east-west) [17]. The advantage of single-axis solutions is the lower complexity of the tracking system, while the efficiency of the system with a movable vertical axis is higher compared to a movable horizontal axis. The issue of energy consumption of the tracking system has been described in many publications, the aim of which was to maximize energy production while limiting its consumption [12]. Solutions that significantly reduce energy consumption include tracking systems that use the change in the shape of the material due to temperature changes. In this solution, the electric motors were replaced with bimetallic steel and aluminum strips, which bend due to temperature changes, causing the photovoltaic panels to move [5].

1. Description of the problem

1.1. Types of algorithms

The apparent position of the Sun on the celestial sphere is determined by the tracking system controller using the differential, algebraic, hybrid and MPP (*Maximum Power Point*) methods [2, 14, 20]. The differential method uses data from photodetectors to determine the apparent position of the Sun on the celestial sphere. The photo-detector function can be performed by photoresistor, photodiode, phototransistor or small photovoltaic cell [6]. The photodetectors converts the solar radiation intensity into voltage, the value of which is measured by a microcontroller and, depending on its difference, the photovoltaic (PV) modules are rotated towards the photodetector with the highest value of solar radiation [8]. The effectiveness of the differential algorithm depends on the quality of the implemented algorithm and method the photodetectors are arranged [11, 18]. The algorithm determines the hysteresis value, which defines the step resolution of the tracking system. Too low resolution will reduce the efficiency of the photovoltaic installation, while too high resolution will increase the electricity consumption of the tracker motors. The voltage value of photodetectors can be used to determine atmospheric conditions and, in the case of heavy cloud cover, arrange the photovoltaic panels horizontally to the ground surface. There are two ways to set up photodetectors. The first involves the use of a partition that limits the amount of incident solar radiation. The second solution is to arrange the photodetectors at an appropriate angle, so that each detector determines one of the geographical directions. The advantages of this solution include low complexity of the measurement system, no calibration of the positioning of photovoltaic panels and response to changing lighting conditions. However, in the case of weak sunlight, no electricity is consumed by the tracker's engines. The disadvantages include susceptibility to interference caused by dirt or shading [1].

The algebraic method involves calculating or using a built-in map of the Sun's path [9, 19, 21]. The map contains information about the Sun's declination and azimuth for each day of the year

at specific time intervals, therefore the tracking system controller does not require the use of external sensors. The lack of photodetectors means that the system does not respond to changing weather conditions [10]. The accuracy of determining the apparent position of the Sun depends on the size of the implemented solar map [7]. The more accurate the solar map causes frequent changes in the position of the PV panels. The simple design causes the system to have several drawbacks: lack of response to changing lighting conditions. The lack of use of detectors makes it impossible to eliminate the shading effect of photovoltaic surface.

The hybrid algorithm uses a combination of the two previously described methods. The positioning of photovoltaic panels is based on the data contained in the solar map. However, to precisely determine the apparent position of the Sun on the celestial sphere, data read from photodetectors is used, which eliminates the use of large solar maps. The applied of photodetectors enables the tracking system to respond to changing lighting conditions, thus eliminating all the above described disadvantages of the algebraic algorithm. The implementation of the solar map in the software of the tracking system controller does not increase its complexity, but it provides protection in the event of damage to the photodetectors. The accuracy of determining the apparent position of the Sun on the celestial sphere is comparable to the differential algorithm that searches for the maximum power point [3, 4].

The MPP algorithm involves searching for the maximum power point – tracking the maximum output power of PV panels [15, 16]. The change in the position of the Sun causes uneven illumination of the photovoltaic installation, which affects the output power of each module. The tracking system controller compares the voltage values and rotates towards the PV module producing the most energy. The presented solution requires the use of modules measuring the parameters of photovoltaic panels, which affects the complexity and costs of the measurement system. The advantage of the extensive measurement system is the ability to eliminate the phenomenon of shading photovoltaic modules. However, the step resolution of the tracking system can be determined based on the maximum difference in output power between individual panels. The advantage of this solution is the elimination of the shading effect of photovoltaic surface. The response to changing lighting conditions and the lack of electricity consumption by the motors in the event of poor sunlight. However, the disadvantages include the high complexity of the control system and susceptibility to interference, e.g. contamination of the photovoltaic surface.

1.2. Calculating the apparent position of the Sun

The azimuth angle and the elevation angle at solar noon are two angles that are used to orient photovoltaic modules relative to the apparent position of the Sun on the celestial sphere. These angles are calculated using "solar time". The regions of the Earth are divided into specific time zones. However, in these time zones, noon does not necessarily correspond to the time when the Sun is at its highest point in the sky. Similarly, sunrise is defined as the stage at which the Sun rises in one part of a time zone. However, due to the distance traveled in one time zone, the time at which the Sun actually leaves the horizon in one part of the time zone may be completely different than the "defined" sunrise (or officially recognized as the time of sunrise). Such assumptions are necessary, otherwise a house located one block from another would actually differ in time by several seconds. On the other hand, solar time is unique for each specific longitude. Consequently, to calculate the Sun's position, the local solar time is first found, and then the Sun's elevation angle and azimuth angle are calculated [10].

Solar declination is angle between the solar radiation falling towards the observer and plane of the Earth's equator. In the northern hemisphere, the declination is positive (from 0° to 90°), in the southern hemisphere it is negative (from 0° to -90°). The Earth's axial tilt is the angle between the Earth's axis and a line perpendicular to its orbit. The current value of the Earth's axial tilt is $\varepsilon = 23°26'$. The value of the declination angle can be written as follows

$$\delta = -23.44° \cdot cos\left[\frac{360°}{365.24} \cdot (N + 10)\right] \quad (1)$$

where N is the number of days since midnight UTC since the beginning of January. The value of +10 in the above equation is due to the fact, that the winter solstice occurs before beginning of year. The equation also assumes that the Sun's orbit is a perfect circle, and the ratio 360/365 determines the day number relative to the Earth's position in its orbit. However, the approximation used results in an error of 0.26°. Moreover, determining the value of the declination angle near the September equinox is subject to an error of +1.5°. Therefore, equation 1. may introduce an error up to 2°. The declination value can be calculated more precisely by not making an approximation and using the parameters of the Earth's orbit to estimate the value more precisely

$$\delta = arcsin[sin(-23.44°) \cdot \cdot cos\left(\frac{360°(N+10)}{365.24} + \frac{360° \cdot 0.0167}{\pi} \cdot sin\left(\frac{360°(N-2)}{365.24}\right)\right)] \quad (2)$$

where the number 2 in *N–2* is the approximate number of days after January 1th until Earth's perihelion. The calculation error does not exceed ±0.2°. Moreover, it may be less than ±0.03° if the value of "10" in the above equation is adjusted according to the date of occurrence of the December solstice in the previous year. These accuracies are comparable to advanced calculations using the Jean Meeus algorithm with an accuracy of 0.01°. The value of the Sun's declination can be determined using the following algorithm: PSA (*Position Sun Algorithm*); NREL – Sun Position Algorithm and Duffie & Beckman algorithm [13]. These algorithms are characterized by very high accuracy in determining the declination, which is desirable in solar radiation concentrator systems that use lenses for each photovoltaic cell.

The angle of elevation is the angular height of the Sun measured relative to a horizontal plane, which at sunrise and sunset is 0°, and when the Sun is directly above the object, it is equal to 90°. The angle of elevation can be calculated using the following relationship

$$\alpha_S = sin^{-1}[sin\,\delta\,sin\,\varphi + cos\,\delta\,cos\,\varphi\,cos(HRA)] \quad (3)$$

where φ – geographical latitude of the object, δ – declination angle depending on the day of the year. The hour angle (HRA) converts local solar time (LST) into a number of degrees that correspond to the Sun's path across the sky. Each hour corresponds to a change in the position of the Sun in the sky by 15°. The hour angle at noon is 0°, in the morning it is negative and in the afternoon it is positive. The HRA value is described by the following equation

$$HRA = 15° \cdot (LST - 12) \quad (4)$$

where LST – local solar time. Twelve noon local solar time is defined when the Sun is highest in the sky. The value of the azimuth angle is determined based on the following relationship

$$A = cos^{-1}\left[\frac{sin\,\delta\,cos\,\varphi - cos\,\delta\,sin\,\varphi\,cos(HRA)}{cos\,\alpha_s}\right] \quad (5)$$

Knowing all the parameters, the time of sunrise ($t_{sunrise}$) and sunset (t_{sunset}) can be calculated as follows

$$t_{sunrise} = 12 - \frac{1}{15} \cdot cos^{-1}\left(\frac{-sin\,\varphi\,sin\,\delta}{cos\,\varphi\,cos\,\delta}\right) - \frac{TC}{60} \quad (6)$$

$$t_{sunset} = 12 + \frac{1}{15} \cdot cos^{-1}\left(\frac{-sin\,\varphi\,sin\,\delta}{cos\,\varphi\,cos\,\delta}\right) - \frac{TC}{60} \quad (7)$$

For the geographical coordinates 50°46′N 20°37′E, the analysis of insolation depending on day of year was performed. The chart below (Fig. 1) shows the sunrise and sunset times and the length of the day expressed in hours.

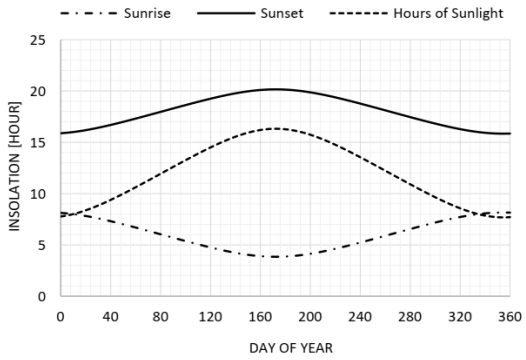


Fig. 1. The value of insolation depending on day of year for geographic coordinates 50°46'N 20°37'E

The time correction factor (TC) [minutes] takes into account changes in local solar time (LST) in a given time zone due to changes in longitude in the time zone

$$\mathrm{TC} = 4 \cdot (\mathrm{longitude} - \mathrm{LSTM}) + \mathrm{EoT} \tag{8}$$

where *EoT* – equation of time, *LSTM* – local standard time meridian. The *LSTM* is a reference meridian used for a specific time zone and is similar to the prime meridian used for Greenwich time. The *LSTM* is shown below

$$LSTM = 15° \cdot \Delta T_{GMT} \tag{9}$$

where ΔT_{GMT} is the difference between local time (LT) and universal time (GMT), expressed in hours. The equation of time (EoT) is used to correct the average time to which the *EoT* value must be added or subtracted to obtain the real time. After calculating *EoT* based on the equation below, the result is expressed in minutes, with an accuracy of ½ minute

$$\mathrm{EoT} = 9.87 \cdot \sin(2\mathrm{N_d}) - 7.53 \cdot \cos(\mathrm{N_d}) - 1.5 \cdot \sin(2\mathrm{N_d}) \tag{10}$$

where N_d parameter is described by the relationship

$$N_d = \frac{360}{365} \cdot (N - 81) \tag{11}$$

where N_d is expressed in degrees and N is the next day of year. The figure 2 shows the average time correction based on the time equation

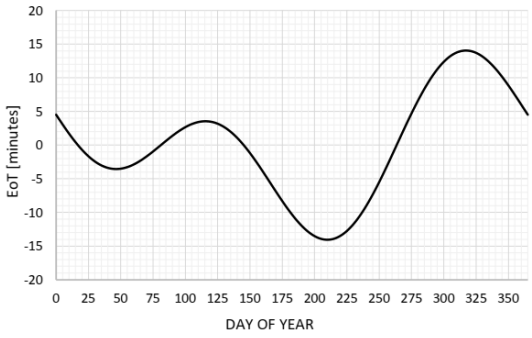


Fig. 2. The value of the equation of time (EoT) depending on day of year for geographical coordinates 50°46'N 20°37'E

Based on the chart it can be concluded that the real solar time and the average solar time are equal when the Sun is near the point of Cancer ($\delta = 23°27'$) or Capricorn ($\delta = -23°27'$) and near the place where $\delta = 10°$. At other times, the Sun passes through the meridian earlier or later so the equation of time takes a negative or positive value. The curve showing the difference between real and average solar time is called an analemma. The cause of the analemma is the movement of the Earth around the Sun and the inclination of the Earth's rotation axis to the orbital plane. The variable length of the day affects the height of the Sun, the highest point on the ecliptic analemma is reached during the summer and winter solstices. The difference between real and average solar time should be taken into account when developing solar maps for tracking system controllers that do not use external photodetectors.

2. Measurement setup

The measurement setup consists of 10 photovoltaic modules was placed on a system that tracks the apparent position of the Sun on the celestial sphere. The total power of the PV installation is equal to 3.5 kWp and the power of the inverter of 3.3 kW. The photovoltaic installation using a tracking system, the power of inverter should be slightly lower than the power of photovoltaic panels, in this case is 94%.

The tracking system was designed by the author of this article in a 3D modeling program. The maximum horizontal rotation angle is equal to 235°, whereas the vertical rotation angle 90°. The frame size is 542×440 cm and is adapted to 10 photovoltaic modules. The design and the finished device are shown in the figure 3.

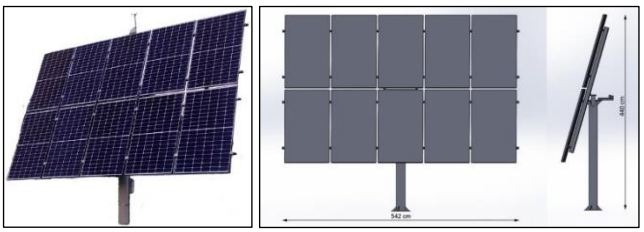


Fig. 3. Photovoltaic modules with tracking system (on the left), model 3D (on the right)

The tracking system has two electric motors with built-in limit switches, which maintain their position in the event of a power outage. To control the tracking system, was designed and created solar controller, that determines the apparent position of the Sun on the celestial sphere. The tracking system controller (Fig. 4) uses several modules, e.g. the GPS module (NEO-6M) to determine local geographic coordinates and the current date and time. The accelerometer and gyroscope module (MPU6050) and the magnetometer module (HMC5883L) determine the current position of photovoltaic panels relative to the Earth's surface. The ESP-12F Wi-Fi module is used to acquire measurement results. Four high-power motor driver module BTS7960 were used to control the electronic motors of the tracking with smooth PWM (*Pulse-Width Modulation*) regulation in the range of 0-100%. The motor voltage range is 6–27 V and the maximum current is 43 A. The BTS7960 module has a number of built-in protections, including: short-circuit, overload and thermal protection. The figure 4 shows the tracking system controller.

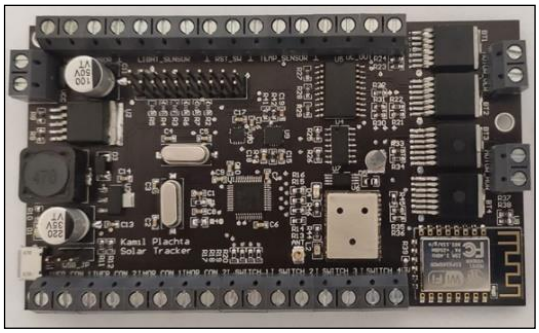


Fig. 4. Photo of the tracking system controller

After initialization, the controller measures the wind speed. If it is higher than the set protection value, the photovoltaic panels are placed in a safe position - parallel to the ground for 5 minutes. If the wind speed does not exceed the protection value, the controller collects data from the GPS module, accelerometer, gyroscope and magnetometer. Based on the received data, the Sun's path is calculated for a specific day of the year. The next stage is to precisely determine the spatial orientation of photovoltaic modules. The controller determines the current position of the PV modules in the horizontal plane based

on the measurement of the change in the magnetic field value and the acceleration. However, the data obtained from the gyroscope and accelerometer are used to calculate the inclination angle relative to the Earth's surface.

In the horizontal system, the horizon plane and the direction of the vertical are determined for a given celestial sphere. The *x* and *y* axes lie in the horizon plane and determine the north and east directions, respectively, while the *z* axis points upwards. The location of a point on the celestial sphere is determined by the coordinates of azimuth *A* and height *h*. The height of an object on the celestial sphere is the angle between the plane of the horizon and the direction toward the object. The zenith distance is defined as the difference of 90° and the height *h*.

The translation of the *xyz* system around the *z* axis by the rotation angle α is shown in the figure 5.

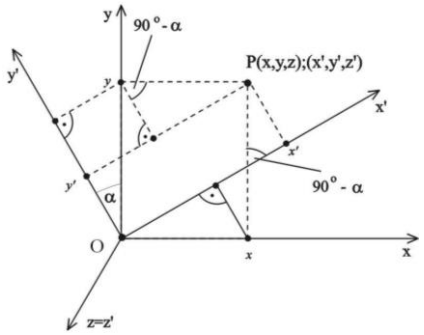


Fig. 5. Translation around the z axis for positive rotation by angle α

The coordinates of a point in the new coordinate system are described by the following equations:

$$x' = \mathrm{x} \cdot \cos\alpha + y \cdot \sin\alpha$$
$$y' = -x \cdot sin\,\alpha + y \cdot cos\,\alpha \quad (12)$$
$$z' = z$$

To transform the horizontal system into the equatorial hour system, rotate the horizontal system around the *y* axis by an angle of 90°-φ, then by an angle of -180° around the *z* axis. In the case of translation in the opposite direction, rotate the hour system by an angle of 90°-φ around *y* axis and by an angle of 180° around the *z* axis. The transformation of the horizontal system into the equatorial hour system can be performed using the following system of equations:

$$\cos\delta\cos t = \sin h\cos\varphi - \cos h\sin\varphi\cos\alpha$$
$$cos\,\delta\,sin\,t = -cos\,h\,sin\,\alpha \quad (13)$$
$$\sin\delta = \sin h\sin\varphi + \cos h\cos\varphi\cos\alpha$$

The inverse transformation can be performed using the equations:

$$\cos h\cos\alpha = \sin\delta\cos\varphi - \cos\delta\sin\varphi\cos t$$
$$cos\,h\,sin\,\alpha = -cos\,\delta\,sin\,t \quad (14)$$
$$\sin h = \sin\delta\sin\varphi + \cos\delta\cos\varphi\cos t$$

Where the *t* can be calculated using the equation

$$t = arc\,tan\frac{sin\,\alpha}{tan\,h\;cos\,\varphi + sin\,\varphi\,cos\,\alpha} \quad (15)$$

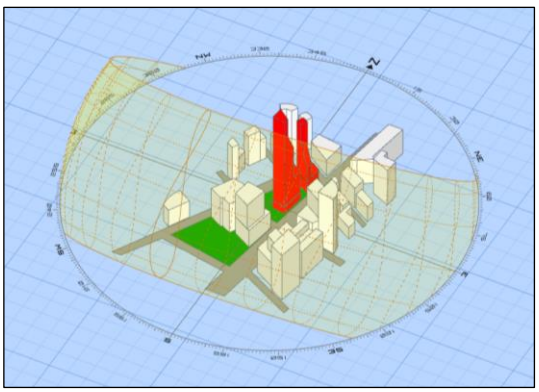


Fig. 6. Sun path for location 50°46'53"N 20°37'16"E (created with andrewmarsh.com)

The controller of tracking system calculates the apparent position of the Sun on the celestial sphere, creating a Sun path for each day of the year.

Visualization of the Sun's path for coordinate locations 50°46′53″N and 20°37′16″E, where the photovoltaic system is located.

3. Measurement results

The efficiency of a photovoltaic installation placed on a two-axis tracking system with a step resolution of 20 min and 120 min was measured. The two warmest months of the year were selected to verify the presented solutions: July and August. In July, the step resolution of the tracking system was 20 minutes, and in August it was 120 minutes. The monthly energy consumption of the electric motors was 3.75 kWh in July and 2.48 kWh in August. Electricity production depending on the day of the month is shown in figures 7 and 8.

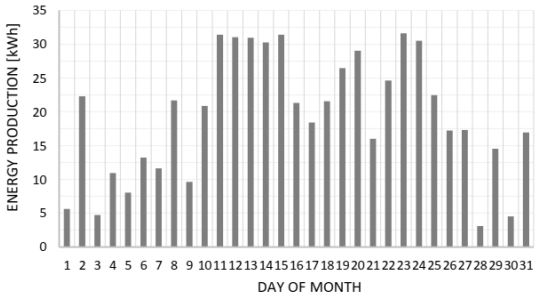


Fig. 7. Energy production with a step resolution of tracking system equal to 20 min

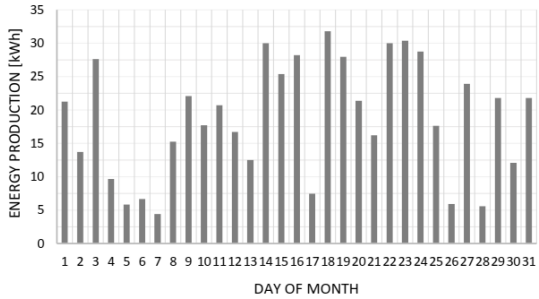


Fig. 8. Energy production with a step resolution of tracking system equal to 120 min

The total energy production by the photovoltaic installation cooperating with the tracking system was 589.5 kWh and 579.85 kWh, for the tracking step resolution of 20 and 120 minutes, respectively. The monthly difference between two presented solutions does not exceed 1.7%. The average daily energy production in July was 19.34 kWh and in August 18.71 kWh. Meteorological conditions were not the same for both tested months, but very similar. There were 16 days in each month when electricity production exceeded the average value. In order to better present the difference in efficiency of tracking system with a step resolution equal to 20 and 120 minutes, figure 9 shows the diffe-rences in the amount of energy produced for each day.

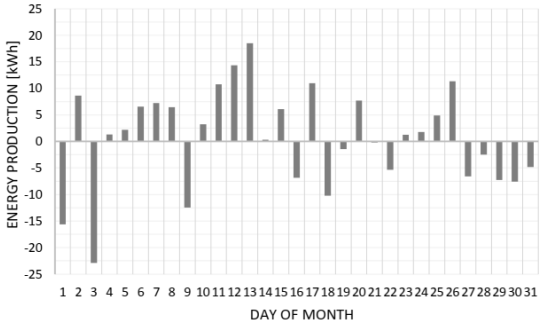


Fig. 9. The difference between the amount of energy produced by the system with a tracking step resolution of 20 min (above the x-axis) and 120 min (under the x-axis)

The data presented in the figure 9 shows that a higher tracking step resolution affects the number of days in which higher energy production was obtained. Discarding differences lower than 1 kWh, i.e. on the 14th and 21st day of the month, the system with a tracking step of 20 minutes ensured greater energy production during 17 days, while in the case of a tracking step of 120 min, the difference was visible during 11 days.

In order to reliably analyse the effectiveness of the presented solutions, it is necessary to compare the energy production results for days with similar solar radiation intensity. The chart below shows the distribution of electricity production in selected ranges.

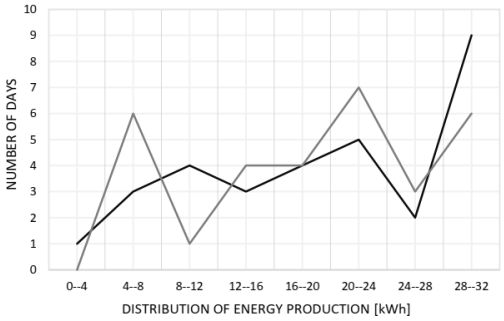


Fig. 10. Distribution of energy production for tracking step resolutions equal to 20 min (black line) and 120 minutes (grey line)

Based on the data presented in the Fig. 10, it can be concluded that the higher resolution of step tracking system enusred 9 days in which the amount of energy produced exceeded 28 kWh, for resolution of 120 minutes, only 6 days were obtained. In the case of a lower frequency of rotation of photovoltaic panels, the number of days with poor (4–8 kWh), medium (20–24 kWh) and good (28–32 kWh) energy production is similar. During the tested period, there were days with different meteorological conditions. Therefore in analysing effectiveness of presented solutions, the days with the highest intensity of solar radiation should be taken into account. During this days, the photovoltaic panels changing its spatial orien-tation with a higher tracking step resolution produced 309.83 kWh, while with a smaller one 259.88 kWh. The difference in the amount of energy produced is 19%. The increase of energy consumption by the tracker motors does not exceed 1% profit of energy produced. Based on the results obtained, it can be concluded that an appropriately selected step resolution of tracking system increases the amount of energy produced by photovoltaic panels.

4. Conclusions

The article presents an energy analysis of the photovoltaic installation placed on the two-axis tracking system with different resolution of step tracking. The measurements were taken during two months characterized by the greatest and similar solar radiation intensity.

The total energy production by the photovoltaic installation cooperating with the tracking system was equal to 589.5 kWh and 579.85 kWh for step tracking resolution equal to 20 and 120 min, respectively. The monthly difference between the two tested months does not exceed 1.7%. However, analyzing sunny, cloudless days in which the value of energy production exceeded 28 kWh/day, the photovoltaic panels changing spatial orientation with a higher step resolution produced 309.83 kWh, while with a smaller one it produced 259.88 kWh, the difference is 19%. During days with lower insolation (e.g. autumn and winter months) the effectiveness of the tested solutions was similar.

Based on the obtained results, it can be concluded that increasing the step resolution of the tracking system increases the amount of energy produced for months with the highest amount of solar radiation intensity. Increasing the frequency of changing the position of photovoltaic panels causes an increase of energy consumption from 2.48 kWh to 3.75 kWh, which constitutes 13.2% of the obtained profit of energy produced over the entire tested period. Considering days with the highest energy production, the increase of electricity consumption by the tracker motors was less than 1%. Therefore, the frequency of changing the position of photovoltaic panels relative to the apparent position of the Sun should be individually selected, depending on the sunlight conditions for each month of the year, so that the increase of energy consumption by the motors is not greater than the energy gained by photovoltaic panels.

References

[1] Bartczak M.: On the capacity of solar cells under partial shading conditions. Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska – IAPGOS 11(4), 2021, 47–50.
[2] Belhachat F., Larbes C.: A review of global maximum power point tracking techniques of photovoltaic system under partial shading conditions. Renewable and Sustainable Energy Reviews 92, 2018, 513–553.
[3] Bentata K., Mohammedi A., Benslimane T.: Development of rapid and reliable cuckoo search algorithm for global maximum power point tracking of solar PV systems in partial shading condition. Archives of Control Sciences 2021, 495–526.
[4] Bollipo R., Mikkili S., Bonthagorla P.: Critical review on PV MPPT techniques: classical, intelligent and optimisation. IET Renewable Power Generation 14(9), 2020, 1433–1452.
[5] Clifford M., Eastwood D.: Design of a novel passive solar tracker. Solar Energy 77, 2004, 269–280.
[6] Dadi V., Peravali S.: Optimization of light-dependent resistor sensor for the application of solar energy tracking system. SN Applied Sciences 2(9), 2020.
[7] Duarte F., Gaspar P., Gonçalves L.: Two axes solar tracker based on solar maps, controlled by a low-power microcontroller. Journal of Energy and Power Engineering 5(7), 2011.
[8] Fathabadi H.: Novel high accurate sensorless dual-axis solar tracking system controlled by maximum power point tracking unit of photovoltaic systems. Applied Energy 173, 2016, 448–459.
[9] Fathabadi H.: Novel online sensorless dual-axis sun tracker. IEEE/ASME transactions on mechatronics 22(1), 2016, 321–328.
[10] Karttunen H. et al.: Fundamental Astronomy. Springer 2014.
[11] Lan J.: Development and performance test of a novel solar tracking sensor. Metrology and Measurement Systems 2023, 2023, 289–303.
[12] Mah A., Ho W., Hassim M., Hashim H.: Optimization of Photovoltaic Array Orientation and Performance Evaluation of Solar Tracking Systems. Chemical Engineering Transactions 83, 2021, 109–114.
[13] Melo K., Tavares L., Villalva M.: Statistical Analysis of Solar Position Calculation Algorithms: SPA and Grena 1–5. IEEE Latin America Transactions 19(7), 2021, 1145–1152.
[14] Mroczka J., Ostrowski M.: A hybrid maximum power point search method using temperature measurements in partial shading conditions. Metrology and Measurement Systems (4), 2014.
[15] Mroczka J., Ostrowski M.: Maximum power point search method for photovoltaic panels which uses a light sensor in the conditions of real shading and temperature. Modeling Aspects in Optical Metrology V, 2015, 371–378.
[16] Mroczka J., Ostrowski M.: A hybrid maximum power point tracking algorithm that uses the illumination and the temperature sensor in solar tracking systems. Nonlinear Optics and Applications XI, 2019, 237–243.
[17] Prinsloo G., Dobson R.: Solar Tracking, 2015.
[18] Salgado-Conrado L.: A review on sun position sensors used in solar applications, Renewable and Sustainable Energy Reviews 82, 2018, 2128–2146.
[19] Sidek M. et al.: Automated positioning dual-axis solar tracking system with precision elevation and azimuth angle control. Energy 124, 2017, 160–170.
[20] Talha A., Boumaaraf H., Bouhali O.: Evaluation of maximum power point tracking methods for photovoltaic systems. Archives of Control Sciences, 2011.
[21] Wu C., Wang H., Chang H.: Dual-axis solar tracker with satellite compass and inclinometer for automatic positioning and tracking. Energy for Sustainable Development 66, 2022, 308–318.

Ph.D. Kamil Płachta
e-mail: kamil.plachta@pwr.edu.pl

Kamil Płachta received Ph.D. degree at the Wroclaw University of Science and Technology, Poland. He has been a member of the teaching and research staff at the Department of Electronic and Photonic Metrology since 2013. His current research interests include renewable energy sources, mainly on systems that tracks the apparent position of the Sun on the celestial sphere.

https://orcid.org/000-0001-6278-1488

http://doi.org/10.35784/iapgos.5373 *received: 27.09.2023 | revised: 21.02.2024 | accepted: 03.03.2024 | available online: 31.03.2024*

DIGITAL IMAGE RESTORATION USING SURF ALGORITHM

Shanmukhaprasanthi Tammineni[1], Swaraiya Madhuri Rayavarapu[1], Sasibhushana Rao Gottapu[1], Raj Kumar Goswami[2]

[1]Andhra University College of Engineering, Department of Electronics and Communication Engineering, Visakhapatnam, India, [2]Gayatri Vidya Parishad College of Engineering for Women, Departement of Electronics and Communication Engineering, Visakhapatnam, India

Abstract. *In contemporary times, the preservation of scientific and creative endeavours often relies on the utilization of film and image archives, hence emphasizing the significance of image processing as a critical undertaking. Image inpainting refers to the process of digitally altering an image in a manner that renders the adjustments imperceptible to a viewer lacking knowledge of the original image. Image inpainting is a technique mostly employed to restore damaged regions within an image by utilizing information obtained from matching characteristics in relevant images. This process involves filling in the damaged areas and removing undesired objects. The SURF (Speeded Up Robust Feature) algorithm under consideration is partitioned into three primary phases. Firstly, the essential characteristics of the impaired image and the pertinent image are identified. In the second stage, the relationship between the damaged image and the relevant image is determined in terms of translation, scaling, and rotation. Ultimately, the destroyed area is reconstructed through the application of the inverse transformation. The quality assessment of inpainted images can be evaluated using metrics such as Structural Similarity Index (SSIM), Peak Signal-to-Noise Ratio (PSNR), and Mean Squared Error (MSE). The experimental findings provide evidence that the suggested inpainting technique is effective in terms of both speed and quality.*

Keywords: SURF, inpainting, image restoration, image features

CYFROWA REKONSTRUKCJA OBRAZÓW Z WYKORZYSTANIEM ALGORYTMU SURF

Streszczenie. *We współczesnych czasach utrwalanie dorobku naukowego i twórczego często opiera się na wykorzystaniu archiwów filmowych i obrazowych, co podkreśla znaczenie przetwarzania obrazu jako przedsięwzięcia krytycznego. Inpainting odnosi się do procesu cyfrowej zmiany obrazu w sposób, który sprawia, że korekty są niezauważalne dla widza nie znającego oryginalnego obrazu. Inpainting to technika stosowana najczęściej w celu przywracania uszkodzonych obszarów obrazu poprzez wykorzystanie informacji uzyskanych na podstawie dopasowania cech odpowiednich obrazów. Proces ten polega na wypełnieniu uszkodzonych obszarów i usunięciu niepożądanych obiektów. Rozważany algorytm SURF (Speeded Up Robust Feature) dzieli się na trzy główne fazy. Po pierwsze, identyfikowane są podstawowe cechy obrazu zaburzonego i obrazu istotnego. W drugim etapie określa się relację pomiędzy obrazem uszkodzonym a obrazem odpowiednim pod względem translacji, skalowania i rotacji. Ostatecznie zniszczony obszar rekonstruuje się poprzez zastosowanie transformacji odwrotnej. Ocenę jakości renowacji obrazów można ocenić za pomocą wskaźników, takich jak wskaźnik podobieństwa strukturalnego (SSIM), szczytowy stosunek sygnału do szumu (PSNR) i błąd średniokwadratowy (MSE). Wyniki eksperymentów dostarczają dowodów na to, że sugerowana technika renowacji jest skuteczna zarówno pod względem szybkości, jak i jakości.*

Słowa kluczowe: SURF, inpainting, renowacja obrazu, cechy obrazu

Introduction

All visual stimuli perceived by the human visual system are stored within the human brain, which can be considered the most extensive storage system in existence. Individuals often make efforts to retain a multitude of information, although there are inevitably certain aspects that elude their memory. As individuals age, there is a gradual deterioration of cognitive function, resulting in the loss of important information from our memory storage system. Every year, a substantial number of images, ranging in the thousands or hundreds, are preserved. Despite the considerable time elapsed since their capture, there is a strong need to access and reminisce upon these cherished instances. Without the presence of a photograph or a well-crafted artistic representation, the true visage of Einstein shall forever remain elusive to humanity. Although historical information is predominantly conveyed through written texts and spoken accounts, the artistic creations produced by artists across different periods of history provide a vivid and tangible representation that enlivens our imaginative faculties. By examining such works, it is possible to gain insight into the lifestyles and environments of various time periods.

The rise in usage of image processing software, such as Photoshop, has presented a growing challenge in distinguishing between an unaltered image and its manipulated counterpart. The presence of numerous manipulated photos that lack discernible evidence has led to a significant need for automated forgery detection systems. These systems are created to assess the reliability and authenticity of a given image [4]. Blind forgeries, unlike digital signatures and watermarks, do not necessitate any prior knowledge of the image. In contrast, the method exclusively relies on the image itself for the purpose of identifying any indications of manipulation [9]. To date, a considerable body of research has been conducted on blind image forensics, encompassing a range of picture tampering techniques such as copy-move forgery, image retouching, image splicing, and even picture steganalysis. The concept of picture inpainting was initially introduced as a means of restoring images and correcting damages. However, picture inpainting can also be employed for the purpose of object removal, wherein the available information in the surrounding areas is utilized to fill the voids created by the process. Image inpainting is a technique that effectively removes objects from an image while maintaining the integrity of the texture and structure. This process ensures that no discernible signs of tampering are left behind. Among all the inpainting techniques, Exemplar based technique is widely used because of its enhanced quality.

The exemplar-based inpainting approach entails iteratively synthesizing the unknown region, referred to as the target region, by employing the most similar patch identified in the source region. Based on the prescribed filling sequence, the approach employs gradient information from adjacent regions to populate structures within the areas that are absent. This methodology demonstrates high efficiency in the reconstruction of extensive target regions. Typically, exemplar-based inpainting methods utilize a pre-determined patch size and do an exhaustive search throughout the whole source region. Nevertheless, the utilization of a patch with a fixed size may present certain disadvantages, as exemplar-based techniques rely on the assumption that textured patterns inside the source region may be discerned adequately with an acceptable patch size. If the patch size is not suitable for adequately covering the designated area of a target region, the accurate assignment of structure and texture information becomes compromised. For instance, in the event that the dimensions of each individual patch are very huge, there is a possibility of erroneous reconstruction of the structure. Conversely, in cases where the patch size is insufficiently large, the process of synthesizing a sizable region with comparable texture patterns becomes too time-consuming [13].

This paper introduces a new methodology for image inpainting that utilizes the Speeded Up Robust Features algorithm and affine transformation to efficiently address computational expenses. It is postulated that the pertinent image depicting the identical situation is already accessible and the data derived

from such image is employed for the rectification. The SURF algorithm demonstrates efficient and precise identification of salient features within both impaired and pertinent images. The establishment of an affine transformation in a repaired image and its corresponding relevant image necessitates the consideration of these important factors. The aforementioned relation serves as a reciprocal transformation for the purpose of retrieving the absent region.

1. Literature review

In the study conducted by Jia et al., the image segmentation process involved the division of an image into many parts, utilizing color and texture information as the basis for this segmentation [8]. Subsequently, each region was subjected to an inpainting procedure on an individual basis. Bertalmio et al., proposed a novel hybrid approach that integrates the diffusion-based scheme with texture creation techniques [2, 3, 7]. The technique demonstrates effective performance in the retrieval of both geometric structures and tiny texture patches.

Criminisi et al., proposed a method that effectively and straightforwardly promotes the completion of missing regions by prioritizing the boundary area where nearby isophotes exhibit high strength [6]. Cheng et al., extended the priority function to enhance the overall performance and reliability of the algorithmic family [5]. Wong et al., proposed a similarity function that use many source patches for the reconstruction of the target patch, as opposed to relying on a single source patch [11]. The method Wu et al., proposes uses local textured information and cross-isophotes diffusion data to calculate the variable size of exemplars in a cross-isophotes exemplar-based model [12]. Inpainting will be used to repair the damaged areas. This technique connects contours by applying curve filling and makes use of the structure information. The majority of contemporary exemplar-based algorithms tend to employ a greedy method, resulting in the inherent drawbacks associated with such algorithms. Specifically, the order in which the exemplars are filled, or their priority, becomes a crucial factor in these algorithms. The effectiveness of exemplar-based inpainting is contingent upon the simplicity of the structure and texture within the missing region. If the number of samples in the image is insufficient, it becomes unfeasible to synthesize the intended image.

The optimization of Markov random fields entails the application of the Scale-Invariant Feature Transform technique to identify interest spots. Furthermore, the direct linear transform approach is utilized to obtain the affine transformation. However, the application of the Scale-Invariant Feature Transform Technique and the integration of multiple sequential stages in the algorithm result in a reduction in overall system efficiency and an escalation in complexity. The absence of a discussion of the performance assessment metric employed for analysing the quality of findings is noted in the report. According to the findings presented, it has been observed that the performance of the SIFT algorithm is comparatively slower when compared to the SURF algorithm. This can be attributed to the fact that SIFT utilizes feature descriptors with larger dimensionality. The performance of SURF is comparable to that of SIFT [1].

2. Proposed method

This study endeavours to address the issue of inpainting by utilizing a pertinent image. When creating the proposed inpainting model, two fundamental assumptions are taken into consideration: The image that is utilized as a reference image is the snapshot obtained at the same place, with potential transformations such as translation, rotation and scaling. Therefore, the system is referred to as the Inpainting Model Based on Affine Transformed Relevant Image. Additionally, it is important to note that the photographs were captured using the same camera and under consistent lighting conditions. The design flow of proposed method is elucidated in figure 1. The corresponding key points in the impaired and pertinent image are acquired utilizing the SURF algorithm. The affine transformations, including scaling, translation, and rotation are applied to the two input images in order to establish a relationship between them. The aforementioned relationship is employed in an inverse manner to determine the extent of the area in the corresponding image that is utilized to fill the region that is absent in the impaired image.

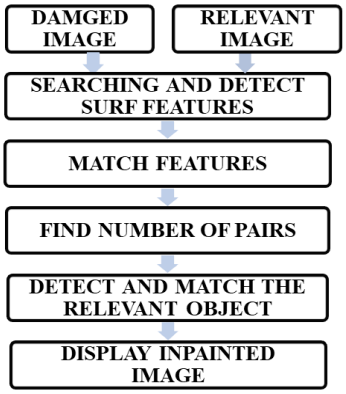


Fig. 1. Flow chart of SURF technique-based image inpainting

The process of designing the transformation function for Inpainting Model Based on Affine Transformed Relevant Image involves the identification and calculation of the scaling, translation and rotation parameters. The scaling transformation refers to the alteration of the camera's zoom level, either by magnifying or reducing the visual field. Rotation, on the other hand, involves the camera's movement in a clockwise or counter clockwise direction. Lastly, translation pertains to the horizontal or vertical displacement of the camera, either to the left or right, or upward or below.

$$\frac{d_1}{d_1^{\,1}} = \frac{d_2}{d_2^{\,1}} = r \tag{1}$$

where d_1 represents the distance between $A(x_1, y_1)$ and $B(x_2, y_2)$, $d_1^{\,1}$ represents the distance between $A'(x_1', y_1')$ and $B'(x_2', y_2')$, $d_2^{\,1}$ represents the distance between $A'(x_1', y_1')$ and $C'(x_3', y_3')$ and d_2 represents the distance between $A(x_1, y_1)$ and $C(x_3, y_3)$.

$$\theta = tan^{-1}\left[\frac{x_2' - x_1' - a(y_2 - y_1)}{x_2 - x_1}\right] \tag{2}$$

$$a = \frac{x_2x_3' - x_2x_1' - x_1x_3' - x_2'x_3 - x_1x_2' - x_3x_1'}{x_2y_3 - x_2y_1 - x_1y_3 - x_3y_1 - x_1y_2 - x_3y_1} \tag{3}$$

$$h = x_1 cos\theta - y_1 sin\theta - x_1' \tag{4}$$

$$K = y_1 cos\theta + x_1 sin\theta - y_1' \tag{5}$$

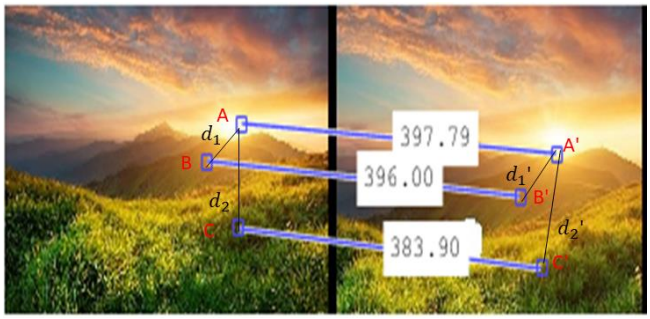


Fig. 2. Detection of key points with SURF algorithm

In figure 2, three-line segments are drawn to detect features and denoted as AA^1, BB^1, CC^1 respectively. As an example, we will choose a set of matched key points $A(x_1, y_1)$ and $A'(x_1', y_1')$. Subsequently, the distances between these locations can be calculated and other corresponding important points, namely $B(x_2, y_2)$ and $B'(x_2', y_2')$, $C(x_3, y_3)$ and $C'(x_3', y_3')$. The distances are represented as d_1 and d_1', d_2 and d_2' respectively. The rotation transformation function is subsequently utilized to compute

the angle of rotation. The symbol θ will be used to denote the angle of rotation between the restored image and the reference image. The variable *h* represents the translation along the x-axis, whereas the variable *k* represents the translation along the y-axis. When the value of *r* is less than 1, it signifies a magnification effect or zoom-in. In the event that the value of r exceeds 1, an observable zoom-out effect occurs. In all other cases, the zoom factor remains constant. Once the zoom factor has been ascertained, it becomes apparent that the appropriate course of action is to adjust the dimensions of the pertinent image by this factor.

The SURF algorithm can be broken down into the following sequential steps:

- identify the corresponding key points in the repaired image and the relevant image with the SURF algorithm, as mentioned in figure 1,
- determine the zoom factor, denoted as *r*, between the repaired and relevant image using equation (1),
- the rescale procedure should be executed on the pertinent image. The rescaling operation is performed using a scaling factor denoted as *r*,
- utilize the SURF algorithm to identify the corresponding key points between a rescaled relevant image and a repaired image.
- in order to determine the transformation parameters, namely θ, *h*, and *k*, the matched key points obtained in step 4 should be utilized in conjunction with the transformation functions in equations (2), (4), & (5),
- the subsequent stage involves the implementation of the inverse transformation. For every pixel within the designated region that requires restoration in the damaged image, perform the following task,
  1) determine the associated *x* and *y* coordinates in the relevant image by employing the inverse transformation function in equation (2),
  2) replace the damaged image pixel with the intensity values found at the new coordinates in the corresponding image obtained in the previous step.

The transformation parameters, denoted as *r*, *h*, and *k*, are determined based on the matched feature points and the utilization of transformation functions in equations (2), (3), (4) & (5). The specific values of these parameters are as follows: 1.008 for *r*, 0° for θ, 357 for *h*, and 20 for *k*. The resizing factor, represented as *r*, is applied to the respective image. The identification of the missing region inside the resized pertinent image can be accomplished by utilizing the inverse transformation function's parameters, namely θ, *k*, and *h*. The process involves identifying the corresponding region in the damaged image and transferring the intensity values from the appropriately scaled relevant image to fill in the missing information.

3. Results and discussion

The outcomes derived from our implementation exhibit promise. Assessing the quality of repair poses a multifaceted challenge. The currently employed metrics, such as Mean Squared Error (MSE) and Peak Signal-to Noise Ratio (PSNR) are straightforward to implement. The computation of the PSNR index involves the utilization of both the original and restored image. By measuring the peak signal strength to the noise strength, in decibels, the PSNR can be utilized as a measure of image quality. The ratio indicated above is commonly used as a criterion for evaluating the quality of an image in relation to its compressed version. A greater PSNR indicates a superior level of quality in the compressed or reconstructed image. The concept of MSE pertains to the average of the squared differences between calculated values and real values.

In the aforementioned cases, distinct values such as the PSNR, MSE were computed. The exemplar-based method and SURF algorithm are two prominent techniques utilized in the field of image inpainting. The values of Signal-to-Noise Ratio and mean square error are 62.8 dB and 0.03 respectively. In this proposed inpainting method, signal ratio is enhanced by 2.5 times. Figure 6 display the input image with visible scratches, followed by the image following the process of exemplar based inpainting.

Fig. 3. Relevant Image

Fig. 4. Damaged Image

Fig. 5. Output of Proposed method (SURF algorithm)

Table 1. Performance Analysis of different Image inpainting Techniques

	PSNR (dB)	MSE
Exemplar method	27	126
SURF extraction method	62,8	0,03

Fig. 6. Exemplar based image inpainting [13]

The objective of the studies is to validate the efficacy of the suggested method and to conduct a comparative analysis with the existing exemplar-based method [10]. The implementation was conducted using MATLAB on a central processing unit (CPU) with an Intel i3 core operating at a clock speed of 2.10 GHz and equipped with 7 GB of random-access memory (RAM). To evaluate the efficacy of the suggested method, the PSNR, MSE is employed as the metrics for performance assessment. The findings of our study indicate that our system yields superior outcomes. The preservation of visual quality is ensured.

4. Conclusion

The method of repairing damaged photos using SURF feature extraction has demonstrated success in the restoration of images that have incurred damage or have been removed. The aforementioned findings demonstrate that the utilization of this approach is a proficient means of image restoration, as it minimizes the loss of the original image's quality to a significant extent. The present study has showcased the efficacy of the Speed Up Robust Features method in the restoration of impaired photos, hence establishing its viability as a dependable tool for digital image repair. The experimental results have shown that our methodology is capable of achieving outcomes that are equivalent to other methodologies that are now considered to be at the forefront of the field [10].

References

[1] Bay H., Tuytelaars T., Van Gool L.: Surf: Speeded up robust features. 9th European Conference on Computer Vision–ECCV 2006, Austria, 2006.
[2] Bertalmio M. et al.: Image inpainting. Proceedings of the 27th annual conference on Computer graphics and interactive techniques, 2000.
[3] Bertalmio M. et al.: Simultaneous structure and texture image inpainting. IEEE transactions on image processing 12(8), 2003, 882–889.
[4] Birajdar G. K., Vijay H. M.: Digital image forgery detection using passive techniques: A survey. Digital investigation 10(3), 2013, 226–245.
[5] Cheng W. H. et al.: Robust algorithm for exemplar-based image inpainting. Proceedings of International Conference on Computer Graphics, Imaging and Visualization, 2005.
[6] Criminisi A., Pérez P., Toyama K.: Region filling and object removal by exemplar-based image inpainting. IEEE Transactions on image processing 13(9), 2004, 1200–1212.
[7] Drori I., Cohen-Or D., Yeshurun H.: Fragment-based image completion. ACM SIGGRAPH 2003 Papers, 2003, 303–312.
[8] Jia J., Tang C. K.: Image repairing: Robust image synthesis by adaptive and tensor voting. IEEE Computer Society Conference on Computer Vision and Pattern Recognition 1, 2003.
[9] Kakar P.: Passive approaches for digital image forgery detection. Thesis, 2015.
[10] Prasanthi T. S. et al.: Performance analysis of different applications of image inpainting based on exemplar technique. International Journal on Recent and Innovation Trends in Computing and Communication 11(4), 2023, 113–117.
[11] Wong A., Orchard J.: A nonlocal-means approach to exemplar-based inpainting. 15th IEEE International Conference on Image Processing, 2008.
[12] Wu J., Ruan Q.: Object removal by cross isophotes exemplar-based inpainting. 18th International Conference on Pattern Recognition (ICPR'06) 3, 2006.
[13] Yogesh Laxman Tonape V.: Faster and Efficient Method for Robust Exemplar Based Inpainting Using Block Processing. International Journal of Computer Science and Information Technologies 6(3), 2015.

M.Sc. Shanmukhaprasanthi Tammineni
e-mail: prashanthitammineni.rs@andhrauniversity.edu.in

She obtained M.Tech. degree in Electronics and Communication Engineering from Jawaharlal Nehru Technological University Kakinada, Andhra Pradesh in 2021. She is currently pursuing Ph.D. degree with Andhra University Visakhapatnam, India. Her research interests include microstrip patch antenna design, VLSI circuit design, image inpainting in image processing.

https://orcid.org/0009-0000-5352-2265

M.Sc. Swarajya Madhuri Rayavarapu
e-mail: madhurirayavarapu.rs@andhrauniversity.edu.in

Currently pursuing Ph.D. in Department of Electronics and Communication, Andhra University. She obtained her M.Tech. degree from CASEST, University of Hyderabad.
Her research interests include deep learning, generative adversarial networks (semi-supervised machine learning) in medical image processing, applying deep learning techniques to 5G-mobile communication (Layer 2 of RAN).

https://orcid.org/0009-0007-7559-2142

Prof. Sasibhushana Rao Gottapu
e-mail: sasigps@gmail.com

He is senior professor in the Department of Electronics & Communication Engineering, Andhra University College of Engineering, Visakhapatnam, India. He is a senior member of IEEE, fellow of IETE, member of IEEE communication Society, Indian Geophysical Union (IGU) and International Global Navigation Satellite System (IGNSS), Australia. Prof. Rao was also the Indian member in the International Civil Aviation organization (ICAO), Canada working group for developing SARPS.

https://orcid.org/0000-0001-6346-8274

Prof. Raj Kumar Goswami
e-mail: rajkumargoswami@gmail.com

He completed his M.Tech. in Radar and Communications from IIT Delhi and subsequently, continued in IIT Delhi to work on a data link project, where he designed and developed the algorithm for RF modem and implemented the same on floating point Digital Signal Processor SHARC, ADSP 21060. Dr. R.K. Goswami subsequently, completed his Ph.D. in Electronic and Communication Engineering from Andhra University during which he designed the forward error correction schemes in respect of multipath channel. His areas of interests include computer networks, signal processing, image processing, artificial intelligence and software engineering.

https://orcid.org/0000-0002-0651-6783

http://doi.org/10.35784/iapgos.5494 *received: 22.10.2023 | revised: 11.02.2024 | accepted: 03.03.2024 | available online: 31.03.2024*

TENSOR AND VECTOR APPROACHES TO OBJECTS RECOGNITION BY INVERSE FEATURE FILTERS

Roman Kvyetnyy[1], Yuriy Bunyak[2], Olga Sofina[1], Volodymyr Kotsiubynskyi[1], Tetiana Piliavoz[1], Olena Stoliarenko[2], Saule Kumargazhanova[3]

[1]Vinnitsia National Technical University, Vinnitsia , Ukraine, [2]Spilna Sprava company, Vinnitsia, Ukraine, [2]Vinnytsia Mykhailo Kotsiubynskyi State Pedagogical University, Vinnitsia, Ukraine, [3]D. Serikbayev East Kazakhstan Technical University, Ust-Kamenogorsk, Kazakhstan

Abstract. *The investigation of the extraction of image objects features by filters based on tensor and vector data presentation is considered. The tensor data is obtained as a sum of rank-one tensors, given by the tensor product of the vector of lexicographic representation of image fragments pixels with itself. The accumulated tensor is approximated by one rank tensor obtained using singular values decomposition. It has been shown that the main vector of the decomposition can be considered as the object feature vector. The vector data is obtained by accumulating analogous vectors of image fragments pixels. The accumulated vector is also considered as an object feature. The filter banks of a set of objects are obtained by regularized inversion of the matrices compiled by object features vectors. Optimized regularization of the inversion is used to expand the regions of object features capture with minimal error. The object fragments and corresponding feature vectors are selected through a training iterative process. The tensor and vector approaches create two channels for recognition. High efficiency of object recognition can be achieved by choosing the filter capture band and creating filter branches according to the given bands. The filters create a convolutional network to recognize a set of objects. It has been shown that the obtained filters have an advantage over known correlation filters when recognizing objects with small fragments.*

Keywords: objects recognition, objects feature, image data tensor, image data vector, inverse filters, optimized regularization

PODEJŚCIE TENSOROWE I WEKTOROWE DO ROZPOZNAWANIA OBIEKTÓW ZA POMOCĄ FILTRÓW CECH ODWROTNYCH

Streszczenie. *Rozważane jest badanie ekstrakcji cech obiektów obrazu przez filtry oparte na prezentacji danych tensorowych i wektorowych. Dane tensorowe uzyskuje się jako sumę tensorów* pierwszego *rzędu,* otrzymanych *przez iloczyn tensorowy wektora leksykograficznej reprezentacji pikseli fragmentów obrazu z samym sobą. Skumulowany tensor jest aproksymowany przez tensor* pierwszego *rzędu uzyskany przy użyciu dekompozycji wartości osobliwych. Wykazano, że główny wektor dekompozycji można uznać za wektor cech obiektu. Dane wektorowe uzyskuje się poprzez akumulację analogicznych wektorów pikseli fragmentów obrazu. Skumulowany wektor jest również uważany za cechę obiektu. Banki filtrów zestawu obiektów są uzyskiwane przez regularyzowaną inwersję macierzy skompilowanych przez wektory cech obiektów. Zoptymalizowana regularyzacja inwersji jest wykorzystywana do rozszerzenia obszarów przechwytywania cech obiektów przy minimalnym błędzie. Fragmenty obiektów i odpowiadające im wektory cech są wybierane w iteracyjnym procesie uczenia. Podejście tensorowe i wektorowe tworzy dwa kanały rozpoznawania. Wysoką skuteczność rozpoznawania obiektów można osiągnąć, wybierając pasmo przechwytywania filtrów i tworząc gałęzie filtrów zgodnie z podanymi pasmami. Filtry tworzą sieć konwolucyjną do rozpoznawania zestawu obiektów. Wykazano, że uzyskane filtry mają przewagę nad znanymi filtrami korelacyjnymi podczas rozpoznawania obiektów z małymi fragmentami.*

Słowa kluczowe: rozpoznawanie obiektów, cechy obiektów, tensor danych obrazu, wektor danych obrazu, filtry odwrotne, zoptymalizowana regularyzacja

Introduction

Tensors are known as multidimensional matrices for coordinate transforms [5]. Recently, they have been used for the description of image object arrays in computer vision and machine learning [3, 6, 10, 13]. Tensors are considered as multidimensional matrices with structural features. These features allow to reduce the dimension or size of a multidimensional data matrix by decomposition into a product of a sequence of tensors of lower dimension and smaller size. The tensors can be approximated by tensors of uncompleted rank to obtain conditions for decomposition. The items of the decomposition can be decomposed in such a way as to create a tensor train [4, 6, 10, 12, 13]. The tensors of incomplete rank are obtained using the tensor or outer product of vectors and matrices [4, 9]. The Singular Value Decomposition (SVD) [20] is used as the tool to decompose matrices and approximate them by the sum of tensor products of orthogonal vectors-columns of two unitary matrices. The number of terms in the sum is defined by the rank of the approximating matrix.

The geometry means of a tensor [5] were modified for binary images as a shape salience detector and a shape descriptor – tensor scale descriptor with influence zones [1, 2]. It is a robust method to compute tensor scale, using a graph-based approach.

The feature of the tensor object can be associated with the core tensor [16] of Tucker decomposition [14, 18]. The decomposition represents the multidimensional matrix as a product of a core matrix and matrices of factors. It allows to reduce a complexity of the problem of objects classification using a least squares approach. Tucker decomposition can be implemented by hierarchical SVD to achieve the required accuracy of the initial tensor approximation [8]. The Canonical Polyadic Decomposition (CPD) [5, 7, 9] represents the tensor as a sum of rank one tensors given by the tensor product of vectors that correspond to each of the dimensions, at least two. One such decomposition is the SVD and its high order extension [3, 13].

The problem of objects recognition is solved by feature filters. The filters create a Convolutional Neural Network (CNN) for recognizing some class of objects. The use of the least squares method for filters creation leads to equations based on the correlation dependencies of the data and the desired results [19]. The simple vector representation of image data is used for this problem.

Multidimensional correlation analysis can be defined for two tensors in a space of some dimensions. The inner product of these dimensions yields the correlation tensor [3, 8]. Correlation tensor analysis, in conjunction with discriminant analysis, is capable of capturing higher order structures in the data patterns by encoding each object as a second- or higher-order tensor. This allows for obtaining a low-dimensional data representation that reflects both class label information and intrinsic manifold structure.

The considered properties of tensors motivate the investigation of the application of the tensor structure of image data in object recognition using correlation filters [7, 11, 19].

1. The aim and objectives of the study

Tensor characteristics of a certain set of objects in the form of images are interesting in the case when each of the objects has its own core feature tensor, and the system of factor tensors is common. The collection of core tensors and factor tensors approximates the image set. Then, by extracting and classifying features from the image, it is possible to recognize the objects contained in it. Such a tensor construction can be created artificially using image data.

The simplest tensor construction in the form of CPD is used for creating a filter bank for object recognition based on Inverse Regularized Feature Filters (IRFF) [11]. Its performance is compared with similar filter banks obtained by the methods considered in [7, 11, 19] using lexicographical vector presentation of object images.

The research method consists of simulating the process of training the filter bank with further filtering of image data in order to recognize objects. Alphabetic and numeric characters with variations in size and perspective were used as a test set for recognizing objects.

The investigation of the influence of the form of data representation – tensor and vector – on the selective properties of the IRFF of image objects is presented.

2. Methods

The problem of object recognition includes the following three main stages: object' feature extraction, feature selection, classifier design. The object' feature characteristics must satisfy such a requirement as stability, it means to be invariant to object variations in order to recognize as many objects of the same type as possible. The fulfilment of this requirement is achieved by training the recognition method using a large number of templates of the target object. The templates arrays of some target objects create multidimensional data structure which can be considered as a tensor. A feature finding and extraction in the form of matrix or vector can be done by tensor decomposition, by lexicographic vectorization over several dimensions and reduction to a matrix form accessible for processing, or by folding dimensions using integration over some dimensions by summation over corresponding indexes [3, 5, 9, 12, 13].

A. Tensor approach

The problem is to recognize a set of image objects. The images are presented by arrays of templates to train for the recognizing. Each of the templates consists of fragments which include desired object features or not. The problem is to select fragments with the object features to create the features filters.

The problem data can be presented as the 7-th order tensor $\mathbf{X}_{i_1,i_2,i_3,i_4,i_5,i_6,i_7}$, where $i_1 = 0....M-1$ is the objects index, $i_2 = 0....N_{T_{i_1}} - 1$ – the templates index of i_1-th object, $i_3, i_4 = 0....N_{F_{i_1,i_2}} - 1$ – i_2-th template of i_1-th object indices, $i_5, i_6 = 0....N-1$ – indices of pixels of fragments of size $N \times N$, $i_7 = 0,\ 1\ ,2$ – index of RGB colors. The solution of the object recognition problem depends on the complexity of the tensor $\mathbf{X}$ data structure. If the tensor includes repeating or cyclic elements, which allows it to be decomposed into outer products of simpler tensors, then this reduces the complexity of the problem and improves recognition accuracy by reducing the number of recognition parameters and their separability by objects. The complexity of the data structure can be characterized by the rank of the tensor. Tensors of incomplete rank lend themselves to tensor decompositions [4, 6, 13]. Tensor $\mathbf{X}$ order can be lowered if to join three last indices by lexicographic representation of image matrices of pixels as a vector, then

$$vec\left(\mathbf{X}_{i_1,i_2,i_3,i_4,i_5,i_6,i_7}\right) \sim \mathbf{X}_{i_1,i_2,i_3,i_4,i=0..3N^2} \tag{1}$$

The target of training process is to obtain a filters bank $F_{i_1,i}$ which classify an image fragments I_i as i_1-th object or not by the next rule:

$$\underset{i_1}{maxarg} \sum_i F_{i_1,i} \cdot I_i = \sigma_{i_1} \begin{cases} \geq threshold : i_1\ true \\ < threshold : i_1\ false \end{cases} \tag{2}$$

The initial filter of k-th object can be found from the equation

$$\sum_i F_{i_1,i} \cdot X_{k,i_2,i_3,i_4,i} = \delta_{i_1,k} 1_{i_2,i_3,i_4}, \tag{3}$$

where $\delta_{i_1,k} = 1:\ i_1 = k;\ \delta_{i_1,k} = 0:\ i_1 \neq k = 1$. The problem statement (3) corresponds to zero-aliasing approach of correlation filters design [7, 19]. The least square solution of the equation (3) is the next:

$$\sum_i F_{i_1,i} \sum_{i_2,i_3,i_4} X_{k,i_2,i_3,i_4,i} X_{k,i_2,i_3,i_4,j} = \delta_{i_1,k} \sum_{i_2,i_3,i_4} X_{k,i_2,i_3,i_4,j} \tag{4}$$

The sum over three indices in the left side of expression (4) is the sum of the matrices of the rank one obtained using the tensor type or the outer product of the image fragment vector by itself. Let denote it as $R_k = \left[R_{k,i,j}\right]_{i,j=0,...,3N^2}$, it has size $3N^2 \times 3N^2$ for k-th object. This sum R_k has rank one if all fragments are mutually collinear, that is, equal up to a factor. Differences in the vectors of fragments form a matrix R_k of higher rank, and with a large number of terms, a matrix of full rank. However, if the problem is to find the feature of each of the set of objects and recognize objects by filtering these features, then, if to assume that each fragment of the object's image templates is the feature with deviations, then this feature should be repeated in the matrix and form a total matrix of rank one. As the matrix R_k is symmetric, its SVD is the follows.

$$\left[R_{k,i,j} = \sum_{m:s_m^{(k)}>0} s_m^{(k)} u_{i,m}^{(k)} u_{j,m}^{(k)} \right]_{i,j=0,...,3N^2} \tag{5}$$

where the unitary matrix $U^{(k)} = \left[u_{i,j}^{(k)}\right]_{i,j=0,...,3N^2}$ of size $3N^2 \times 3N^2$ consists of orthogonal columns, singular values $s_0^{(k)} > s_1^{(k)} > ... \geq 0$. The number of nonzero singular values is equal to the rank of matrix (5). The SVD (5) gives an approximation of the matrix R_k by the rank one matrix as

$$\left[\tilde{R}_{k,i,j} = s_0^{(k)} u_{i,0}^{(k)} u_{j,0}^{(k)}\right]_{i,j=0...3N^2-1} \tag{6}$$

Therefore, the vector

$$\Phi_k = \sqrt{s_0^{(k)}}\, N_{TF_k}^{-1} \left[u_{i,0}^{(k)}\right]_{i=0...3N^2-1} \tag{7}$$

where N_{TF_k} is total number of fragments of k-th object' templates array, is the feature vector of the object. The equation (3) for feature vector (7) can be written as

$$\sum_i F_{i_1,i} \Phi_{k,i} = \delta_{i_1,k} \tag{8}$$

As it follows from (8), the bank of M filters can be defined as the matrix inverse to matrix compelled by M vectors (7):

$$\left[F_{i_1,i}\right] = \left[\Phi_{i_1,i}\right]_{i_1=0...M-1,i=0...3N^2-1}^{-1} \tag{9}$$

The feature vectors are defined on the base of SVD of the total sum of all fragments in (4). The second step of the filter bank (9) creation is basing on the training process that is the selection by operation (2) that fragments which give true result. Also, the fragments which give false result are selected too. The sum of selected fragments gives two matrices $R_{i_1}^{tr}$ and $R_{i_1}^{fl}$. The SVD of the matrices gives feature vectors (7) $\Phi_{i_1}^{tr}$ and $\Phi_{i_1}^{fl}$ of the extended filters bank

$$\left[F_{i_1,i}^{tr}\ F_{i_1,i}^{fl}\right] = \left[\Phi_{i_1,i}^{tr}\ \Phi_{i_1,i}^{fl}\right]_{i_1=0...M-1,i=0...3N^2-1}^{-1} \tag{10}$$

The using of filters bank (10) gives new arrays of fragments which include features of true and false results in (2). The procedure of features extraction and filters bank (10) creation can be repeated up to the desired result in terms of the ratio of correct and incorrect recognitions results will not be achieved.

B. Vector approach

The convergence of the iterative scheme for determining the filter bank (10) to a given level of recognition errors depends on the structure of the image data, which effects on the ability to approximate matrices $R_{i_1}^{tr}$ and $R_{i_1}^{fl}$ by matrices of unit rank. The feature vector is the result of the projection of the data vectors onto one of the vectors of the vector space of the SVD. Therefore, each of the data vectors can only be partially considered when features are determining.

Let's consider a scheme for generating feature vectors and forming a filter bank in such a way that the data of templates fragments are fully considered. The equation (3) can be rewritten in the manner

$$\sum_i F_{i_1,i} \cdot \frac{1}{N_{TFk}} \sum_{i_2,i_3,i_4} \mathbf{x}_{k,i_2,i_3,i_4,i} = \delta_{i_1,k} \tag{11}$$

where

$$\left|\mathbf{x}_{k,i_2,i_3,i_4,i} = vec_{i=i_5,i_6,i_7}\left(X_{k,i_2,i_3,i_4,i_5,i_6,i_7}\right)\right|_{i=0...3N^2-1}$$

i is the index of lexicographic representation of items indexed by i_5, i_6, i_7. The sum over three indices can be denoted as $\overline{\mathbf{x}}_k$. The feature vector is the average image data vectors obtained by lexicographic representation of image frames matrices of pixels in the case of scheme (11), that is,

$$\Psi_{i_1} = N_{TFi_1}^{-1}\left[\overline{x}_{i_1,i}\right]_{i=0...3N^2-1} \tag{12}$$

As it follows from condition (11), it can be obtained the filter bank for feature vectors (12) by the same iterative procedure as the filters bank (10).

$$\left[H_{i_1,i}^{tr}\ H_{i_1,i}^{fl}\right] = \left[\Psi_{i_1,i}^{tr}\ \Psi_{i_1,i}^{fl}\right]^{-1}_{i_1=0...M-1, i=0...3N^2-1} \tag{13}$$

C. Filters bank optimization

The inverse filters like (9) and (13) were investigated in [11]. The experiments have shown that filters on the base of exact inversion are narrow and select only objects which vary from the feature vectors by small values at the level of computational error. The inverse filters' band of selection can be expanded by regularization of the inversion. The SVD of features matrix for inversion in (10) with account regularization parameter λ is the follows.

$$\left[\Phi_{i_1,i}^{tr}\ \Phi_{i_1,i}^{fl}\right]_\lambda = \sum_m u_{\Phi i,m}(s_{\Phi m} + \lambda)\left[v_{\Phi i_1,m}^{tr}\ v_{\Phi i_1,m}^{fl}\right] \tag{14}$$

The inversion of (14) to define the filters bank (9) yields the next expression.

$$\left[F_{i_1,i}^{tr}\ F_{i_1,i}^{fl}\right]_\lambda = \sum_m \left[v_{\Phi i_1,m}^{tr}\ v_{\Phi i_1,m}^{fl}\right] u_{\Phi i,m}\left(s_{\Phi m} + \lambda\right)^{-1} \tag{15}$$

The regularization in (15) degrades the separation ability of the filter (9). The separation ability will be high if mutual relation of the feature filters

$$\sum_m F_{i,m} \cdot F_{k,m} \approx \delta_{i,k}$$

is minimal as possible at $i \neq k$. This condition demands growth of λ in (15). At the same time, the energy of the regularized feature matrix (14) increases indefinitely under the growth of λ. Therefore, an optimal regularization should balance these two mutually contradictory conditions. The condition of the balance for matrices of 2*M* features vectors (14) and vector of filters (15) may be presented by the functional [11]

$$\mathrm{I}(\lambda) = \underset{\lambda}{argmin}\left\{\frac{\sum_i \left|\Phi_i\right|_\lambda^2}{\sum_i \left|\Phi_i\right|_0^2} + \frac{\sum_{i,k;i\neq k}\left|\sum_m F_{i,m}F_{k,m}\right|_\lambda^2}{\sum_{i,k;i\neq k}\left|\sum_m F_{i,m}F_{k,m}\right|_0^2}\right\} \tag{16}$$

The matrices in the left side of (14) and (15) are incompatible because they are mutually inverse, therefore the normalization of square norms of the matrices by initial unperturbed norms was used. The condition $\partial \mathrm{I}(\lambda)/\partial\lambda = 0$ yields the following equation in terms of matrix (14) SVD.

$$-\sum_{i;i\neq k}\left|\sum_m \tilde{F}_{i,m}\tilde{F}_{k,m}\right|_0^{-2} \sum_{i,k:i\neq k}\sum_{m=0}\frac{v_{i,m}v_{k,m}}{(s_m+\lambda)^3} + \tag{17}$$

$$+\sum_i\left|\Phi_i\right|_0^{-2}\sum_{i,k}\sum_m v_{i,m}v_{k,m}(s_m+\lambda) = 0,$$

where $v_{i,m}$ are elements of joined matrices $\left[v_{\Phi i_i,i}^{tr}\ v_{\Phi i_i,i}^{fl}\right]$. The closest to zero value of the expression (17) and corresponding to it λ_{opt} can be found by iterative way numerically.

The optimization procedure for filters bank (13) is similar. The filters banks (9) and (13) are not identical and therefore can be used as two channels of recognizing. The two channels allow to expand the range of capturing desired objects and increase a number of recognitions, which makes it possible to improve the accuracy of recognition by analysing the distribution statistics of recognized objects and selection of most probable ones.

3. Experimental analysis

Analysis of an effectivity of the IRFF on the base of tensor and vector approaches was made using 2600 images of 35 characters of the size from 13×13 to 41×41 pixels, cropped from car license plate images. Up to 200 thousand fragments with the size 13×13 were used for recognition. Methods of errors elimination considered in [11] were used, such as: spectrums combination; amplitude selection; filters brunching by recognized groups of objects.

A. Errors elimination

Initially, the recognition by spectrum (2) in the basis of filters bank (15) is not effective, the number of errors is slightly less then number of successful recognitions even after completed training by selection of the feature true and false frames.

The true features vectors $\Phi_{i_1}^{tr}$ in (14) can be divided on some parts. The filters (15) subbank can be defined for each part. Since the operation of inversion (15) is ill-posed and regularized its results are ambiguous for matrices compiled by different number of columns $\Phi_{i_1}^{tr}$. Therefore, the filters of the full bank (15) will differ from subbanks' filters. This feature allows to check the correctness of the object spectrum (2) estimation by combining the spectrum of filter bank (15) with the spectrum in the corresponding subbanks. If the spectrum matches according to the assessment of the object, then such the assessment is assumed to be correct. Five subbanks by seven filters and seven subbanks by five filters were used. Thus, the evaluation of the object using expression (2) was carried out by combining the results for the filter bank and two subbanks. This procedure allowed to reduce the number of false recognitions by 20-30%.

Let the fragment $\mathbf{x}_{k,i_2,i_3,i_4,i}$ of *k*-th object has the spectrum (2)

$$\sum_i F_{i_1,i}^{tr} \cdot \mathbf{x}_{k,i_2,i_3,i_4,i} = \sigma_{i_1,i_2,i_3,i_4} \tag{18}$$

which is recognized as *j*-th object in the training process. Such mistake was caused by the form of the spectrum (18) in the basis of filters bank. The spectrum can be accumulated by averaging as the matrix $\left[\varsigma_{i,j,m}\right]_{i,j,m=0..M-1}$. The first index relates with the current object, second index with recognized object and third with *M* spectral samples (2), (18). This matrix fixes the cases when different objects can include similar fragments. Such frames

should be eliminated. The true spectrum can be selected by finding the minimal spectral distance. If spectrum (2) points on the *i*-th object and

$$\min_{k} \arg \sum_{m=0}^{M-1} \left| \sigma_m - \varsigma_{i,k,m} \right| = i$$

then object is recognized as true, otherwise as false.

The purpose of the training process is to obtain true recognition with minimal error level. Not all objects templates will be recognized by created filters basis. The objects which are recognized create the branch of the recognizing. Remained templates can be used for obtaining the next branches while all templates have been recognized. Each true object recognition can be in a single branch. This is a criterion to eliminating errors as well.

B. Dynamic of the filters training

The dynamic of the training process for filters banks of Tensor IRFF, TIRFF (9), vector IRFF, VIRFF (13), and joined two channels of IRFF banks (JIRFF) are presented in table 1. The number of selected as features fragments, the number of true and false recognitions are shown by three corresponding rows for some numbers of training steps.

As it can be seen from the table, the training processes are stable and gradually improve the recognition result. However, the accuracy of using the filters (9) is somewhat lower comparing to using of the filters (13) for the same number of training steps. The reason for this is the fact that feature vectors (7) are projections (6) of the total data (5) onto the main SVD vector. The use of several SVD vectors for data approximation in (6) deteriorate the recognition result – the range of captured objects narrows and the rating of correctly recognized objects also decreases. The range of captured objects by a two-channel filters narrows in the process of training, as well as for filters (9) and (13), which shows that sets of selected features by filters (9) and (13) gradually converges.

Table 1. Dynamic of training process for TIRFF, VIRFF, JIRFF

Training step	Feature fragments	True fragments	False fragments	True rating
0	1462 1163 1783	1043 868 1291	419 295 492	0.7134 0.7463 0.7240
20	1026 943 1446	852 845 1228	174 98 218	0.8304 0.8960 0.8492
40	960 864 1267	819 805 1137	141 59 130	0.8531 0.9317 0.8974
60	920 839 1187	810 791 1090	110 48 97	0.8804 0.9427 0.9182
80	841 806 1050	761 768 977	80 38 73	0.9048 0.9528 0.9301
100	774 794 945	713 760 893	61 34 52	0.9211 0.9571 0.9452

Table 2. Comparison of IRFF with Correlation Filters

Filter type	Feature fragments	True fragments	False fragments	True rating
TIRFF	758	727	30	0.9591
VIRFF JIRFF MECF	782 912 1285	759 878 851	23 34 434	0.9705 0.9627 0.6622
ZACF	1064	923	141	0.8674

The results of recognition by TIRFF, VIRFF, TIRFF and joined two channel JIRFF, obtained as the result of 200 optimization steps, are presented in Table 2. The residual error depends on the bandwidth of the objects captured by the filters. It can be eliminated by decreasing the regularization parameter in (15). Its optimal assessment value is $\lambda_{opt} \sim (0.007...0.012) \cdot s_0$. The results of recognition by Minimum Energy Correlation Filter (MECF) [19] and Zero-Aliasing Correlation Filter (ZACF) [7] are shown in Table 2 for comparison. The IRFFs have the highest selectivity, the number of selected fragments is the smallest, this can explain the smallest number of errors of these filters.

4. Conclusions

The tensor (3) and vector (11) forms of the problem to define features and filter object sets are considered. The tensor form is based on the approximation of the image's template set data matrix by a matrix of unit rank. The data matrix is obtained as the sum of matrices of unit rank given by the tensor product of the vectors of lexicographic representation of image pixels with itself. The vector form is obtained by the accumulation of the analogous vectors (12). The tensor matrix is symmetrical and can be presented as the tensor product of vectors (7). These vectors are the feature vectors. They allow the creation of filters and the recognition of the image object.

The regularized inversion (15) of the matrix compiled by feature vectors as columns in (9) yields the feature filter bank for recognizing the set of objects with given features. The regularization level defines the selectivity property of the filters – the range of objects captured and the level of error.

The feature vectors are selected by a training iterative process. The two approaches considered create two channels of recognition because they are not identical in their selectivity properties.

The inverse regularized feature filters based on tensor and vector forms allow for the selection of image fragments corresponding to desired objects with regulated accuracy. It is possible to achieve high efficiency in object recognition by choosing a capture range for the filter and creating filter branches according to the given ranges.

The known correlation filters [7, 19] have high recognition ability when the full object is captured and foreign objects are absent. We considered an approach to recognize objects by fragments or by some feature fragments. The IRFF can be used in CNN for recognizing complex objects by filtering some branches of their fragments.

References

[1] Andaló F. A. et al.: Shape feature extraction and description based on tensor scale. Pattern Recognition 43(1), 2010, 26–36 [https://doi.org/10.1016/j.patcog.2009.06.012].
[2] Avrunin O. G. et al.: Features of image segmentation of the upper respiratory tract for planning of rhinosurgical surgery. 39th International Conference on Electronics and Nanotechnology, ELNANO 2019, 485–488.
[3] Deng Y., Tang X., Qu A.: Correlation Tensor Decomposition and Its Application in Spatial Imaging Data. J. of the American Statistical Association 118(541), 2023, 440–456 [https://doi.org/10.1080/01621459.2021.1938083].
[4] De Lathauwer L.: Signal Processing based on Multilinear Algebra. PhD thesis, Katholieke Universiteit Leuven, 1997.
[5] Dubrovin B. A., Fomenko A. T., Novikov S. P.: Modern Geometry – Methods and Applications Pt. 1. Springer, New York 1992.
[6] Comon P.: Tensor decomposition: State of the art and applications. V. J. G. McWhirter, I. K. Proudler (eds): Mathematics in Signal Processing, Oxford University Press, Oxford 2002.
[7] Fernandez J. A. et al.: Zero-Aliasing Correlation Filters for Object Recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence 37(8), 2015, 1702–1715.
[8] Fu Y., Huang T. S.: Image Classification Using Correlation Tensor Analysis. IEEE Trans on Image Processing 17(2), 2008, 226–234.
[9] Grasedyck L.: Hierarchical Singular Value Decomposition of Tensors. SIAM Journal on Matrix Analysis and Applications 31(4), 2010 2029–2054 [https://doi.org/10.1137/090764189].
[10] Kolda T. G., Bader B. W.: Tensor decompositions and applications. SIAM Rev. 51, 2009, 455–500.

[11] Kvyetnyy R. et al.: Inverse correlation filters of objects features with optimized regularization for image processing. Proc. SPIE 12476, 2022, 124760Q [https://doi.org/10.1117/12.2664497].
[12] Orazayeva A. et al.: Biomedical image segmentation method based on contour preparation. Proc. SPIE 12476, 2022, 1247605 [https://doi.org/10.1117/12.2657929].
[13] Oseledets I. V.: Tensor-train decomposition. SIAM Journal on Scientific Computing 33(5), 2011, 2295–2317 [https://doi.org/10.1137/090752286].
[14] Pavlov S. V.: Information Technology in Medical Diagnostics. W. Wójcik, A. Smolarz (eds), CRC Press, 2017.
[15] Panagakis Y. et al.: Tensor Methods in Computer Vision and Deep Learning. Proceedings of the IEEE 105(5), 2021, 863–890 [https://doi.org/10.1109/JPROC.2021.3074329].
[16] Phan A. H., Cichocki A.: Tensor decompositions for feature extraction and classification of high dimensional datasets. Nonlinear Theory and Its Applications IEICE 1(1), 2010, 37–68 [https://doi.org/10.1587/nolta.1.37].
[17] Timchenko L. I. et al.: Q-processors for real-time image processing. Proc. SPIE 11581, 2020, 115810F [https://doi.org/10.1117/12.2580230].
[18] Tucker L. R.: Some mathematical notes on three mode factor analysis. Psychometrika 31(3), 1966, 279–311 [https://doi.org/10.1007/BF02289464].
[19] Vijaya Kumar B. V. K., Mahalanobis A., Juday R. D.: Correlation pattern recognition. Cambridge University Press, Cambridge 2005.
[20] Wilkinson J. H., Reinsch C.: Handbook for Automatic Computation. Linear Algebra. Heidelberg New York, Springer Verlag, Berlin, 1974.

Prof. Roman Kvyetnyy
e-mail. rkvetny@sprava.net

Prof. of Department of Automation and Intelligent Information Technologies, Vinnytsia National Technical University. Scientific interests include modeling of complex systems and decision-making under conditions of uncertainty (probabilistic and interval methods), modern methods of data processing. The main direction of scientific activity is development of methods and tools for mathematical modeling and information processing in computerized systems of automation and control.

https://orcid.org/ 0000-0002-9192-9258

Ph.D. Yuriy Bunyak
e-mail: iuriy.buniak@gmail.com

Spilna Sprava company, Vinnytsya. The main direction of scientific activity is investigation of the methods of optimization in big data, signals and image processing – denoising, deblurring, object recognition, also using big data.

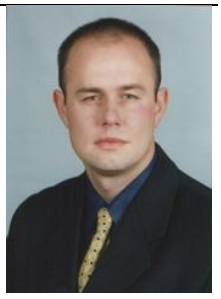
https://orcid.org/0000-0002-0862-880X

Ph.D. Olga Sofina
e-mail: olsofina@gmail.com

Ph.D., senior lecturer of Department of Automation and Intelligent Information Technologies, Vinnytsia National Technical University.
The main direction of scientific activity is modern methods of data and image processing, namely methods of filtering textured images and identifying extraneous objects on their background, as well as methods of removing blurring of the image.

https://orcid.org/0000-0003-3774-9819

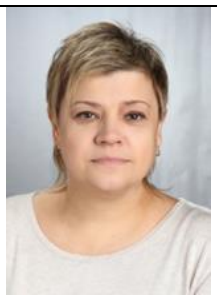
Ph.D. Volodymyr Kotsiubynskyi
e-mail: Vkotsyubinsky@gmail.com



Ph.D., senior lecturer of Department of Automation and Intelligent Information Technologies, Vinnytsia National Technical University.
The main direction of scientific activity is information technologies and management of software projects.

https://orcid.org/0000-0001-6759-5078

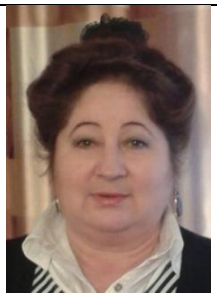
Ph.D. Tetiana Piliavoz
e-mail: pilyavoz@vntu.edu.ua



Assistant professor of the Department of Entrepreneurship, Logistics and Management, Vinnytsia National Technical University, Vinnytsia, Ukraine.
Research interests: innovation, entrepreneurship, clustering, logistics, information economic.

https://orcid.org/0000-0001-7535-7360

D.Sc. Olena Stoliarenko
E-mail: olena-best@ukr.net



Doctor of Pedagogic Sciences, professor at the Department of Pedagogics, Vinnytsia Mykhailo Kotsiubynskyi State Pedagogical University, Vinnytsia, Ukraine.
Research interests: innovation, entrepreneurship, logistics, pedagogics.

https://orcid.org/0000-0002-1899-8089

Ph.D. Saule Kumargazhanova
e-mail: SKumargazhanova@gmail.com

She is currently the dean of the Department of Information Technologies and Intelligent Systems of D. Serikbayev East Kazakhstan Technical University. She is a co-author over 50 papers in journals and conference proceedings.
Her professional interests are software engineering, data processing and analysis.

https://orcid.org/0000-0002-6744-4023

http://doi.org/10.35784/iapgos.5615 *received: 27.11.2023 | revised: 29.02.2024 | accepted: 07.03.2024 | available online: 31.03.2024*

ARCHITECTURAL AND STRUCTURAL AND FUNCTIONAL FEATURES OF THE ORGANIZATION OF PARALLEL-HIERARCHICAL MEMORY

Leonid Timchenko[1], Natalia Kokriatska[1], Volodymyr Tverdomed[1], Iryna Yu. Yepifanova[2], Yurii Didenko[1], Dmytro Zhuk[1], Maksym Kozyr[1], Iryna Shakhina[3]

[1]State University of Infrastructure and Technology, Kyiv, Ukraine, [2]Vinnytsia National Technical University, Vinnytsia, Ukraine, [3]Vinnytsia Mykhailo Kotsiubynskyi State Pedagogical University, Vinnytsia, Ukraine

Abstract. *Parallel hierarchical memory (PI memory) is a new type of memory that is designed to improve the performance of parallel computing systems. PI memory is composed of two blocks: a mask RAM and a tail element RAM. The mask RAM stores the masks that are used to encode the information, while the tail element RAM stores the actual information. The address block of the PI memory is responsible for generating the physical addresses of the cells where the tail elements and their masks are stored. The address block also stores the field of addresses where the array was written and associates this field of addresses with the corresponding external address used to write the array. The proposed address block structure is able to efficiently generate the physical addresses of the cells where the tail elements and their masks are stored. The address block is also able to store the field of addresses where the array was written and associate this field of addresses with the corresponding external address used to write the array. The proposed address block structure has been implemented in a prototype PI memory. The prototype PI memory has been shown to be able to achieve significant performance improvements over traditional memory architectures. The paper will present a detailed description of the PI transformation algorithm, a description of the different modes of addressing organization that can be used in PI memory, an analysis of the efficiency of parallel-hierarchical memory structures, and a discussion of the challenges and future research directions in the field of PI memory.*

Keywords: parallel hierarchical memory, PI memory, address block, mask RAM, tail element RAM, performance improvement

ARCHITEKTONICZNE, STRUKTURALNE I FUNKCJONALNE CECHY RÓWNOLEGŁO-HIERARCHICZNEJ ORGANIZACJI PAMIĘCI

Streszczenie. *Równoległa pamięć hierarchiczna (pamięć PI) jest nowym typem pamięci zaprojektowanym w celu poprawy wydajności równoległych systemów obliczeniowych. Pamięć PI składa się z dwóch bloków: maski RAM i ogon RAM. Maska RAM przechowuje maski używane do kodowania informacji, podczas gdy ogon RAM przechowuje rzeczywiste informacje. Blok adresowy pamięci PI jest odpowiedzialny za generowanie fizycznych adresów komórek, w których przechowywane są elementy końcowe i ich maski. Blok adresowy przechowuje również pole adresu, w którym tablica została zapisana i kojarzy to pole adresu z odpowiednim adresem zewnętrznym użytym do zapisu tablicy. Proponowana struktura bloku adresowego jest w stanie efektywnie generować fizyczne adresy komórek, w których przechowywane są elementy ogonowe i ich maski. Blok adresowy może również przechowywać pole adresu, w którym tablica została zapisana i powiązać to pole adresu z odpowiednim adresem zewnętrznym użytym do zapisu tablicy. Zaproponowana struktura bloku adresowego została zaimplementowana w prototypie pamięci PI. Wykazano, że prototyp pamięci PI jest w stanie znacznie poprawić wydajność w porównaniu z tradycyjnymi architekturami pamięci. W artykule zostanie przedstawiony szczegółowy opis algorytmu konwersji PI, opis różnych trybów adresowania, które mogą być używane w pamięci PI, analiza wydajności równoległo-hierarchicznych struktur pamięci oraz omówienie wyzwań i przyszłych kierunków badań w dziedzinie pamięci PI.*

Słowa kluczowe: równoległa pamięć hierarchiczna, pamięć PI, blok adresowy, maska RAM, ogon RAM, poprawa wydajności

Introduction

In recent years, much effort has been made to create parallel computing systems that use the power of the Internet and the availability of computers in many homes and businesses. The main advantage of these approaches is that they provide a cheap environment for parallel computing for those who cannot afford the costs of supercomputers and hardware parallel processing. However, most of the solutions on offer are not very flexible in the use of available resources and are very difficult to install and configure [21]. When designing high-performance computers and systems, it is necessary to make many trade-offs, such as the sizes and technologies for each level of the hierarchy.

Analyzing previous work [9, 12, 14] on parallel-hierarchical structures, we can identify the main disadvantages of these systems [8, 10, 15].

For example, the lack of a method for evaluating peak performance as the number of instructions executed by the computer per unit of time (MIPS, Million Instruction Per Second) gives only a general idea of the speed, since it does not take into account the specifics of specific programs (it is difficult to predict how many and which instructions a user program will be displayed by the processor) [2, 5, 20].

Parallel computing systems are excessively expensive. According to the law of diminishing returns, the performance of a computer grows proportionally to the square of its cost; as a result, it is much more profitable to obtain the required computing power by purchasing one high-performance processor than by using several less powerful processors [4, 9, 10].

When organizing parallelism, performance losses grow too quickly. According to the Minsky hypothesis (Marvin Minsky), the acceleration of calculations achieved when using a parallel system is proportional to the binary logarithm of the number of processors (with 1000 processors, the possible acceleration is only 10).

Counter-argument. The given acceleration estimate is true for parallelizing certain algorithms. However, there are a large number of tasks, the parallel solution of which achieves close to 100% use of all available processors of a parallel computing system.

Sequential computers are constantly being improved. According to the well-known Moore's law, the complexity of sequential microprocessors doubles every 18 months, so the required performance can also be achieved on "ordinary" sequential computers.

However, the use of parallelism allows obtaining the necessary acceleration of calculations without waiting for the development of new faster processors. The efficiency of parallelism strongly depends on the characteristic properties of parallel systems. All modern sequential electronic computers operate in accordance with the classical von Neumann scheme; parallel systems differ significantly in architecture and the maximum effect from the use of parallelism can be obtained by full use of all the features of the hardware (consequence – the transfer of parallel algorithms and programs between different types of systems is difficult, and sometimes impossible).

For decades of operation of sequential electronic computers, a huge amount of software has been accumulated, oriented at sequential electronic computers; its processing for parallel computers is practically unrealistic.

There is a limit on the acceleration of calculation in the parallel – hierarchical implementation of the algorithm in comparison with the sequential one.

Parallel-hierarchical memory (PHM) is a promising approach to addressing the challenges of memory organization in parallel computing systems. PHM systems are organized in a hierarchy of levels, with each level providing different levels of performance and capacity. This allows the system to adapt to the needs of different applications. In this paper, we present an overview of the architectural and structural and functional features of the organization of PHM systems.

1. Examples of network parallel-hierarchical structures

The network structure allows to simulate the principle of operation of a distributed PI network and, due to spatial separation over time, processes information in a deterministic pyramidal PI network (Fig. 1). The network, built on the basis of PI transformation, consists of a set of subnetworks (Fig. 1) for forming features about the states of the spatio-temporal environment (PTE), the structure of which is homogeneous and consists of a number of interdependent hierarchical levels.

The network operation algorithm is universal and consists in the PI formation of sets of common and different states of the PTE. Generalization of all types of sensory information occurs at the very final stage of the transformation outside the hierarchical processing of each type of sensory information. So, the process of generalization between different types of sensory information begins only when the construction according to a certain set of features is completed [9].

On the figure 1 there is an example of transformed states on each level: ☐ – set of shared states on each level, ☒ – define common state on separate level.

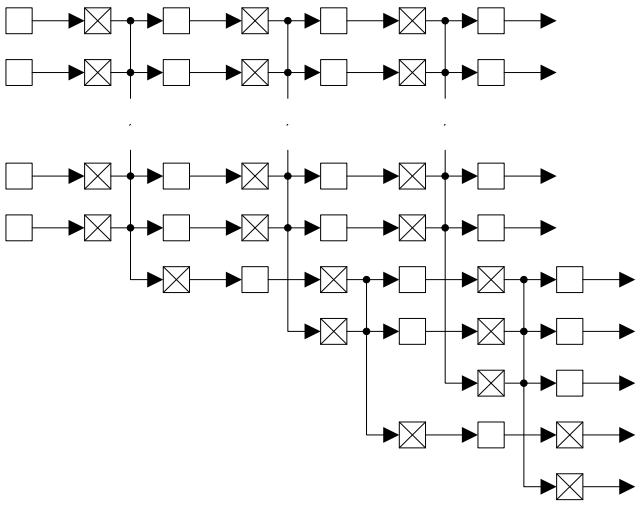


Fig. 1. Model of neural structur based on the PI transformation block

To associate the masks with their corresponding branches at each level, after their sequential writing starting from the lower branches, two additional bytes of information need to be written. This information includes the number of masks generated at that level and the highest branch number from which the last mask is selected at that level. This information is also generated in the PI transformation block.

On the figure 2 you see the structure of a parallel hierarchical memory (PI memory). The core of the structural organization of PI memory consists of three main components: the storage medium (mask RAM, tail element RAM), the PI transformation network, and the address block. The storage medium is composed of two blocks, each of which is a conventional RAM with a sequential data access structure. The word size of the mask RAM corresponds to the memory's access width, while the word size of the tail element RAM matches the bit depth of the image frame [1, 5].

The PI transformation network is responsible for encoding the information during the write operation and decoding it during the read operation from the memory.

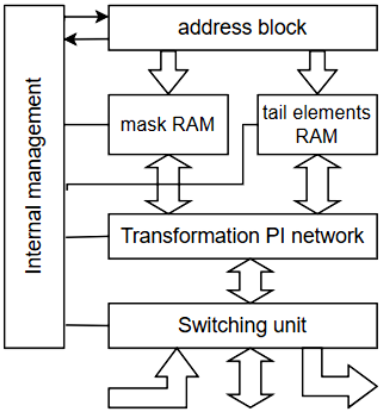


Fig. 2. Structure of PI memory with PI transforming network

2. Organization of the architecture of the address block parallel to the hierarchical memory

The address block of the PI memory implements the functions of generating internal addresses for the storage medium blocks based on the external address and the state of the PI transformation network.

In a linear sequential memory, the memory capacity (N), which represents the number of addressable data elements (cells), plays a crucial role in determining the address size. The dependency $H_{addr} = INT(log_2 N + 0.5)$, is based on the H_{addr} – number of binary digits in the address code, INT – represents the integer part operation. The address structure may contain additional fields used for control information, such as specifying base or index register numbers, address modification indicators, etc. Thus, the description of accessing conventional memory typically has a two-component structure $[C, adr]$, including C – control code and adr – address code [6, 7].

For the considered PI memory, the structure of the access description follows the conventional form and includes the address code. Unlike the cell address structure, here the address of the array or array package is used instead of the cell address. The control code in the address structure incorporates information about the memory operation mode: a) standard mode with the ability to access any cell, and b) array mode [3, 18].

Let's examine the structure of the address block (Fig. 3) in the PI memory that stores information arrays. The tasks of the address block are as follows:

1) During array write operation, generate physical addresses of the cells where the tail elements and their masks are stored.
2) After completing the write operation, store the field of addresses where the array was written and associate this field of addresses with the corresponding external address used to write the array.

3) During read operation using the external address value, determine the field of physical addresses where the array is stored and generate the necessary data physical addresses in the required sequence [1, 13].

The proposed address block structure, presented in Fig. 3, performs this function. To illustrate the operation of the suggested address structure in the PI memory, figure 3 also shows the control scheme and the mask RAM and single-element RAM blocks. The address block of the PI memory is outlined with a dashed line.

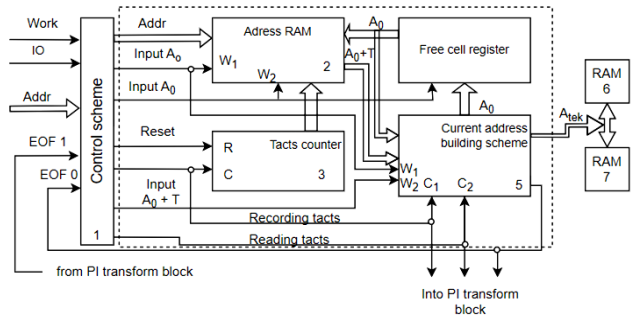


Fig. 3. Functional scheme of PI memory address block

To write an array to the PI transformation network, the array itself is digitally input to the data bus, and the control signals "Work" (selection of the memory), "I/O" (write information), and "Addr" (address) are sent to the control circuit. The control circuit routes the content of the "Addr" bus to the input of Address RAM 1 [5, 6]. Address RAM 1 has a capacity of $2 \times k$ cells, where k – is the address bus width. The control circuit generates the "Write Address" signal, which writes the address of the first available cell in Address RAM 1 and the current address generation circuit 4 from the free cell register 3. This address will be the same for Address RAM 6 and Address RAM 7 because the number of tail elements for decomposing one array using the PI transformation method is equal to the number of masks for that array. In other words, each selected element will have its corresponding mask. In the circuit, the first available cell address is required to generate the current physical addresses for writing the encoded array to Address RAM 6 and Address RAM 7.

In Block 2, the value A_0 is written to the cell with an address equal to the address on the "Addr" bus. This establishes the connection between the external address on the "Addr" bus and the internal addresses of the storage environment. Then, the control scheme generates a sequence of pulses on the "Write Clocks" line to perform the array write operation [8, 9, 21]. Using these pulses, Control Scheme 4 generates the current address by incrementing the value of A_0. The Clock Counter 2 counts the number of pulses on the "Write Clocks" line. When the decomposition of the array in the PI transformation block is completed, a signal "End of Write" is sent to the control scheme, and the pulses on the "Write Clocks" line cease. With the control signal "Write Time", generated by the control scheme, the content of Clock Counter 2 is written to Block 1 under the same address on the "Addr" bus, and the new address of the first available cell from Block 4 is written to Block 3. Then, the "Reset" signal resets Clock Counter 2, completing the write procedure. During the read operation, the control signals "Work", "Wr/Rd", and the corresponding address from the "Addr" bus are applied to the control scheme [7, 25].

The control scheme 1 generates the "Write" signal $A_0 + T$, which writes the values A_0 and T from Address RAM 2 to Current Address Generation Scheme 5, corresponding to the code on the "Addr" bus. In Scheme 5, the first current address is formed by adding the values A_0 and T. The control scheme generates a series of pulses on the "Read Clocks" line. These pulses are sent to the PI transformation block and Scheme 5, where for each pulse, the values A_0 and T are decremented by one. This process reads the array from Address RAM 6 and Address RAM 7. With each pulse on the "Read Clocks" line, when the value becomes zero, Scheme 5 generates the "End of Read" signal, which, when received by the control scheme, causes the cessation of pulses on the "Read Clocks" line [7, 19]. The decoding of the read array is completed in the PI transformation block, and the read process is finished.

When organizing the address block of the PI memory, significant difficulties arise due to the significant difference in addressing between the tail element memory block and the mask memory block [9, 31]. This is because the number of masks in processing an array packet, according to the PI transformation method, is much larger than the number of tail elements.

For algorithms where only one tail element can appear at any given time (parallel-sequential processing, selecting only the first tail element at each level, etc.), the addressing of the tail element RAM can be sequential. In this case, one bit in the masks is allocated as a flag indicating the presence or absence of a tail element for that mask. If the mask RAM contains information about the current transformation step and the number of masks at each level of that step, the flag indicating the presence of a tail element is not necessary [19, 23]. A mask corresponds to a tail element when that mask is the only one at its level in the current transformation step.

For algorithms where all tail elements are considered in the encoding result, and multiple tail elements can occur at different levels in a particular transformation step, their storage in memory is done sequentially, starting from the lowest level. During decoding, masks are selected starting from the highest level, and tail elements are read in reverse order, also starting from the highest levels [7, 17].

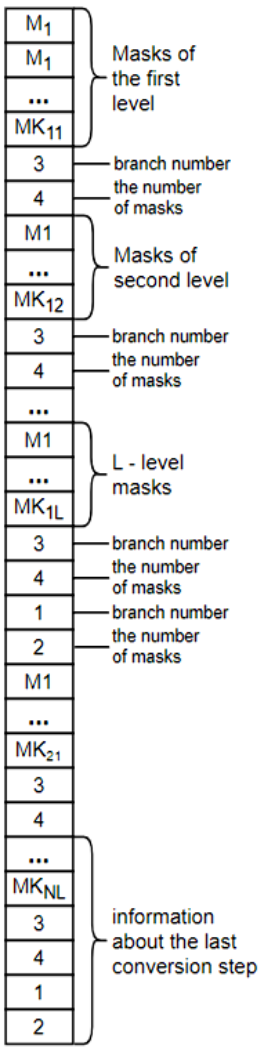


Fig. 4. Structure of information composition in the RAM mask by memory cell

The process of writing masks to the PI memory is further complicated by the fact that, according to the PI transformation algorithm, masks are generated in parallel at multiple levels, and each level can have multiple masks. A regular storage environment is used to store the mask codes. Therefore, at each transformation step, the mask codes are captured in buffer registers and are sequentially rewritten from the lower levels to the mask RAM, starting with the lowest levels and branches [12, 16].

The described association procedure is valid only when all masks belong to neighboring branches. It should be noted that on any level, there may be a situation where masks from one transformation step are not present on all branches located between the initial and final branches. In such a case, additional information about the location of masks within the level needs to be stored in the mask RAM. The arrangement of information in the mask RAM cells for one array packet is presented in figure 4 [13, 22, 29].

Masks have different sizes, so it is inefficient to allocate cells with pre-determined sizes to store them. The following structure is proposed for mask addressing:

a) The four most significant bits of the mask indicate the length of the mask in half-bytes.
b) The mask bits themselves follow, occupying the number of half-bytes indicated in the mask size. The length of the mask is determined by the PI transformation block.

Figure 5 illustrates the bit-wise structure for storing one mask.

By using this structure, the size of the area where the mask is stored can be varied widely. The minimum size for storing a mask would be 1 byte. At the same time, you can record $4 \le k \le 60,\ mod\,{}^{k}/_{4} = 0$.

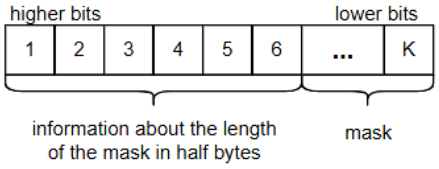


Fig. 5. Structure of information composition for separate mask (mod – euclidean division)

Here, half a byte is used to indicate the length of the mask, and the other half a byte represents the code of the mask. The maximum size for the mask area is 15 (4 + 4) = 64 bits, or 8 bytes. In this case, 4 bits indicate the mask length, and the remaining 60 bits represent the mask code. In regular memory, the word size is 16 bits, and the addressing allows accessing individual bits.

During the write mode, the addressing block of the PI memory sequentially increments the addresses for writing the masks until the "End of Write" signal is received from the PI transformation block. The same process is used to generate addresses for writing the tail elements into their respective RAM [17, 23, 26].

During the read mode, the addressing block of the PI memory uses the mask length information and the information stored in cells 1, 2, 3, 4 (Fig. 5) to generate physical addresses for the mask RAM. The masks are selected in sequence, starting from the highest levels and branches. For the tail element RAM, the addressing block generates addresses by decrementing the current address by "one" when selecting a mask from the mask RAM that requires a tail element.

The organization of the addressing part becomes simpler if the encoding of the array packet is done in a parallel-sequential manner, where each level is processed in parallel while the levels themselves are processed sequentially. In this case, an additional buffer memory is required to store arrays from the previous level.

Indeed, there can be different modes of addressing organization depending on the transformation algorithm based on the PI transformation method and the corresponding structural organization of the address block and the entire memory system [16]. This topic is a separate subject of research.

The approach presented here outlines a general method of addressing the tail elements and masks in a parallel memory based on the PI transformation principle.

When designing a PI memory, the development of the address block should consider the requirements imposed on it, such as performance, hardware costs, cost, power consumption, etc. [13, 18]. Therefore, in a specific implementation of a particular address block algorithm, there is a question of what should be performed at the hardware level and what can be offloaded to software. To make a decision in each specific case, a special analysis is required.

The generalized structural diagram of a parallel memory implemented according to the PI transformation algorithm is presented in Fig. 6. The PI transformation block not only performs encoding and decoding of the original information arrays but also controls the memory units (RAM) and generates the addresses for data retrieval. RAM1 stores the values of tail elements for each level, while RAM2 stores the masks generated during the transformation process in the PI transformation block.

This generalized structural diagram illustrates the main components and connections in a PI memory system, which can be further adapted and expanded for specific implementations and requirements.

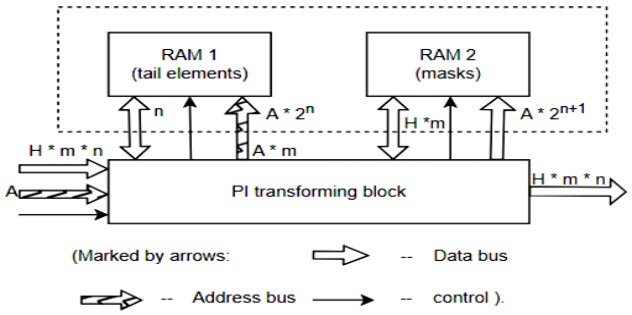


Fig. 6. Structural scheme of PI memory (A – bit size address bus, m – array size, n – word size in array)

The address bus width of the RAM is shown for the case when the PI transformation block performs transformations according to the expressions (15), (16), (17), and the level processing is done sequentially.

3. Analysis of the efficiency of parallel-hierarchical memory structures

Let us derive an expression that allows calculating the number of cycles required for the complete write (read) operation in a PI memory with a word dimension of $H \cdot m$ [16, 17, 27].

The number of levels until the packet of arrays fully converges, excluding only one initial single element on each level, is calculated using the formula:

$$U = H \cdot m - \sum_{i=1}^{u}\sum_{j=1}^{W_i} r_{i,j} + \sum_{i=1}^{u}\sum_{j=1}^{W_i} L_{i,j} + 1 = \\ = H \cdot m - \sum_{i=1}^{u}\sum_{j=1}^{W_i}\left(r_{i,j} - L_{i,j}\right) + 1 \quad (1)$$

where U is the number of levels formed as a result of the PI processing of arrays, i – is the ordinal number of the level, W_i – is the number of branches (original arrays) for the i-th level, j – is the branch number on the current level, $r_{i,j}$ and $L_{i,j}$ are the respective counts of identical elements and groups with identical elements on the i-th level of the j-th branch.

The term 1 in formula (1) accounts for the first level where there is no initial tail element in the PI transformation network. When excluding all tail elements from further processing at each level, formula (1) takes the following form:

$$U = H \cdot m - \sum_{i=1}^{u}\sum_{j=1}^{W_i}\left(r_{i,j} - L_{i,j}\right) - \sum_{i=1}^{u}\sum_{j=1}^{W_i} z_{i,j} + 1 \quad (2)$$

where $z_{i,j}$ – number of tail elements in j-th branch of i-th level (value of $z_{i,1}$ does not account for separated element $a_{1,1}$).

The number of cycles required to write (read) the initial tail elements in RAM1 is equal to the number of levels minus one, and it is generally less than the number of cycles required to write (read) the masks in RAM2, especially for certain algorithms [10, 27, 28]. Therefore, the determining factor influencing the performance of the PI memory is the time it takes to write (read) the masks obtained from encoding using the PI transformation method. The following are the formulaic dependencies for calculating the number of cycles required to write (read) the masks for all levels:

a) for the first level:

$$T_1 = m - \min_{j=1,W_1}\left(r_{1j} - L_{1j}\right) \quad (3)$$

b) second level:

$$T_2 = \max_{j=1,W_2}\left(\omega_{2j} + (j-1) - \left(r_{2j} - L_{2j}\right)\right) \quad (4)$$

where ω_{2j} is the number of words (elements) in the initial array of the j-th branch of the 2nd level. Since each subsequent branch on all levels, except the first, starts processing one step later than the previous one, and this step corresponds to the cycle of writing (reading) the mask, the term *(j – 1)* in expression (4) represents the shifting process. The shifting starts from the second branch, so one is subtracted from j. Thus, expression (4) determines the maximum length of the decomposition among all branches of the second level, taking into account the shifting of each subsequent branch and the convolutions $\left(r_{2j} - L_{2j}\right)$ in each branch, if applicable [17, 19, 30].

c) for the sequential levels will be correct formula (4).

$$T_i = \max_{j=1,W_i}\left(\omega_{ij} + j + L_{ij} - r_{ij} - 1\right)$$
$$i = \{1,2,\dots,u\}, \quad j = \{1,2,\dots,W_i\} \quad (5)$$

Number of leaves on each level counts by formulas:

$$W_1 = H$$
$$W_2 = T_1$$
$$W_3 = T_2 - 1 \quad (6)$$
$$\dots$$
$$W_i = T_{i-1} - 1, \; i = \{3,4,\dots,u\}$$

If all tail elements are taken into account as a result of expansion, then expressions (6) will take the form:

$$W_1^{э} = H$$
$$W_2^{э} = T_1 - z_{1j}, \quad z_{1j} = \{0,1,2,\dots,m-1\}$$
$$W_3^{э} = T_2 - \sum_{j=1}^{W_2} z_{2j} - 1 \quad (7)$$
$$\dots$$
$$W_i^{э} = T_{i-1} - \sum_{j=1}^{W_{i-1}} z_{i-1j} - 1, \; i = \{3,4,\dots,u\}$$

With sequential processing of levels, the total number of cycles for writing (reading) masks is:

$$T_{\Sigma}^{pos} = \sum_{i=1}^{u} T_i = m - \min_{j=1,W_1}\left(r_{1j} - L_{1j}\right) + \sum_{i=2}^{u}\max_{j=1,W_i}\left(\omega_{ij} + j + L_{ij} - r_{ij} - 1\right) \quad (8)$$

With a parallel encoding process, the number of cycles for writing (reading) an information array with a dimension of H·m words will be:

$$T_{\Sigma}^{pair} = U + max\left(\omega_{uj} + j + L_{uj} - r_{uj} - 1\right) - 2 = U + \max_{j=1,W_u}\left(\omega_{uj} + j + L_{uj} - r_{uj}\right) - 3 \quad (9)$$

The subtraction of two in the first part of expression (9) is done to exclude the first and last levels. The first level does not involve a cycle of writing, and the last level is already accounted for in the term that determines the maximum length of the last level.

For regular sequential memory, writing (reading) a two-dimensional array of words will require cycles:

$$T^{st} = H \cdot m \quad (10)$$

From the analysis of expressions (1), (8), (9) and (10) it follows that T_{Σ}^{pair} less than T^{ST} for $\sum_{i=1}^{u}\sum_{j=1}^{W_i}\left(r_{i,j} - L_{i,j}\right) - \max_{j-1,W_u}\left(\omega_{uj} + j + L_{uj} - r_{uj}\right) + 2$, a T_{Σ}^{pos} depending on the variables K, i, ω and W can be either less or more T^{st}.

The capacity of the PI memory (Fig. 6) depends on the type of F^* and Q^* transformation and the method of mask generation during array processing.

The capacity of RAM1 is relatively small with a fixed word size of n bits. RAM2, on the other hand, has a much larger capacity, and the word size varies from m to H bits for the first and second levels, to 2 bits for the last level. This is achieved through additional technical and software means, compromises, and flexible addressing that allows access to memory cells of different lengths. The number of levels and branches, as seen from (8) and (9), directly affects the memory performance and required capacity [24, 31].

The PI memory (Fig. 5) allows storing 2^A two-dimensional arrays with a dimension of $H \cdot m$, and during reading, it can generate the entire original array in parallel. It can be used, for example, to store information about image fragments with dimensions $H \cdot m$ samples. The PI transformation block in the PI memory is a complex device, which may not always be justified in practice for hardware implementation of parallel memory. A simpler transformation algorithm can be used for encoding a one-dimensional array, i.e., processing a single branch using the PI transformation method [1, 24, 28]. A parallel-hierarchical RAM with pre-processing of a one-dimensional array can have a single RAM block (Fig. 7), into which the selected element value (data-1) and the mask of that element (data-2) are simultaneously written (read) under the same address [11].

The number of clock cycles for writing (reading) in this case is determined by the formula:

$$T^{brn} = m - r + L \quad (11)$$

where r – number of the same words in the array; L – number of pairs with the same words.

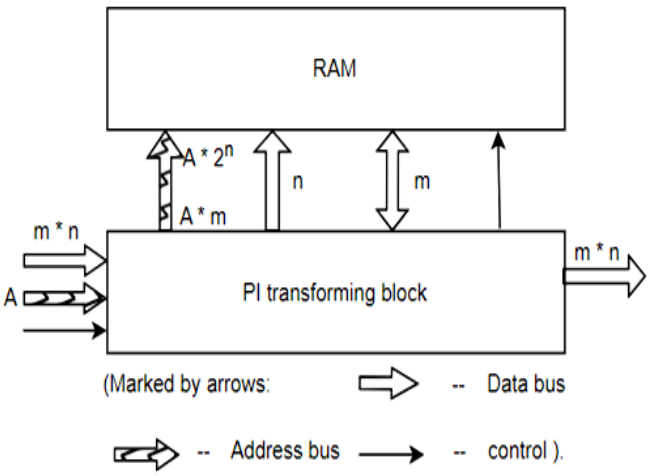


Fig. 7. Structural scheme of PI memory (A – Address bus registry size; m – address' registry size; n – arrays' word registry size)

4. Research results

Comparing formulas (10) and (11), it can be seen that the PI memory with preliminary array transformation is more efficient compared to regular memory by the amount of $r - L$.

To store information about one array in the PI memory, the required memory volume is:

$$O^{brn} = T^{brn} \cdot n + T^{brn} \cdot m = T^{brn} \cdot (n + m) = (m - r + L)(m + n) \quad (12)$$

Required value for usual memory:

$$O^{st} = m \cdot n \tag{13}$$

Subtracting (12) from (13), we get:

$$P = O^{st} - O^{st} = r \cdot (n+m) - L \cdot (n+m) - m^2 \tag{14}$$

From (14) as consequence:

$$\begin{aligned} r \cdot (n+m) - L \cdot (n+m) - m^2 = 0, \ \text{to} \ O^{st} = O^{brn} \\ r \cdot (n+m) - L \cdot (n+m) - m^2 > 0, \ \text{to} \ O^{st} > O^{brn} \\ r \cdot (n+m) - L \cdot (n+m) - m^2 < 0, \ \text{to} \ O^{st} < O^{brn} \end{aligned} \tag{15}$$

From equation (15), it follows that the volume of the PI memory compared to traditional memory structures will be smaller, the higher the values of r and n, and the smaller the values of L and m.

The efficiency of reducing the volume of PI memory can be assessed by the ratio:

$$R = \frac{O^{st}}{O^{brn}} = \frac{mn}{(m-r+L)(n+m)} \tag{16}$$

Consider a specific example: let there be an array with such parameters $m = 15$, $n = 8$, $r = 10$ and $L = 4$:

$$T^{brn} = 15 - 10 + 4 = 9\tau, T^{st} = 15\tau$$
$$O^{brn} = 9(15+8) = 207\text{bits}$$
$$O^{st} = 15 \times 8 = 120 \text{ bits}$$

For an array of the same dimension but with different parameters $m = 15$, $n = 8$, $r = 13$, $L = 3$:

$$T^{brn} = 15 - 13 + 3 = 5\tau, \ T^{st} = 15\tau$$
$$O^{brn} = 5(15+8) = 115\text{bits}$$
$$O^{st} = 15 \times 8 = 120 \text{ bits}$$

In the first case, the performance of the PI memory is higher than that of the standard memory, but the memory capacity is larger. In the second case, both the performance is higher and the memory capacity of the PI memory is smaller than that of the standard memory. In general, the number of clock cycles required for writing (reading) information using the proposed PI transformation method is within the range of:

$$\tau \leq T^{brn} \leq T^{st}$$

Therefore, the explored architectural features of the parallel-hierarchical memory classify it as a non-von Neumann computational structure. The proposed structure of the PI memory, focused on parallel and compact processing of data fields, enables real-time transformation. The researched architectural characteristics allow for efficient memory organization for array processing and transformation, which can be valuable in various applications requiring fast and parallel data processing.

5. Conclusions

In this paper we demonstrate the potential of parallel-hierarchical memory (PHM) systems to significantly improve the performance of parallel computing systems. The results show that PHM systems can achieve significant performance improvements over traditional memory systems, particularly for applications with a high degree of data parallelism.

Research also highlight the challenges that need to be addressed in order to realize the full potential of PHM systems. These challenges include the design of efficient memory access policies, the development of scalable and efficient memory management schemes, the development of fault-tolerant PHM systems, and the development of PHM systems that are compatible with existing programming models and languages.

Provided information is a foundation for further research on PHM systems. The results suggest that PHM systems have the potential to revolutionize the field of parallel computing, and the challenges identified in this paper provide a roadmap for future research in this area.

Presented results is significant because they demonstrate the potential of PHM systems to revolutionize the field of parallel computing.

References

[1] Aboutabl A. E., Elsayed M. N.: A Novel Parallel Algorithm for Clustering Documents Based on the Hierarchical Agglomerative Approach. International Journal of Computer Science & Information Technology – IJCSIT 3(2), 2011, 152–163.

[2] Bisikalo O. et al.: Parameterization of the Stochastic Model for Evaluating Variable Small Data in the Shannon Entropy Basis. Entropy 25(2), 2023, 184.

[3] Bykov M. et al.: Neural network modelling by rank configurations. Proc. of SPIE 10808, 2018, 1080821.

[4] Kim S., Wunsch D. C.: A GPU based Parallel Hierarchical Fuzzy ART clustering. IJCNN IEEE, 2011, 2778–2782.

[5] Kohonen T.: Self Organization and Associative Memory: Third Edition. Springer-Verlag, New York, 1989.

[6] Kovtun V., Izonin I.: Study of the Operation Process of the E-Commerce Oriented Ecosystem of 5Ge Base Station, Which Supports the Functioning of Independent Virtual Network Segments. Journal of Theoretical and Applied Electronic Commerce Research 16(7), 2021, 2883–2897.

[7] Kukharchuk V. V. et al.: Features of the angular speed dynamic measurements with the use of an encoder. Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Srodowiska – IAPGOS 12(3), 2022, 20–26.

[8] Kukharchuk V. V. et al.: Information Conversion in Measuring Channels with Optoelectronic Sensors. Sensors 22(1), 2022, 271.

[9] Kuusilinna K. et al.: Configurable parallel memory architecture for multimedia computers, Journal of Systems Architecture 47(14–15), 2002, 1089–1115.

[10] Kvyetnyy R. et al.: Inverse correlation filters of objects features with optimized regularization for image processing. Proc. SPIE 12476, 2022, 124760Q.

[11] Li Z., Li K., Xiao D., Yang L.: An Adaptive Parallel Hierarchical Clustering Algorithm. Perrott, R., Chapman, B.M., Subhlok, J., de Mello, R.F., Yang, L.T. (eds): High Performance Computing and Communications. HPCC 2007. Lecture Notes in Computer Science 4782. Springer, Berlin, Heidelberg 2007.

[12] Nere A., Lipasti M.: Optimizing Hierarchical Algorithms for GPGPUs. Master's Project Report. University of Wisconsin Madison, 2010.

[13] Orazayeva A. et al.: Biomedical image segmentation method based on contour preparation, Proc. SPIE 12476, 2022, 1247605.

[14] Osman A. A. M.: A Multi-Level WEB Based Parallel Processing System: A Hierarchical Volunteer Computing Approach. World Academy of Science, Engineering and Technology 13, 2006, 66–71.

[15] Pavlov S. V. et al.: The use of Bayesian methods in the task of localizing the narcotic substances distribution. International Scientific and Technical Conference on Computer Sciences and Information Technologies 2, 2019, 8929835, 60–63.

[16] Rajasekaran S.: Efficient Parallel Hierarchical Clustering Algorithms. IEEE Transactions on Parallel and Distributed Systems 16(6), 2005, 497–502.

[17] Romanyuk S. A. et al.: Using lights in a volume-oriented rendering. Proc. SPIE 10445, 2017, 104450U.

[18] Rose K.: Deterministic Annealing, Clustering and Optimization. Ph.D. Thesis, California Institute of Technology, Pasadena, 1991.

[19] Sobota B.: Parallel Hierarchical Model of Visualization Computing. Journal of Information, Control and Management Systems 5(2), 2007, 345–350.

[20] Sudarshan R. Lee S. E.: A Parallel Hierarchical Solver for the Poisson Equation, May 14, 2003,.

[21] Timchenko L. et al.: New methods of network modelling using parallel-hierarchical networks for processing data and reducing erroneous calculation risk. CEUR Workshop 2805, 2020, 201–212.

[22] Timchenko L. I., Kokriatskaia N. I., Pavlov S. V., Tverdomed V.: Method of indicators forecasting of biomedical images using a parallel-hierarchical network. Proc. of SPIE 11176, 2019, 111762Q.

[23] Timchenko L. I.: A multistage parallel-hierarchic network as a model of a neuronlike computation scheme. Cybern Syst Anal. 36, 2000, 251–267.

[24] Tolegen G., Toleu A., Mamyrbayev O., Mussabayev R.: Neural Named Entity Recognition for Kazakh. Lecture Notes in Computer Science 13452, 2023, 3–15.

[25] Tymkovych M. et al: Ice crystals microscopic images segmentation based on active contours. IEEE 39th International Conference on Electronics and Nanotechnology – ELNANO 2019, 493–496 [https://doi.org/10.1109/ELNANO.2019.8783332].

[26] Vasilevskyi O. et al.: A new approach to assessing the dynamic uncertainty of measuring devices. Proc. of SPIE 10808, 2018, 108082E.

[27] Vysotska O. V., Nosov K.: An approach to determination of the criteria of harmony of biological objects. Proc. of SPIE, 10808, 2018, 108083B

[28] Wójcik W., Pavlov S., Kalimoldayev M.: Information Technology in Medical Diagnostics II. Taylor & Francis Group, CRC Press, Balkema book, London 2019.

[29] Ybytayeva G. et al.: Creating a Thesaurus "Crime-Related Web Content" Based on a Multilingual Corpus. CEUR Workshop Proceedings 3396, 2023, 77–87.

[30] Zeki S.: A Vision of the Brain. Blackwell Scientific Publications, Oxford 1993.

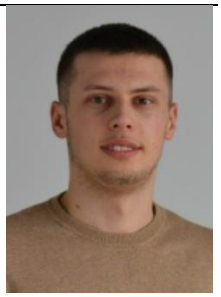
[31] Zhao X., Guo Y., Feng Z., Hu S.: Parallel Hierarchical Cross Entropy Optimization for On-Chip Decap Budgeting. Design Automation Conference, Anaheim, CA, USA, 2010, 843–848.

Prof. Leonid Timchenko
e-mail: tumchenko_li@gsuite.duit.edu.ua

Doctor of Technical Science, professor. 56 articles published in Scopus, 227 citations in 112 articles (h-index = 8).

https://orcid.org/0000-0001-5056-5913

Yurii Didenko
e-mail: didenk.y.v@gmail.com



Post-graduate student at State University of Infrastructure and Technology.
Research interests: systems of artificial intelligence, image processing systems.

https://orcid.org/0009-0008-1033-4238

Ph.D. Natalia Kokriatska
e-mail: nkokriatskaia@gmail.com

Ph.D., associate professor. 34 articles published in Scopus, 119 citations in 82 articles (h-index = 7).

https://orcid.org/0000-0003-0090-3886

Dmytro Zhuk
e-mail: zhuk_do@ukr.net

Post-graduate student at State University of Infrastructure and Technology.
Research interests: systems of artificial intelligence, image processing systems.

https://orcid.org/0000-0001-8951-5542

Ph.D. Volodymyr Tverdomed
e-mail: tverdomed@gsuite.duit.edu.ua

Ph.D., associate professor, Director of Kyiv Institute of Railway Transport, State University of Infrastructure and Technology, Ukraine.
13 articles published in Scopus, 13 citations in 16 articles (h-index = 3).
Research interests: development of methods for diagnosing the technical condition and forecasting the duration of operational work of railway track elements and track devices.

https://orcid.org/0000-0002-0695-1304

Maksym Kozyr
e-mail: jettab3@gmail.com

Ph.D. student. Artificial Intelligence Systems and Telecommunication Technologies Department, State University of Infrastructure and Technology, Ukraine.
Research interests: systems of artificial intelligence, image processing systems.

https://orcid.org/0009-0007-2564-6552

Prof. Iryna Yu. Yepifanova
e-mail: yepifanova@vntu.edu.ua

Doctor of economic sciences, professor, Vice-rector by science work, Faculty of Management and Information Security of Vinnytsia National Technical University, academician of the Academy of Economic Sciences of Ukraine.
Scientific interests: financial support of innovative activities of domestic enterprises, enterprise potential, competitiveness, personnel management, digital economy, energy saving.

https://orcid.org/0000-0002-0391-9026

Ph.D. Iryna Shakhina
e-mail: rom.shahin@gmail.com

Ph.D. (in Pedagogy), associate professor at the Department of Innovation and Information Technologies in Education Vinnytsia Mykhailo Kotsiubynskyi State Pedagogical University, Vinnytsia, Ukraine.
Scientific direction: information technologies, image processing, innovation in pedagogic

https://orcid.org/0000-0002-4318-6189

http://doi.org/10.35784/iapgos.5516 *received: 27.10.2023 | revised: 10.03.2024 | accepted: 15.03.2024 | available online: 31.03.2024*

SIMULATION AND COMPUTER MODELING OF BRIDGE STRUCTURES DYNAMICS USING ANSYS

Anzhelika Stakhova, Adrián Bekö
Slovak University of Technology, Department of Structural Mechanics, Bratislava, Slovakia

Abstract. *This study focuses on utilizing computer modeling and simulation techniques, specifically the ANSYS software, to analyze the dynamics of bridge structures. The primary objective was to study the vibrations of a riverbed metal bridge structure and determine their characteristics. The research involved theoretical dynamic calculations considering the design features of the bridge components and the materials used in their construction. The obtained results enabled the determination of resonance frequencies for the vibration modes. By utilizing the ANSYS software, a three-dimensional virtual model of the bridge structure was created, allowing for a detailed analysis of its dynamic behavior. The first three vibration modes of the riverbed metal bridge structure were calculated, and numerical results were obtained for six modes. The findings of this research have practical significance as they provide informed decision-making support during the construction, maintenance, and modernization of bridge structures. The study of bridge dynamics using advanced technologies contributes to enhancing the safety, reliability, and longevity of these vital infrastructure assets.*

Keywords: ANSYS software package, computer modeling, bridge structures, dynamics, virtual model

SYMULACJA I MODELOWANIE KOMPUTEROWE DYNAMIKI KONSTRUKCJI MOSTÓW Z WYKORZYSTANIEM ANSYS

Streszczenie. *Niniejsze badanie koncentruje się na wykorzystaniu technik modelowania komputerowego i symulacji, w szczególności oprogramowania ANSYS, do analizy dynamiki konstrukcji mostowych. Głównym celem było zbadanie drgań metalowej konstrukcji mostu w korycie rzeki i określenie ich charakterystyki. Badania obejmowały teoretyczne obliczenia dynamiczne uwzględniające cechy konstrukcyjne elementów mostu oraz materiały użyte do ich budowy. Uzyskane wyniki umożliwiły wyznaczenie częstotliwości rezonansowych dla postaci drgań. Wykorzystując oprogramowanie ANSYS stworzono trójwymiarowy wirtualny model konstrukcji mostu, pozwalający na szczegółową analizę jego dynamicznego zachowania. Obliczono trzy pierwsze postacie drgań metalowej konstrukcji mostu koryta rzeki i uzyskano wyniki numeryczne dla sześciu postaci. Wyniki tych badań mają znaczenie praktyczne, ponieważ zapewniają świadome wsparcie w podejmowaniu decyzji podczas budowy, utrzymania i modernizacji obiektów mostowych. Badanie dynamiki mostów przy użyciu zaawansowanych technologii przyczynia się do poprawy bezpieczeństwa, niezawodności i trwałości tych kluczowych elementów infrastruktury.*

Słowa kluczowe: pakiet oprogramowania ANSYS, modelowanie komputerowe, konstrukcje mostowe, dynamika, model wirtualny

Introduction

Studying the dynamics of bridge structures is of great importance in evaluating their operational characteristics.

The investigation of the dynamics of bridge structures involves the study and analysis of various aspects. Firstly, an analysis of the loads and forces acting on the structure is conducted, such as road traffic, wind effects, train vibrations, or other dynamic factors [3, 4]. Subsequently, modelling and calculations are carried out to predict the dynamic behaviour of the structure, including the analysis of resonance phenomena and possible resonance frequencies [5, 7]. Furthermore, an assessment of the vibrational characteristics of the structure is performed, including the amplitude of oscillations, natural frequencies, and damping [8, 15].

Researching the dynamics of bridge structures requires the application of various methods and tools. Sensors, accelerometers, deformation sensors, and other instruments can be used for data collection and vibration monitoring [12]. Modern computer programs enable numerical modeling and calculations, which help predict the dynamic behaviour of the structure under different conditions and optimize its parameters [5].

Research on the dynamics of bridge structures using modern technologies and software tools, such as ANSYS, also enables virtual testing and optimization of bridge designs. This reduces costs associated with physical testing and allows engineers and designers to more efficiently utilize resources in the construction of new bridges or the upgrading of existing ones.

The results of studying the dynamics of bridge structures can have a wide range of applications. They can be used to optimize the design of new bridges and the reconstruction of existing bridge structures to enhance their durability, safety, and efficiency. The obtained data and analytical findings from the research can be utilized in developing engineering solutions for structural reinforcement, selecting optimal materials, determining parameters of damping systems, and formulating maintenance strategies and regular condition monitoring of bridges.

Furthermore, investigating the dynamics of bridge structures contributes to the development of more accurate models and predictions for determining allowable loads and forecasting the behaviour of structures under operational conditions.

Thus, research on the dynamics of bridge structures using modern methods and software is of great importance for infrastructure development and ensuring the safety of transportation networks. It enables obtaining more accurate data on the dynamic behavior of bridge structures, making informed decisions in design and maintenance, and reducing the risks of unforeseen circumstances.

1. Dynamic testing of bridge structures

During dynamic testing of bridge structures, various types of dynamic actions are employed, which can be categorized as follows: moving dynamic loads, impact dynamic loads, and continuous vibration loads. It should be noted that dynamic actions on bridges result in both vertical and horizontal displacements. However, the effects of these displacements on the load-carrying capacity of bridges vary depending on their purpose. In the case of road bridges, vertical oscillations are typically the focus of investigation. For certain types of bridges, such as suspension bridges with low horizontal stiffness, knowledge of both types of oscillations is crucial. For railway bridges, horizontal vibrations are of primary concern.

Two types of moving dynamic loads are distinguished: loads generated by the natural flow of vehicles [11, 15] and loads generated by specialized vehicles [10, 13]. Loads in regular traffic flow are uncontrolled, while those created by specialized vehicles (e.g., vehicles loaded with ballast) are controlled. The passage of vehicles occurs at a specified constant speed along different lanes to create maximum dynamic loads (figure 1). Impact loads are typically generated by mechanical and hydraulic hammers [6], loads dropped from a certain height [14], specialized machines for inducing impact vibrations [6], and other devices (figure 2). However, the most common type of dynamic load applied during dynamic testing of bridge structures involves the use

of specialized vehicles moving at different speeds. Combinations of different types of vibration excitations are also employed.

Depending on the types and schemes of loading, different types of sensors are used to obtain information about the response of structures to the applied actions.

For moving dynamic actions such as natural traffic flow, movement of specialized vehicles, pedestrian movement, wind flow, and temperature, only electrical sensors are used. In fact, the same sensors used for static testing are employed, including strain gauges, inductive sensors, optical sensors, piezoceramic sensors, fiber optic sensors, and others [1, 2]. Instruments and systems such as MEGADAC 3108, Si425-500, STS, SHM, OPTIM MEGADAC, 3852A (HP), PSM-R, and others are utilized for processing signals from these types of sensors. Additionally, piezoceramic accelerometers and induction-type velocity sensors are widely used. In this case, vibrations are recorded in three directions – vertical, horizontal, and transverse.

Fig. 1. Dynamic loading of bridge structures using specialized sand-loaded vehicles: a) passage through the central roadway of a split reinforced concrete bridge, b) passage through the outer roadway lane of a split reinforced concrete bridge

For non-moving dynamic actions in the form of impact loads, accelerometric, strain gauge, fiber optic sensors, as well as displacement sensors, are used [6, 13]. However, in the latter case, the registered response of the structure is not typically a vibrational process but rather an impulsive response to the impact action. For non-moving continuous loads, accelerometers [6], optical lasers [2], and vibration velocity sensors [16] are used to measure the resulting vibrations. Acceleration and velocity sensors such as V401 CR, TCU102, Dytran 3100B, 5801A4, 353B03, 3031-100, WR-731A, PCB, CXL02, QA-700, CB, and SG, produced in Japan, the USA, Canada, Norway, Australia, Finland, Russia, are employed. The frequency range of acceleration sensors typically ranges from 0.5 Hz to 1000 Hz, with sensitivities ranging from 0.1 V/m/s² to 5.0 V/m/s². For velocity sensors, the frequency range is typically -5 Hz to 400 Hz, with sensitivities of -10 V/m/s to 20 V/m/s.

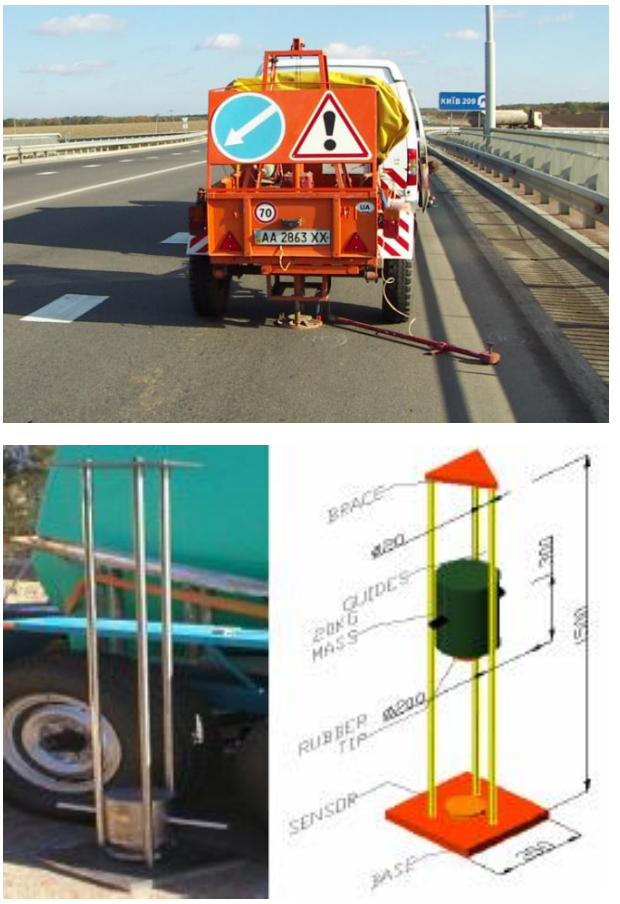


Fig. 2. Special devices for dynamic loading of bridge structures: a) machine for inducing impact loads, b) drop hammer with falling load

When processing signals obtained from strain gauges, optical sensors, and displacement transducers, the systems discussed above can be used. However, a wide range of devices and systems are employed for processing signals from accelerometers and velocity sensors, which are designed specifically for vibration signal analysis. Typically, these include multichannel systems such as TEACR-280 FM, SA-390, MCBS, VFC-05, USS8, Diamos, Jaguar, Cougar, SigLab, VX2842, HP3566A, and others. However, these systems are primarily intended for continuous vibration signals generated by operating equipment, such as machine tools, ventilation units, power generators, gas turbines, and so on. Consequently, they impose certain limitations on the amount of recorded information (1024, 2048, or 4096 data points), assuming the possibility of subsequent recording and processing of continuous signals. Similar conditions can be created when dealing with non-moving vibration loads, allowing for the full utilization of the mentioned systems.

However, the frequencies of vibrations in bridge structures are significantly lower compared to the frequencies generated by continuously operating equipment. In these cases, the limitation on the sample size, imposed by increasing the sampling interval to ensure the registration of information over a longer time span (the need to capture multiple periods of the resulting signal), leads not only to a restriction on the number of processed parameters but also to significant errors in the obtained results. Therefore, systems developed by researchers [1, 6, 9, 13] are widely used.

Standard mathematical packages that include data processing using fast Fourier transform are widely employed for information processing. Primary data acquisition devices are used as well. More complex systems are being developed, where stationary sets of primary transducers are employed on the structures, and information transmission is carried out through telephone communication lines or using GPS systems. However, in these systems, despite the use of modern technologies, standard information is recorded, and standard methods are applied for its processing.

2. Modeling of bridge structures dynamics

In addition to the aforementioned methods, there are other approaches to studying the dynamics of bridge structures that can provide information about their characteristics. Since analysis is often required for voluminous and complex models, solving such problems requires the use of modern software packages such as ANSYS. Analysis in this software package is performed using the finite element method.

Modern software tools allow for the creation of three-dimensional virtual models of structures and conducting various scenarios of loads and operating conditions. This enables researchers to analyze the dynamic behavior of the structure, identify potential issues, optimize the design, and make informed decisions regarding maintenance and bridge modernization.

By utilizing software packages like ANSYS, the complexity and size of bridge models can be effectively considered, providing more accurate and realistic results. This is particularly important when analyzing complex bridges, taking into account factors such as environmental changes, dynamic loads, and interactions with surrounding objects.

Therefore, the application of modern software tools, including ANSYS, is an efficient tool for studying the dynamics of bridge structures, enabling valuable insights into their behaviour and supporting informed engineering decisions.

The ANSYS software package consists of a large family of specialized programs that share common functions. However, the mathematical capabilities of each program are designed to solve specific classes of problems. One notable feature of the software is the file compatibility among all members of the ANSYS family across different platforms.

For the purpose of analysis within the ANSYS software package, a bridge structure was selected as the object of study. The bridge structure consisted of eight spans: seven approach spans and one main span (see figure 3). The bridge scheme employed a deck-girder configuration for the approach spans, with a truss system in the main span (3×16.76 + 53.30 + 4×16.76). The total length of the bridge was 171.64 meters.

By using the ANSYS software, engineers were able to conduct detailed analyses of the bridge structure, taking into account its geometry, material properties, and boundary conditions. The software allowed for the simulation of various loading scenarios, such as static loads, dynamic loads, and environmental conditions, to assess the structural behavior and performance of the bridge.

Fig. 3. Bridge with a metal truss

The river section of the bridge consists of a single span structure supported by 3-4 intermediate piers. The span structure includes upper and lower metal trusses (see Fig. 4) with a roadway at the bottom, metal connections, a beam grid with a monolithic reinforced concrete slab for the roadway section, and prefabricated sidewalks. The metal beam grid of the river section is composed of transverse and longitudinal beams with I-shaped cross-sections. The transverse beams are located at the lower nodes of the truss with a spacing of 7.5 m. Three longitudinal beams are installed at the level of the upper flange of the transverse beams with a spacing of 2.76 m. The cross-section of the upper and lower flanges of the trusses consists of two channel sections (No. 30) joined together by top and bottom welded cover plates (320×240×8.1 mm) with a spacing of 0.84–0.87 m along the axes. The trusses are connected by horizontal diagonal bracing members, whose cross-section consists of two angle sections (126×80 mm) joined by cover plates. The length of the metal river span is 53.50 m.

Fig. 4. The general view of the beam cage of channel span 3-4 s

The roadway section of the river span consists of a monolithic reinforced concrete slab with a thickness of approximately 0.17 m, which is placed on top of the beam grid along the transverse beams. The width of the roadway section is 7.0 m. The road surface is made of multi-layered asphalt concrete with a thickness of up to 0.3 m.

In the scope of this study, theoretical dynamic calculations of the main modes of vibration were conducted for the river-spanning superstructure. The calculations took into account the structural peculiarities of the spans and the materials they are made of. Graphical representations of the vibrations with corresponding numerical data were generated, which allowed for determining the resonance frequency of each identified mode of vibration. The obtained results were used for further calculations and comparison with the actual vibrations of the superstructure.

A three-dimensional virtual model of the bridge structure was built, taking into account its geometry, materials and connections. The dimensions, shape, and material properties of each structural element, such as channel girders, supports, and connectors, were specified.

The analysis of the forms and frequencies of natural vibrations in the ANSYS software involves following a specific sequence of actions. Calculations were performed for the first three vibration modes of the river-spanning metal superstructure using ANSYS (see Fig. 5). Table 1 presents the numerical results of the calculations for the first six vibration modes of the river-spanning metal superstructure.

Table 1. Calculations of vibrations for a metal deck truss bridge structure

No oscillation form	1	2	3	4	5
Resonance frequency, fp, Hz	2.51	3.24	3.57	3.96	4.74

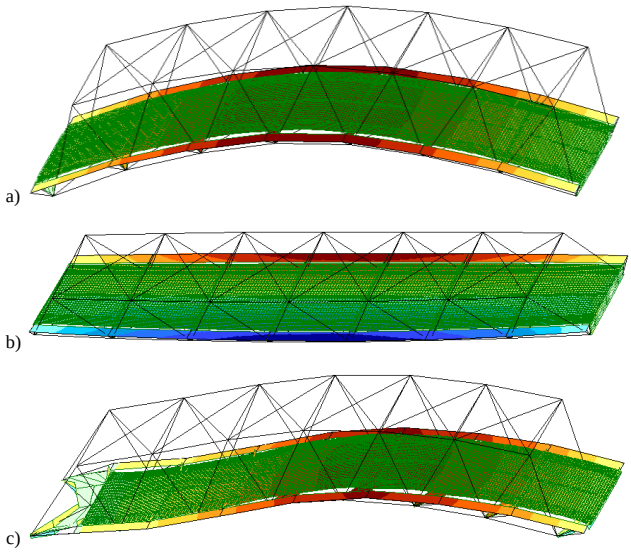


Fig. 5. Calculated forms of vibrations of a metal truss bridge structure

Thus, with the use of modern technologies and research methods, the study of the dynamics of bridge structures becomes more precise, detailed, and efficient. These methods allow for considering complex physical interactions and operational conditions, conducting virtual tests, and optimizing the design to ensure its safety and durability. Such research holds significant importance for engineers and designers, assisting them in making informed decisions during the construction, maintenance, and upgrade of bridge structures, thereby ensuring the reliability and stability of these vital infrastructure assets.

3. Conclusion

This study involved the investigation of the dynamics of bridge structures using modern methods and the ANSYS software package. The main objective was to study the vibrations of the riverbed metal bridge superstructure and determine their characteristics.

During the research, theoretical dynamic calculations of the main modes of vibration were conducted, taking into account the structural features of the superstructure and the materials from which it was constructed. The obtained results allowed for the determination of the resonance frequencies of the vibration modes.

The utilization of the ANSYS software package enabled the creation of a three-dimensional virtual model of the structure and a detailed analysis of its dynamic behaviour. The first three vibration modes were calculated for the riverbed metal bridge superstructure, and numerical results were obtained for six modes.

The research findings hold practical significance as they enable engineers and designers to make informed decisions during the construction, maintenance, and modernization of bridge structures. The study of bridge dynamics using modern technologies contributes to enhancing their safety, reliability, and durability.

Thus, this study provides valuable scientific and practical outcomes that can be utilized in the design and operation of bridge structures, contributing to the development of infrastructure and ensuring the safety of transportation systems. The integration of ANSYS software in the analysis process enhances the accuracy, comprehensiveness, and efficiency of the research, further emphasizing its relevance in the field of bridge engineering.

Acknowledgements

Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V01-00104.

We (also) gratefully acknowledge the financial support by the VEGA Grant Agency of the Slovak Republic, grant number 1/0230/22.

References

[1] Abdulkarem M. et al.: Wireless sensor network for structural health monitoring: A contemporary review of technologies, challenges, and future direction. Structural Health Monitoring 19(3), 2020, 693–735.
[2] Bado M. F., Casas J. R.: A review of recent distributed optical fiber sensors applications for civil engineering structural health monitoring. Sensors 21(5), 2021, 1818.
[3] Chang P. C., Flatau A., Liu S.-C.: Health monitoring of civil infrastructure. Structural health monitoring 2(3), 2003, 257–267.
[4] Chen W.-F., Duan L. (eds): Bridge engineering handbook: construction and maintenance. CRC Press, 2014.
[5] Chopra A. K.: Dynamics of structures: Theory and Applications to Earthquake Engineering. Prentice Hall, 2012.
[6] Cunha A. et al.: Recent perspectives in dynamic testing and monitoring of bridges. Structural Control and Health Monitoring 20(6), 2013, 853–877.
[7] de Sá Caetano E.: Cable vibrations in cable-stayed bridges. IABSE, 2007.
[8] Fujino Y.: Vibration, control and monitoring of long-span bridges—recent research, developments and practice in Japan. Journal of Constructional Steel Research 58(1), 2002, 71–97.
[9] Fukuda Y. et al.: Vision-based displacement sensor for monitoring dynamic response using robust object search algorithm. IEEE Sensors Journal 13(12), 2013, 4725–4732.
[10] Hasegawa S. et al.: Bridge damage detection utilizing dynamic force obtained from moving vehicle acceleration. Bridge Safety, Maintenance, Management, Life-Cycle, Resilience and Sustainability. CRC Press, 2022.
[11] Kim C.-W. et al.: Ambient and vehicle-induced vibration data of a steel truss bridge subject to artificial damage. Journal of Bridge Engineering 26(7), 2021, 04721002.
[12] Kvasnikov V., Stakhova A.: Vibration Measurement Technologies and Systems. Safety in Aviation and Space Technologies: Select Proceedings of the 9th World Congress" Aviation in the XXI Century". Springer International Publishing, 2022.
[13] Oksen E.: Defining the parameters of loading of concrete bridges superstructures basing on the level of vibroacoustic emission signals. Transportation Research Procedia 14, 2016, 3935–3942.
[14] Samali B. et al.: Load rating of impaired bridges using a dynamic method. Electronic Journal of Structural Engineering 1, 2007, 66–75.
[15] Yau J. D., Yang Y. B.: Vibration reduction for cable-stayed bridges traveled by high-speed trains. Finite elements in analysis and design 40(3), 2004, 341–359.
[16] Zhu X. et al.: Damage identification in bridges by processing dynamic responses to moving loads: features and evaluation. Sensors 19(3), 2019, 463.

Ph.D. Eng. Anzhelika Stakhova
e-mail: anzhelika.stakhova@stuba.sk

Researcher of Department of Structural Mechanics, Slovak University of Technology in Bratislava.
Main scientific direction – systems for measuring mechanical quantities, the control and forecasting of the technical condition, dynamic studies of bridge structures.

https://orcid.org/0000-0001-5171-6330

Ph.D. Eng. Adrián Bekö
e-mail: adrian.beko@stuba.sk

Senior lecturer at Department of Structural Mechanics, Slovak University of Technology in Bratislava.
Main scientific direction – nonlinear modelling and analysis of structures, structural dynamics, vibration analysis.

https://orcid.org/0009-0009-3389-0648

http://doi.org/10.35784/iapgos.5626 *received: 30.09.2023 | revised: 21.12.2023 | accepted: 11.02.2024 | available online: 31.03.2024*

ENHANCING CROP HEALTH THROUGH DIGITAL TWIN FOR DISEASE MONITORING AND NUTRIENT BALANCE

Sobhana Mummaneni, Tribhuvana Sree Sappa, Venkata Gayathri Devi Katakam
Velagapudi Ramakrishna Siddhartha Engineering College, Department of Computer Science and Engineering, Vijayawada, India

Abstract. *Digital twins is a digital replica of a physical object to observe its real-time performance, gather data, and recommend corrective actions if required to enhance its performance. This fascinating technological idea is now reaching the agriculture fields to transform farming, by creating digital twins of entire farms. This initiative presents an innovative strategy to enhance crop health and yield by creating a digital twin for paddy fields. The aim is to enable early detection of nutrient deficiencies and leaf blast disease, leading to a transformation in agriculture. Creating virtual replicas of plants and fields, the digital twin harnesses real-time data and advanced analytics to transform the way agricultural systems are managed. By integrating remote sensing, data analytics, and various Internet of Things devices like pH, nitrous, potassium, and phosphorus sensors, coupled with a gateway system, the digital twin provides real-time monitoring and analysis of crop health and nutrient levels. Employing advanced machine learning algorithms, notably Convolutional Neural Networks ensures precise and early detection of nutrient deficiencies and crop diseases. This ground-breaking technology provides timely alerts and actionable insights to farmers, enabling proactive decision-making for optimal crop management. This farmland digital twin represents a transformative approach towards agricultural sustainability and enhancing productivity.*

Keywords: agricultural sustainability, convolution neural networks, digital twin, internet of things, nutrient deficiency detection

POPRAWA ZDROWIA UPRAW ZA POMOCĄ CYFROWEGO BLIŹNIAKA DO MONITOROWANIA CHORÓB I BILANSU SKŁADNIKÓW ODŻYWCZYCH

Streszczenie. *Cyfrowe bliźniaki to cyfrowa replika obiektu fizycznego, która umożliwia obserwację jego działania w czasie rzeczywistym, gromadzenie danych i rekomendowanie działań naprawczych, jeśli jest to wymagane w celu poprawy jego wydajności. Ta fascynująca koncepcja technologiczna dociera obecnie do dziedzin rolnictwa, aby przekształcić rolnictwo, tworząc cyfrowe bliźniaki całych gospodarstw. Inicjatywa ta przedstawia innowacyjną strategię mającą na celu poprawę zdrowia i plonów upraw poprzez stworzenie cyfrowego bliźniaka pól ryżowych. Celem jest umożliwienie wczesnego wykrywania niedoborów składników odżywczych i zarazy liści, co doprowadzi do transformacji rolnictwa. Tworząc wirtualne repliki roślin i pól, cyfrowy bliźniak wykorzystuje dane w czasie rzeczywistym i zaawansowane analizy, aby zmienić sposób zarządzania systemami rolniczymi. Dzięki integracji teledetekcji, analizy danych i różnych urządzeń Internetu rzeczy, takich jak czujniki pH, azotu, potasu i fosforu, w połączeniu z systemem bramek, cyfrowy bliźniak zapewnia monitorowanie i analizę stanu upraw i poziomów składników odżywczych w czasie rzeczywistym. Zastosowanie zaawansowanych algorytmów uczenia maszynowego, w szczególności konwolucyjnych sieci neuronowych, zapewnia precyzyjne i wczesne wykrywanie niedoborów składników odżywczych i chorób upraw. Ta przełomowa technologia zapewnia rolnikom aktualne alerty i przydatne informacje, umożliwiając proaktywne podejmowanie decyzji w celu optymalnego zarządzania uprawami. Ten cyfrowy bliźniak pól uprawnych reprezentuje transformacyjne podejście do zrównoważonego rozwoju rolnictwa i zwiększania produktywności.*

Słowa kluczowe: zrównoważony rozwój rolnictwa, konwolucyjne sieci neuronowe, cyfrowy bliźniak, internet rzeczy, wykrywanie niedoboru składników odżywczych

Introduction

The need for food and agriculture is paramount, with rice serving as the primary staple for almost half of the global population and growing across all six continents, including Asia, Africa, North America, South America, Europe and Australia except the frigid continent, Antarctica [20]. In the previous year (2022/23), global rice production reached 513.68 million tons, and projections for this year estimate a potential rise to 518.14 million tons, reflecting the constant effort to meet increasing demand [18]. India, cultivating rice across 43 million hectares, faces the challenge of maintaining an average productivity of 2.6 tons per hectare [19].

The impressive yield potential of rice relies heavily on the availability of key nutrients, with nitrogen (N), phosphorus (P), and potassium (K) playing vital roles in every stage of growth. These elements serve as the lifeblood of healthy rice plants, directly impacting various life cycles, including sprouting, leafing, blooming, fruiting, and reproducing [22]. Nitrogen, a crucial nutrient, accelerates crop growth and enhances grain yield and quality. Phosphorus, another essential nutrient, serves various functions in crops, including involvement in photosynthesis, respiration, energy storage and transfer, cell division and enlargement, and internal operations. Potassium, as the third essential nutrient, is absorbed by crops in the form of K+ ions from the soil. It serves as an activator for numerous enzymes involved in various crop metabolic processes, including photosynthesis [14].

Rice cultivation, despite its significance as a global staple, faces formidable challenges due to diseases, presenting biotic hurdles that can result in yield reductions ranging from 20% to 100% [19]. Chief among these challenges is rice blast, triggered by the fungus Magnaporthe oryzae (formerly Magnaporthe grisea) standing out as the most impactful disease that inflicts annual yield losses ranging from approximately 10% to 30%. Under favourable conditions, this disease has the potential to rapidly devastate entire rice plants within 15 to 20 days, resulting in staggering losses of up to 100% [2]. The rice blast fungus, with its ability to infect and create lesions on a significant portion of the crop, progresses through various stages, commencing with leaf blast and subsequently involving collar, panicle, and node blast [16]. In the initial stages of growth, symptoms of rice blast primarily manifest on leaves referred to as leaf blast, posing a significant threat to the crop [4]. A leaf blast infection, with the potential to eliminate seedlings or plants up to the tillering stage, poses a severe threat and can lead to significant yield losses, emphasizing the critical need for effective disease management strategies in rice cultivation [5].

Routine visual assessments of rice plants stand out as one of the easiest and most efficient methods for detecting leaf blasts and nutrient deficiencies early. This process can be further enhanced with the assistance of a digital twin, offering real-time monitoring and a visual representation of the field to facilitate timely disease detection and management.

Digital twins, representing virtual counterparts of physical entities, go beyond mere representation by incorporating real-time monitoring and data integration. Through the utilization of real-time data collection, processing, and analysis, Digital Twins (DTs) provide a comprehensive digital portrayal of physical systems. This capability enables meticulous monitoring and prediction of both current and future states. It allows for the improvement of existing models and the reevaluation of systems and procedures, thus playing a pivotal role in the evolution of smart agricultural systems [15].

The characteristic feature of Digital Twins lies in the bidirectional exchange of data between tangible reality and its virtual representation. The Digital Field Twin, a key component, not only stores historical data but also serves as an interface for accessing

current sensor data, including satellite images and weather data [3]. Digital twin creation is rooted in Internet of Things (IoT) devices and sensors, capturing extensive data from the physical farming environment. These sensors collect diverse data points, ranging from soil moisture levels and temperature variations to nutrient levels, crop health indicators, and machinery performance metrics. This data serves as a foundational element for agriculture digital twins, where advanced algorithms and modelling techniques analyse and interpret the information, generating a real-time reflection of the current state of the agricultural system.

In navigating the sensitive landscape of agricultural data, prioritising data security and privacy becomes paramount. Robust encryption, access controls, and strict adherence to data protection regulations effectively address these concerns. Additionally, achieving seamless integration of physical and digital systems necessitates standardised protocols, IoT integration, and middleware solutions.

At the core of this agricultural revolution is the seamless integration of digital twin technology, machine learning algorithms, and sensor data. These components synergise to offer an innovative approach to monitoring and managing nutrient deficiencies and diseases in paddy crop fields. The forthcoming sections of this paper will delve into the methodology employed in the development of the digital twin, present results and findings, and conclude with a discussion of the implications and prospects of this pioneering research.

1. Literature review

Anton Terentev et al. [21] delved into the synergy of metabolomic approaches with hyperspectral remote sensing, Raman spectroscopy, and mass spectrometry for enhanced early plant disease detection. The review navigates through proven data acquisition techniques, showcasing the potential of metabolomics in optimizing contemporary methods. The system faces challenges in achieving stability in ultra-sensitive remote sensing, particularly in hyperspectral and Raman spectrometry. The complexity and cost of advanced sensors like GC-MS and LC-MS are limiting factors.

Abbas et al. [1] explored the pivotal role of drones in plant disease assessment, emphasising the drones efficiency in monitoring and detection for smart agriculture. It utilizes LiDAR, SfM photogrammetry, multispectral, and thermal cameras to capture plant morphological information and detect diseases. The research underscores the effectiveness of deep learning models over traditional methods in disease classification. Susceptibility to environmental conditions affecting drone functionality is a key limiting factor.

Han Yih Lau et al. [11] presented DNA-based point-of-care diagnostic methods for detecting plant diseases, emphasizing characteristics like specificity, sensitivity, and multiplexing. While PCR is prevalent, isothermal techniques like LAMP and RCA offer advantages for field applications with constant temperature operation. In summary, nucleic acid-based methods offer higher specificity, with isothermal techniques presenting potential solutions, and advanced technologies like NGS requiring further refinement for practical point-of-care applications. Challenges include size variations in amplification products, and NGS, while promising, faces obstacles like complex equipment and high costs.

Bravo et al. [6] investigated the potential of spectral reflectance to differentiate between healthy and Puccinia striiformis-infected wheat plants. In-field spectral images, taken with a spectrograph at spray boom height, underwent normalization for reflectance and illumination adjustments. Leveraging a quadratic discrimination model on selected wavebands significantly reduced confusion rates from 12% to 4%. Simplifying the system while maintaining a 96% success rate, the research enhances understanding of disease detection in wheat fields. The method's reliance on ambient illumination conditions poses a potential challenge, as it may exhibit sensitivity to variations in weather and lighting situations.

Zhe Xu et al. [23] investigated the efficacy of Deep Convolutional Neural Networks (DCNNs) for the diagnosis of nutrient deficiencies through image classification in rice plants. By utilizing 1818 leaf photographs obtained from hydroponic experiments, encompassing a spectrum of nutritional conditions, the researchers fine-tuned four DCNNs: NasNet-Large, Inception-v3, DenseNet121 and ResNet50. Impressively, all DCNNs achieved accuracy levels exceeding 90%, with DenseNet121 emerging as the top performer (with a validation accuracy of 98.62% and test accuracy of 97.44%). Notably, the DCNNs outperformed traditional methods such as colour features with SVM and HOG with SVM.

Anu Jose et al. [8] delved into nutrient deficiency detection in tomato plants using neural networks. The study employed artificial neural networks to classify nutrient deficiencies in tomato plants, analysing leaf characteristics such as colour and shape. Two segmentation schemes, hue-based and threshold-based, were compared, with hue-based segmentation outperforming and achieving an 88% accuracy. The study's reliance on user identification, primarily farmers, for specific nutrient deficiency input hinders automatic detection and may introduce human errors.

Hazem M. Kalaji et al. [9] explored the utility of chlorophyll fluorescence parameters for early detection of nutrient deficiencies in rapeseed plants. Using 60 soil samples and rapeseed plants, the study employs principal component analysis, hierarchical k-means, and machine-learning methods to identify distinct groups representing different nutrient deficiency levels. The results reveal adverse effects on the photosynthetic machinery in nutrient-deficient groups, emphasizing the potential of chlorophyll fluorescence combined with machine learning for early detection. However, the research relies on soil samples from a specific geographic location and soil characteristics can vary significantly across regions, impacting the transferability of the study's findings to diverse soil compositions.

M. V. Latte et al. [10] focused on a rule-based approach to detect nutrient deficiencies utilizing HSV (Hue, Saturation, Value) colour features in paddy crops. Using HSV colour features and rigorous experimentation to establish rules, the approach achieves impressive accuracy – 100% for healthy leaves and an overall 95.39% for nitrogen, phosphorus, and potassium deficiencies. A potential drawback of the rule-based approach is its reliance on predefined rules, limiting adaptability to real-world variations. It may struggle with ambiguous symptoms and environmental factors, requiring constant rule adjustments for improved robustness and effectiveness.

Anshuman Nayak et al. [12] explored the realm of leveraging image processing and transfer learning techniques, specifically applied to the detection of rice diseases and nutrient deficiencies using smartphone images. Leveraging 2259 smartphone images, the research adeptly classified 12 rice diseases and deficiencies, pinpointing MobileNetV2 as the optimal CNN model for the 'Rice Disease Detector' Android application. Despite promising results, a notable limitation is that the user or farmer must detect the disease or deficiency, as the app focuses solely on identification and does not offer automatic detection.

Jesus David Chaux et al. [7] presented a Digital Twin Architecture designed to enhance productivity in Controlled Environment Agriculture. Utilizing simulation software, the framework enhances climate control and crop management, showcasing bidirectional communication in a greenhouse prototype. Noteworthy advantages include scalability for industrial deployment and replicability in educational settings. An additional advantage is the continuous monitoring capability enabled by the digital twin, providing real-time insights for prompt decision-making in agricultural practices. However, a potential limitation lies in the exclusive focus on controlled environments, urging future research to extend the digital twin concept to open-field agriculture.

2. Proposed methodology

The methodology employed in developing a digital twin for the paddy field focuses on guaranteeing robustness, accuracy, and user-friendly crop management tools in the designed system architecture. It begins with rigorous data handling, encompassing collection, cleaning, and augmentation. A hybrid CNN-EfficientNetB1 model is designed and trained with an Adam optimizer and key metrics for evaluation. Simultaneously, a dynamic 3D digital twin is crafted, integrating real-time sensors for effective agricultural monitoring. Fig. 1 shows the process flow of this methodology.

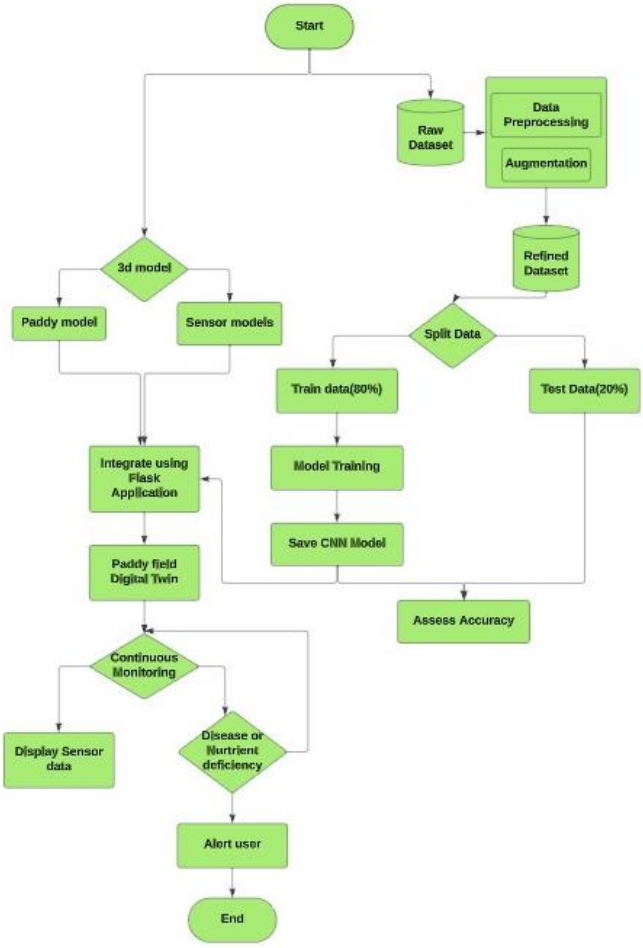


Fig. 1. Process Flow model

2.1. Data collection and preprocessing

The data collection and preprocessing procedures ensured the integrity and relevance of the dataset, comprising essential agricultural images. Two datasets were considered, both sourced from Kaggle [13, 17]. The first dataset focuses on nutrient deficiency symptoms and is organised into three folders, each representing a specific deficiency type: nitrogen (n) with 440 images, phosphorus (p) with 333 images, and potassium (k) with 383 images. The second dataset comprises 779 images depicting leaf blast disease and 1,488 images containing healthy leaves. A careful data cleaning process was carried out, involving tasks like handling missing data, duplicate removal, label consistency checks, and noise filtration. This data cleaning improved the overall quality of the dataset, providing a strong foundation for subsequent model training. The resulting dataset is categorised into distinct classes, including "Leaf Blast," "Nutrient Deficiency," and "Healthy," effectively encapsulating the diversity of conditions observed in paddy fields.

2.2. Data augmentation

Data augmentation, including random flips and rotations, was applied to diversify the dataset, improve the model's ability to generalize and mitigate the risk of overfitting.

2.3. Data loading and processing

In the data loading and processing stage, the process begins with loading the pre-trained model architecture and weights for nutrient deficiency and disease detection using the Keras library. This involves storing the model in Hierarchical Data Format (H5) and JSON files, enabling easy retrieval for subsequent predictions. An essential aspect of this stage is the implementation of an image-processing function. This function is designed to ensure that input images are appropriately resized, converted, and rescaled to meet the model's specifications. The ImageDataGenerator class from Keras facilitates the creation of data generators for both the training and validation sets. These generators play a crucial role in enhancing the model's ability to generalize by augmenting the training data with various transformations. Constants such as image size, batch size, epochs, and the number of classes are defined to maintain consistency throughout. Overall, the data loading and processing stage is fundamental for preparing the dataset, configuring the model, and establishing a robust pipeline for subsequent training and evaluation steps.

2.4. Model creation

In Model Creation, an image classification framework was built. A Convolutional Neural Network (CNN) was strategically chosen, harnessing its well-established proficiency in image classification tasks. This decision aligns with the objective of achieving robust predictions in diverse scenarios. Furthermore, to expedite the learning process and leverage prior knowledge encoded in large datasets, pre-trained models were seamlessly integrated. This integration encompassed the deployment of EfficientNetB1, recognised for its superior image classification capabilities, along with a custom-designed CNN. Incorporating these pre-trained models not only capitalizes on their domain expertise but also ensures a versatile and efficient framework capable of handling a spectrum of prediction tasks. The CNN model comprises Convolutional layers (Conv2D) for feature extraction, MaxPooling layers for downsampling and retaining essential information, and Dropout layers to prevent overfitting during training. Flattening layers convert the 2D feature maps into a 1D vector, preparing the data for Dense layers that contribute to the final classification. This combination of layers plays a crucial role in discerning intricate patterns in agricultural images, enabling the model to efficiently classify and identify nutrient deficiencies and diseases in plants.

The integration of these layers optimizes the model for image analysis tasks, enhancing its capability to recognize diverse conditions in agricultural settings. The layers in the model, encompassing both the Convolutional Neural Network (CNN) and EfficientNetB1, employ Rectified Linear Unit (ReLU) activation functions. Specifically, these functions are applied to convolutional layers, ensuring non-linearity in the model's learning. For multi-class classification, categorical cross-entropy serves as the chosen loss function, complemented by the Adam optimizer during compilation. To facilitate effective convergence, a judicious learning rate, typically set between 0.001 to 0.01, was selected. The inclusion of Batch Normalization contributes to stabilizing and expediting training by normalizing layer inputs, enhancing overall model performance. In the final compilation step, the model is configured with the specified optimiser, loss function, and metrics, laying the foundation for efficient training and robust predictions.

2.5. Model training

In the model training phase, mini-batch training is implemented with a batch size of 32 for both the training and validation sets, optimizing computational efficiency. The mini-batch sizes for both the training and validation sets are determined using a data generator, established through the 'flow_from_directory' method, an inbuilt method provided by the Keras library within the 'ImageDataGenerator' class. Throughout the training, the Adam optimizer is employed to minimize the categorical cross-entropy loss, a key factor in updating the model's weights effectively. The advancement of training is tracked by assessing accuracy as a metric, offering valuable insights into the model's proficiency in generating precise predictions. The inclusion of the 'ModelCheckpoint' callback ensures that the model weights are saved when the validation loss is at its minimum, preserving the best-performing configuration. This methodology ensures not only the efficient learning of underlying patterns but also the preservation of the optimal model state for subsequent use. The visual analysis of training and validation accuracy and loss over epochs further enhances the understanding of the model's performance dynamics. As a final step, the trained model is saved for future deployment, enabling seamless predictions of new data with the acquired knowledge.

2.6. Model evaluation

In the evaluation phase, various metrics are employed to comprehensively assess the performance of the model. The 'Accuracy' metric measures the overall correctness of the model's predictions. Precision gauges the accuracy of positive predictions, this metric is particularly relevant in scenarios where precision is of utmost importance. 'Recall', also known as sensitivity, signifies the model's ability to capture all relevant instances, providing insights into the model's ability to identify true positive cases. The 'F1 Score' acts as a harmonic mean between precision and recall, providing a balanced assessment of the model's performance. It is especially useful when there is a need to balance precision and recall in the evaluation. The 'Confusion Matrix' provides a tabular summary, offering a comprehensive breakdown of the model's performance across different classes, aiding in understanding potential areas for improvement. The 'Loss Function' value during training and validation reflects the model's learning performance, while 'Validation Accuracy' signifies the model's accuracy on the validation set during training. These metrics collectively furnish a holistic understanding of the model's effectiveness and areas for potential improvement. The specific formulas below provide insight into the quantitative measurement of different aspects of the model's predictive capabilities.

ALGORITHM 1: LEAF BLAST AND NUTRIENT DEFICIENCY DETECTION

Input: Pre-processed dataset of leaf images.
Output: Trained model for detection of Leaf Blast and Nutrient Deficiencies.
1) Import libraries: TensorFlow, Keras, NumPy, Matplotlib, scikit-learn.
2) Load and process image data using ImageDataGenerator.
3) Split data into training and validation sets.
4) Build and compile a CNN model with Adam optimiser.
5) Train the model and save it.

ALGORITHM 2: DIGITAL TWIN

Input: Crop images and sensor data.
Output: Digital Twin Display.
1) Create an html template.
2) Include hotspots for various sensors for data retrieval and implement routes to update sensor visuals.
3) Import necessary libraries and modules into the Flask app.
4) Load trained detection model from JSON and H5 files.
Implement routes to detect nutrient deficiencies and leaf blasts using the prediction function from the loaded model on crop images.

$$Accuracy \leftarrow \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \quad (1)$$

$$Precision \leftarrow \frac{True\ Positives}{True\ Positives+False\ Positives} \quad (2)$$

$$Recall \leftarrow \frac{True\ Positives}{True\ Positives+False\ Negatives} \quad (3)$$

$$F1\ score \leftarrow 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

2.7. Digital twin

In the initial phase of the digital twin process, a comprehensive 3D model of the paddy field is crafted, capturing its layout and features. This model serves as the core component of the complete digital twin, incorporating strategically located hotspots representing sensors like pH, EC, humidity, moisture, NPK, and cameras. These sensors actively gather real-time data from the physical field, ensuring an accurate representation within the digital twin. The seamless integration of this data reflects the ever-changing conditions in the agricultural environment.

The digital twin's disease and nutrient deficiency detection capabilities are powered by an integrated model that combines Convolutional Neural Network (CNN) and EfficientNetB1. Captured images from field cameras play a pivotal role in facilitating the identification of key issues and providing timely insights for informed decision-making. Hotspot annotations in the digital twin dynamically reflect sensor status, offering at-a-glance updates on the rice crop's well-being. A user-friendly colour-coded system, with red indicating anomalies and green for normal conditions, ensures intuitive interpretation.

These hotspots, strategically positioned in the digital twin, act as informative markers, allowing users to quickly identify areas of concern and monitor specific aspects. Clicking on hotspots triggers script executions, providing detailed information about sensor readings.

Moreover, the digital twin doesn't stop at static representation; it incorporates dynamic elements like changing colours and annotations to convey the real-time nature of agricultural data. This responsiveness enhances the user experience, making the monitoring process engaging, insightful, and aligned with the dynamic nature of agriculture. The system is designed to be a user-friendly tool, enabling farmers to efficiently monitor and manage their crops.

3. Results

3.1. Training accuracy and loss graphs

The training accuracy and loss graphs provide insights into the learning progress of the combined CNN and EfficientNetB1 model over epochs. The accuracy of the model exhibits consistent improvement throughout the training process, signifying its adeptness in learning and generalizing from the provided dataset. The loss values follow a consistent trend, steadily decreasing throughout training, indicating effective convergence. Notably, the validation loss remains relatively low, affirming the model's ability to generalize well to unseen data. Fig. 2 and 3 respectively show the training accuracy and loss over epoch curves for nutrient deficiency and leaf blast detection.

Fig. 2. Nutrient deficiency detection training accuracy and loss over epochs

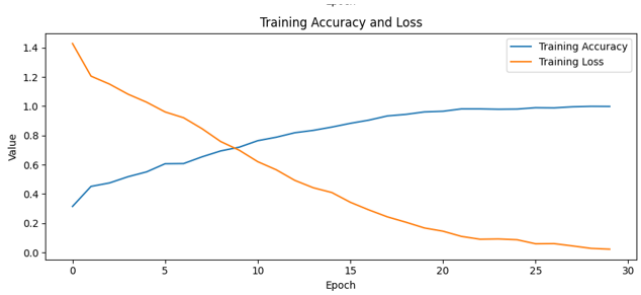


Fig. 3. Disease detection training accuracy and loss over epochs

3.2. Confusion matrix

The confusion matrix in Fig. 4, and 5 respectively illustrate the model's performance in detecting nutrient deficiencies and leaf blast in the rice crop.

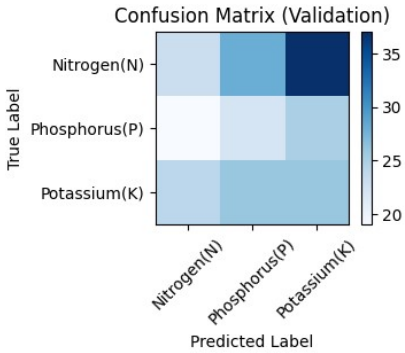


Fig. 4. Nutrient detection confusion matrix

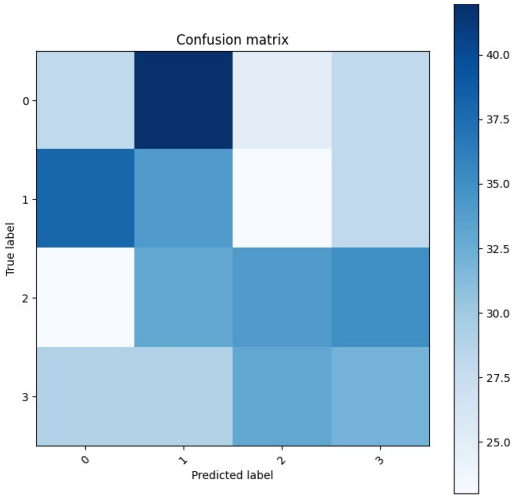


Fig. 5. Disease detection confusion matrix

3.3. Digital twin

The digital twin excelled in providing real-time insights for precision agriculture. The 3D model accurately represented the paddy field, integrating live data from strategically placed sensors. The disease and nutrient deficiency detection model, powered by CNN and EfficientNetB1, demonstrates robust performance, enabling timely decision-making in crop management. The following Fig. 6, 7 and 8 are the images of the paddy field's digital twin.

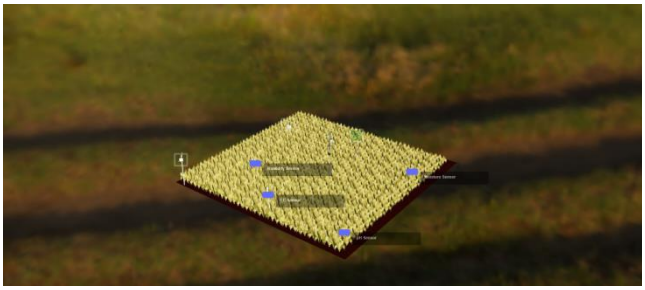


Fig. 6 . Digital Twin for Paddy Field

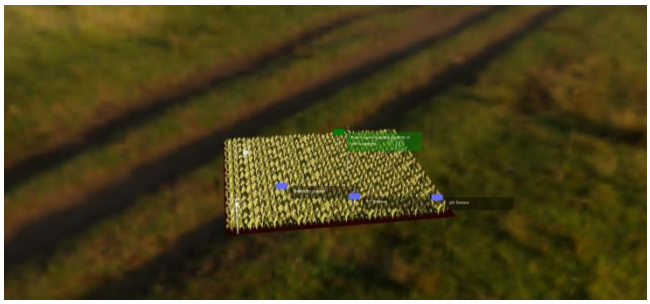


Fig. 7. Digital twin showing Healthy annotation

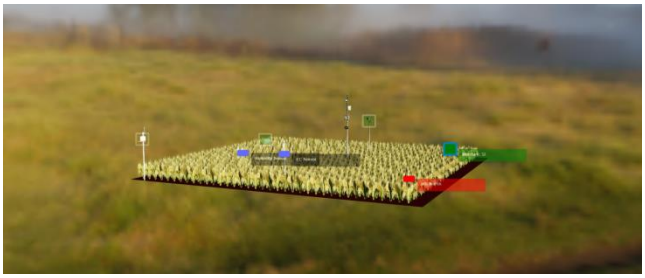


Fig. 8. Digital twin showing Imbalance of pH value

4. Conclusion

In conclusion, the digital twin of the paddy field transforms rice crop management by offering farmers the ability to monitor their fields in near real-time, providing constant insights for informed decision-making. The early detection of nutrient deficiency and leaf blast disease is a pivotal outcome, empowering farmers to address issues promptly. The machine learning model, comprising a combination of Convolutional Neural Network (CNN) and EfficientNetB1, forms the backbone of this innovative approach. Integrated into a Digital Twin framework, the model enhances the comprehensive understanding and monitoring of the agricultural landscape. This study holds the potential for expansion to encompass various crops and diseases, broadening its applicability and impact on diverse agricultural contexts. The significance of the digital twin is in its ability to enhance crop monitoring, contributing significantly to sustainable and efficient agricultural practices.

References

[1] Abbas A. et al.: Drones in Plant Disease Assessment, Efficient Monitoring, and Detection: A Way Forward to Smart Agriculture. Agronomy 13(6), 2023, 1524.
[2] Asibi A. E., Chai Q., Coulter J. A.: Rice blast: A disease with implications for global food security. Agronomy 9(8), 2019, 451.
[3] Awan J.: Digital Twins for Agriculture - Blog Des Fraunhofer IESE. Fraunhofer IESE, 25 Nov. 2020 [www.iese.fraunhofer.de/blog/digital-twins-agriculture] (avaible 29.09.2023).
[4] Bastiaans L.: Effects of leaf blast on growth and production of a rice crop: 1. Determining the mechanism of yield reduction. Netherlands Journal of Plant Pathology 99, 1993, 323–334.
[5] Blast (Leaf and Collar), IRRI Rice Knowledge Bank. [www.knowledgebank.irri.org,www.knowledgebank.irri.org/training/fact-sheets/pest-management/diseases/item/blast-leaf-collar] (avaible 29.09.2023).
[6] Bravo C. et al.: Early disease detection in wheat fields using spectral reflectance. Biosystems Engineering 84(2), 2003, 137–145.
[7] Chaux J. D. et al.: A digital twin architecture to optimize productivity within controlled environment agriculture. Applied Sciences 11(19), 2021, 8875.
[8] Jose A. et al.: Detection and classification of nutrient deficiencies in plants using machine learning. Journal of Physics: Conference Series 1850(1), 2021.
[9] Kalaji H. M. et al.: Chlorophyll fluorescence as a tool for nutrient status identification in rapeseed plants. Photosynthesis Research 136, 2018, 329–343.
[10] Latte M. V., Shidnal S., Anami B. S.: Rule based approach to determine nutrient deficiency in paddy leaf images. International Journal of Agricultural Technology 13(2), 2017, 227–245.
[11] Lau H. Y., Botella J. R.: Advanced DNA-based point-of-care diagnostic methods for plant diseases detection. Frontiers in plant science 8, 2017.
[12] Nayak A. et al.: Application of smartphone-image processing and transfer learning for rice disease and nutrient deficiency detection. Smart Agricultural Technology 4, 2023, 100195.
[13] Nutrient-Deficiency-Symptoms-In-Rice. [www.kaggle.com/datasets/guy007/nutrientdeficiencysymptomsinrice] (avaible 27.09.2023).
[14] Paiman J. et al.: Maximizing the Rice Yield (Oryza Sativa L.) Using NPK Fertilizer. The Open Agriculture Journal 15(1), 2021, 33–38, [https://doi.org/10.2174/1874331502115010033].
[15] Peladarinos N. et al.: Enhancing smart agriculture by implementing digital twins: A comprehensive review. Sensors 23(16), 2023, 7128.
[16] Rice Blast, Rice, Agriculture: Pest Management Guidelines. UC Statewide IPM Program (UC IPM) [ipm.ucanr.edu/agriculture/rice/rice-blast] (avaible 29.09.2023).
[17] Rice Diseases Image Dataset [www.kaggle.com/datasets/minhhuy2810/rice-diseases-image-dataset] (avaible 29.09.2023).
[18] Rice Production by Country. World Agricultural Production 2023/2024 [www.worldagriculturalproduction.com/crops/rice.aspx] (avaible 29.09.2023).
[19] Shivappa R. et al.: Emerging minor diseases of rice in India: losses and management strategies. Integrative Advances in Rice Research, 2021.
[20] Talukder Md S. H. et al.: An Improved Model for Nutrient Deficiency Diagnosis of Rice Plant by Ensemble Learning. 4th International Conference on Sustainable Technologies for Industry 4.0 (STI). IEEE, 2022.
[21] Terentev A., Dolzhenko V.: Can Metabolomic Approaches Become a Tool for Improving Early Plant Disease Detection and Diagnosis with Modern Remote Sensing Methods? A Review. Sensors 23(12), 2023, 5366.
[22] Wang C. et al.: Classification of nutrient deficiency in rice based on CNN model with Reinforcement Learning augmentation. International Symposium on Artificial Intelligence and its Application on Media (ISAIAM). IEEE, 2021.
[23] Xu Z. et al.: Using deep convolutional neural networks for image-based diagnosis of nutrient deficiencies in rice. Computational Intelligence and Neuroscience, 2020, 7307252.

Prof. Sobhana Mummaneni
e-mail: sobhana@vrsiddhartha.ac.in

Dr. Sobhana Mummaneni is currently working as an associate professor in the Department of Computer Science and Engineering, V. R. Siddhartha Engineering College, Vijayawada, India. She received Ph.D. degree in Computer Science and Engineering in 2018 from Krishna University. She has 16 years of teaching experience. Her research interests lie in areas such as artificial intelligence, machine learning, data analytics, cyber security, and software engineering. She published 35 papers in National and International journals and published 7 patents.

https://orcid.org/0000-0001-5938-5740

Eng. Tribhuvana Sree Sappa
e-mail: tribhuvanasree@gmail.com

Tribhuvana Sree Sappa is a third-year B.Tech. student specializing in Computer Science and Engineering at V. R. Siddhartha Engineering College, Vijayawada, India. She is passionate about machine learning, the Internet of Things and web programming.

https://orcid.org/0009-0009-8594-429X

Eng. Venkata Gayathri Devi Katakam
e-mail: gayathridevikatakam1@gmail.com

Venkata Gayathri Devi Katakam is a third-year B.Tech. student specializing in Computer Science and Engineering at V. R. Siddhartha Engineering College, Vijayawada, India. Her interests lie in the fields of Internet of Things, machine learning and web programming.

https://orcid.org/0009-0002-3492-4386

http://doi.org/10.35784/iapgos.5572 *received: 15.11.2023 | revised: 16.01.2024 | accepted: 11.03.2023 | available online: 31.03.2024*

REVIEW OF MODELLING APPROACHES FOR WEBSITE-RELATED PREDICTIONS

Patryk Mauer
Opole University of Technology, Opole, Poland

Abstract. *This paper researches various modelling approaches for website-related predictions, offering an overview of the field. With the ever-expanding landscape of the World Wide Web, there is an increasing need for automated methods to categorize websites. This study examines an array of prediction tasks, including website categorization, web navigation prediction, malicious website detection, fake news website detection, phishing website detection, and evaluation of website aesthetics.*

Keywords: machine learning, web sites, prediction methods, classification algorithms

PRZEGLĄD PODEJŚĆ DO MODELOWANIA PRZEWIDYWAŃ ZWIĄZANYCH Z WITRYNAMI INTERNETOWYMI

Streszczenie. *Ten artykuł naukowy przeprowadza analizę różnorodnych metod modelowania stosowanych do prognozowania aspektów witryn internetowych, zapewniając przegląd tej dynamicznie rozwijającej się dziedziny. Podczas gdy Internet nieustannie się powiększa, nabiera wagi potrzeba stosowania automatycznych metod do klasyfikacji nowo powstających stron internetowych. Zbadano metody zastosowane w szerokim zakresie przewidywań, obejmujących kategoryzację witryn internetowych, prognozowanie zachowań nawigacyjnych użytkowników online, identyfikację stron o złośliwym charakterze, wykrywanie fałszywych informacji, rozpoznawanie prób phishingu oraz ocenę estetycznych aspektów witryn internetowych.*

Słowa kluczowe: uczenie maszynowe, witryny internetowe, metody przewidywania, algorytmy klasyfikacji

Introduction

The estimated total number of existing websites in the World Wide Web exceeded 2 billion stated by Statsfind.com [4]. The same platform states that there are 252 thousand websites created everyday worldwide which gives 175 websites every minute. These statistics only showcase the need of automated ways to indexing and categorization of these websites. Content categorization falls into use cases of machine learning or probabilistic models as they can be trained to automatically categorize or classify content into predefined categories based on the content's characteristics, features, or patterns. Techniques needed for achieving a success in these fields have been developed by researchers generally in purpose of cracks and malicious websites detection [2], web navigation prediction [5], fake news detection [1], Search Engine Optimization [11]. All the purposes of websites or generally content classification utilized different methods of data acquisition and preprocessing, feature extraction and machine learning algorithms. This paper covers the need of a structured review of the techniques and approaches used particularly for website categorization.

1. Modelling approaches

The following chapters are structured in the way that each point in the chapters number lists corresponds to a particular approach used by researchers. The numeration is consistent throughout the paper. The table 1 shows the corresponding data.

2. Data acquisition and preprocessing

Data acquisition and preprocessing phases differed significantly from each other and are presented in the following subchapters.

2.1. Data sources and sizes

The data in the examined research papers was gathered from the following sources in the following sizes:

1) Manually labelled set of 450 web pages that was uniformly distributed among six categories. In total, the pages contained 3890 images.
2) Publicly available web log datasets that include Cyber Threat Intelligence (CTI) Dataset: DePaul University web server logs from April 2002, with 13,745 sessions, 683 pages, and 16 categories; Microsoft Anonymous Web Data (MSWEB) Dataset: Microsoft logs from February 1998 with 38,000 random users and 294 unique Vroots (virtual directories – a feature that allows the web server to serve content from a directory other than the server's root directory); BMS WebView1 Dataset: E-commerce website logs from Gazelle.com with 59,601 sequences and 497 distinct items.
3) The experimental data comprised: (i) the "UK Web Archive" that contributed an initial dataset consisting of 14922 categorized websites. (ii) A manually gathered list of 510 URLs containing malware content to enhance the dataset's coverage of malicious sites. (iii) Dataset created utilizing the "Google Safe Browsing" API to classify and label websites, including the identification of "Malware" and other safety-related categories.

Table 1. Corresponding numbers of points in numbered lists to researchers' papers and goals of the prediction. Indication on how to read the further parts of the paper

No.	Goal of the prediction	Researchers
1	Website categorization – the proces of assigning a website to predefined categories.	Nandanwar A., Choudhary J., 2020 [8]
2	Web Navigation Prediction – the anticipation of the next web page or resource that a user is likely to access during their online browsing session.	Jindal H., Sardana N., 2022 [5]
3	Crack and Malicious Website detection – the forecast of whether a website is engaged in distributing illegal software.	Cohen D., Naim O., Toch E., Ben-Gal, 2021 [2]
4	Fake News Website detection – the classification of whether a website falls under the category of authentic or fabricated news providers.	Bozarth L., Budak C., 2019 [1]
5	Phishing Website detection – the categorization of a website as either legitimate or a potential phishing threat	Opara C., Chen Y., Wie B., 2023 [7]
6	Evaluation of website aesthetics – the prediction of the mean subjective user score regarding the design aesthetics of a website.	Delitzas A., Chatzidimitriou K.C., Symeonidis A.L., 2023 [11]
7	Web Page Classification into low, medium and high degree of adjustment to Search Engine Optimization (SEO) guidelines.	Matošević G., Dobša J., Mladenić D., 2021 [10]
8	Predicting users' intention of potential online purchase.	Sakar C.O., Polat S.O., Katircioglu M., 2018 [9]

4) The data source comprises (i) Lists of Fake and Traditional News Sites: Collected by Bozarth et al. in 2019, resulting in 1800 unreliable news sites and 8200 reliable news sites (ii) Homepage and Subpages: Collected using scrapy for 7600 homepages, followed by recursive crawling of subpages, resulting in 2 million distinctive subpages. (iii) News Article Webpages and Tweets: obtained from Bozarth et al. (2019), including 170 thousand unique articles shared in 700 thousand Tweets.
5) Data sources encompass Alexa.com for genuine websites, Phishtank.com for fraudulent web pages, and a set of 22 thousand web pages collected in Korkmaz et al. (2020) [6].
6) A publicly accessible dataset collected by researchers in 2014 includes 398 webpage screenshots with aesthetics ratings on a scale of 1 to 9. Additionally, a second dataset, generated by the authors, consists of 100 webpages randomly selected from Alexa's top 5000 websites and includes pairwise comparisons of web designs obtained through crowdsourcing via a web application.
7) The data in the research is a random sample of 600 pages extracted DMOZ – open directory of web pages maintained by volunteers. These pages were manually classified by three independent SEO experts into low, medium and high adjustment level to SEO guidelines. A Python script was used to extract website features and keywords.
8) The data was curated with emphasis on its diversity in order not to show any tendency related to marketing campaigns or particular days. The dataset consisted of over 12 thousand user navigation sessions with restriction that each session belong to a unique user. Over 15% of the dataset contained user sessions that ended with a purchase. Ten features were numerical that mainly consisted of spent time on defined websites, number of visited pages and bounce and exit rates. The eight categorical features contained information about the user agent, geographical location, and data if the user is a new or returning one.

2.2. Data types and categories

The data collected from previously mentioned sources appeared in the following formats and meaning:

1) Numerical data: The patterns from extracted images from manually labelled websites were used as feature vectors for training classifier models.
2) Text data, Numerical data: Web Log Files – a unit of user activity on a website usually characterized by a sequence of interactions between a web browser and a web server during a specific time period. These web log session files are primarily text-based and include numerical data for various attributes like session size, average session duration, as well as counts of unique items and Vroots.
3) Numerical data: Website design features that encompass attribute values of all elements, both visible and invisible, that are present on a webpage. This data was acquired by employing a web scraper, which was fed with a list of URLs.
4) Text data: Content extracted from homepages and subpages, content of news article webpages, text in Tweets. Categorical data: Lists of websites categorized as either fake news or traditional news, classification of domains as having little content or being for sale, classification of URLs according to the domains in the news sites lists.
5) Text data: Raw URLs and HTMLs were collected and processed to extract relevant information such as metadata, text content, and structural elements. Raw URLs were parsed to identify domains, paths, and parameters, while HTML documents were analysed to extract text, tags, attributes, and the hierarchical structure of web pages.
6) The data types in the paper include numerical data (user ratings, standard deviations), categorical data (user IDs), binary data (pairwise comparisons, control questions), image data (webpage screenshots), and temporal data (timestamps).
7) Numerical data – all 21 independent features, mainly html tags and keywords, were described as numbers of occurrence or length or word/character count. The only categorical data was a dependent variable representing a level of adjustment to SEO guidelines.
8) Numerical data – statistic of users' activity in terms of spent time on website and bounce rate. Text data – the user identification information, including the operating system type, geographical location and traffic source as well as the visitor type determined as either new or returning one.

To summarize, the researchers made their predictions based on the following web related elements : Images, Web Log Files, Web elements and their attributes, raw HTMLs and URLs, user data and user activity, timestamps, Packet Capture Files. The data used in by them can be assigned to a particular goal of the prediction what is depicted in the table 2.

Table 2. Association between data categories and their respective served prediction purposes

Goal of the prediction	Data categories
Website categorization	Images, Web elements and their attributes, Text on websites, Raw HTML and URL, User ratings, Webpage screenshots.
Web Navigation Prediction	Web Log Files, Statistic of users' activity, User identification information.
Evaluation of website aesthetics	User ratings, Webpage screenshots.
Detection of encrypted malicious activity	Packet capture files.
Predicting users' intention	Statistic of users' activity, User identification information.

2.3. Preprocessing

The collected data often contains noise, irrelevant content, might be in an unsuitable format for processing. Preprocessing steps the referenced literature included are:

1) Elimination of images utilized for non-page-descriptive purposes, especially for advertisements and navigation items. The feature vectors were derived from these images using a pre-trained Convolutional Neural Network (CNN) model.
2) Removal of HTTP requests for files, differentiating the traffic to user and spider generated, organizing user interactions into sessions, identifying sessions (as time-stamped user click-streams) and users (assigning unique user IDs to web sessions), generating representations of user navigation patterns, and classifying links between these patterns. This preprocessing was performed to prepare the data for building a Markov Model.
3) The values of the collected attributes, such as element areas, text lengths and rates, colours, etc., were calculated. These values were then subjected to statistical functions including maximum, minimum, summation, average, count, colour classification, standard deviation, normal distribution, interquartile range, and normalization.
4) Domains with little content and links directing to external websites were filtered out. URL Characteristics: suffix count, domain length, top-level domain. Homepage Auxiliary Data: Quality, content, and specific links. Homepage Style and Scripts: HTML tags for scripts, style, metadata. Homepage Link Categories: Routine links. Homepage HTML Tags and Path: DOM tree elements, tag counts, characteristics. Network Characteristics: Structure, connectivity metrics. Motifs: Recurring subgraph patterns.
5) Removal of HTTP:// and HTTPS:// prefixes from the URLs. Creation of dictionary of words out of the HTML documents, where all punctuation characters were treated as separate tokens. Then word-level corpus of HTMLs and character-level corpus of URL were tokenised to a one-dimensional digital vector. Equalization of the length of the URLs – filling the lacking characters with a token of no significance. Concatenation of URL and HTML embedded matrices into a two-dimensional matrix.

6) As the crowdsourced data was generated as pairwise comparison, the date preprocessing phase included creating a count matrix that captures the number of times one webpage is preferred over another. Then by applying the Bradley-Terry model the authors estimated the aesthetics ranking out of the calculated probabilities of preferred websites and normalized the score to a range of 1 to 10 to match the first dataset.
7) The researchers did not mention any significant preprocessing steps that were needed to be taken on the dataset. It may be assumed that the nature of data (small numerical data representing the number of occurrence or length of html tags) did not require any manipulation. One action, directed to decrease computational complexity and provide uniformity to a spectrum of algorithms used in the paper, was to use Min-Max Scaler that preserve the original distance and relationship between the datapoints.
8) The preprocessing steps contained encoding the categorical variables with one-hot encoding and standardizing (centralizing) the numerical data the way that mean of the values is equal zero and standard deviation to one. The researchers applied filter-based feature selection methods and compared them to feature extraction. The latter which involves transforming features into linear combinations of attributes, was considered impractical due to the need to track and update features during user interactions. They employed techniques such as Correlation, Mutual Information (MI), and the mRMR algorithm for feature selection. MI, which measures mutual dependence between variables and captures both linear and nonlinear relationships, was used to rank features. Continuous variables were discretized to apply MI effectively. The mRMR algorithm was utilized to select a subset of features that maximized relevance to the class variable while minimizing redundancy among selected features. The study aimed to find the most informative features for classification, ultimately enhancing the system's performance. After first evaluation of the classification model, the authors detected an imbalance problem – the dataset contained much more sessions where user does not make purchase comparing to where the purchase was made. The oversampling approach was applied and resulted in adding more purchase-made session to the data set.

2.4. Dataset size overview

The size of dataset is an important factor in obtaining high scores in models' evaluation metrics. The data sizes and its meaning in the papers is shown in table 3.

Table 3. Data size and meaning of datasets in corresponding papers

No.	Data size and meaning
1	450 Web pages, 6 categories, 3890 images
2	Three datasets were utilized with (number of sessions: number of different pages): (13 754: 683), (38 000: 294), (59 600: 497).
3	15432 URLs**,** 2522 features for each webpage.
4	1800 unreliable news sites, 8200 reliable news sites, 7627 homepages, 2 million subpages, 174 thousand articles.
5	45373 phishing and benign instances that contain URLs and HTMLs. 321 thousand unique words from the set of HTML codes.
6	As the dataset was relatively small – 400 screenshots of web pages rated by 40000 users and 100 screenshots of pages rated by 174 users, the authors used transfer learning methods to increase the efficiency of the score predictions.
7	The dataset consisted of randomly sampled 600 websites from a public directory of websites. Out of these websites 21 independent features were extracted that stored small numerical values describing mainly html tags.
8	Over 12 thousand unique user navigation sessions that end with either purchase or leaving a website.

3. Models used for predictions

The researchers utilized the following models in their experiments:

1) Transfer learning method was used that utilized already trained Deep Convolutional Network VGG-19 model that extracted feature vectors out of images. Subsequently the vectors were used as independent variables for Logistic regression, Support Vector Machine (SVM), K-nearest neighbour (KNN) and Naïve Bayes machine learning algorithms.
2) The models used for web navigation prediction were two threshold-based All-Kth Modified Markov Models which are probabilistic models. One of them with geometric and second with branching factor thresholds. Geometric employs properties like states, outlinks, sessions, and transition. In the contrast, branching factor bases on a ratio between a number of outlinks and number of states.
3) Five well-known machine learning models are employed for website categorization in the second experiment. These models include logistic regression, k-nearest neighbours (KNN), an artificial neural network (ANN), adaptive boosting, and a CART decision tree.
4) The researchers combined custom structure-based models with NELA and RDEL content based classifying models. This way they updated the baseline models by enhancing them with structural features resulting in NELA+ and RDEL+ models.
5) The WebPhish model designed for a binary classification task, aiming to predict between two distinct classes: legitimate or phishing. Consists of two convolutional layers and two FC layers that apply ReLu activation function.
6) The authors decided to use a deep neural network with an architecture inspired by AlexNet that is known to perform well in aesthetics assessment tasks involving web and photo images. The original AlexNet architecture was designed for classification tasks, but in this study, it was adapted for regression. Specifically, the authors replaced the original output layer with a single neuron output for regression. The goal was to predict an aesthetic score for each webpage screenshot. Since the datasets used in the study were relatively small, the authors took steps to prevent overfitting – they decreased the neuron count in the fully connected layers of the network, which improved the model's ability to generalize effectively when working with a limited amount of data.
7) The researchers used five machine learning methods: Logistic Regression, KNN, SVM, Naïve Bayes and J48 Decision Trees. The models had their hyperparameters tuned with Grid Search approach. The researchers put emphasis on statistical tests and techniques of the dataset, i.e. Fleis Kappa to assess reliability of agreement between the SEO experts and correlation analysis between the independent variables. This emphasis aimed to reveal the importance of particular features in the SEO guidelines adjustment level classification.
8) The comparison of three models has been conducted that included: Multilayer Perceptron (MLP) employed with a single hidden layer with resilient backpropagation; SVM algorithm, found its place in the study, despite lacking a mechanism for continuous updates with objective to identify an optimal hyperplane for effective class separation, and the authors utilized both linear and radial basis function kernels to handle linear and nonlinear data relationships. The authors also encompassed the C4.5 algorithm for decision tree construction. This choice was made due to C4.5's ability to handle numerical attributes, address missing values, and perform tree pruning. Additionally, the study explored the Random Forest algorithm, which involves constructing an ensemble of decision trees using bagging resampling and aggregating their predictions via a voting mechanism.

4. Evaluation metrics

The choice of evaluation metric is critical as it directly affects performance assessment. It must align with specific goals, consider trade-offs, adapt to context, reflect real-world impact, address bias and imbalance, and suit the task type (regression or classification) and multiclass complexities. Some domains may require specialized metrics tailored to unique challenges. The papers authors in their models used the following metrics:

1) F1-Score, Accuracy, Precision and recall, Confusion matrix. Accuracy assesses the overall correctness by determining the proportion of accurate predictions among all the predictions made. Precision measures the accuracy of positive predictions, finding the ratio of true positives to all positive predictions. Recall assesses the model's ability to identify all actual positives, calculated as the ratio of true positives to all actual positives. F1-Score is a balance between precision and recall, providing a single score that considers both false positives and false negatives.
2) Coverage – average number of predictions possible for the given test dataset – the ratio of total possible predictions from each session to total number of session in the testset; Prediction Accuracy – a measure calculated by dividing the number of correct predictions by the total number of sessions in the test set.
3) The proposed approach's performance was evaluated using the following metrics: Classification Accuracy: These measure how accurately the classifier predicted website categories. It's computed by dividing the count of correctly classified instances by the total number of instances. True Positive Rate (Recall), True Negative Rate, Precision and F1-Score.
4) The study primarily uses the ROC AUC metric to evaluate the performance of the classifiers. This metric evaluates the classifiers' capacity to differentiate between positive and negative classes, while considering the class imbalance between mainstream and fake news. Additionally, the study conducts an error analysis based on websites attributes such as age, popularity, and ideological leaning to further assess model performance.
5) Confusion matrix was generated to describe the performance of the classification model that allowed to visually evaluate the model, as well as calculate F1-Score, Accuracy, Precision and Recall.
6) The evaluation metrics mentioned include Precision, Recall, F1-Score, and Accuracy. All the metrics were provided for each of the trained classifiers with an additional differentiation for the applied method of image feature extraction.
7) The models were evaluated on Accuracy of the predictions with used of both Holdout and 10-Folds Cross Validation methods. The researchers also evaluated the suitability of the dataset size by plotting the learning curve of each model. To prove the difference between pre-set baseline model, the McNemar and Wilcoxon tests were performed to determine the level of similarity of the predictions.
8) The researchers provided a comprehensive classification report that contained F1-Score, Accuracy, True Positive Rate, True Negative Rate determined for each class in the classification.

5. Conclusions

This paper has examined a spectrum of prediction tasks within the subject of website-related predictions. These tasks encompass website categorization, web navigation prediction, crack and malicious website detection, fake news website detection, phishing website detection, and the assessment of website aesthetics.

Researchers carried out acquisition of data from various sources, which range from manually labelled sets to publicly available datasets or combination of both. Furthermore, the datasets themselves exhibit significant diversity in terms of size and format, with some encompassing hundreds of web pages and images, while others include massive collections of URLs, web content, or session logs. In cases where dataset sizes are constrained, transfer learning and model adaptation have proven beneficial, enhancing model performance. In essence, data handling in the context of web-related predictions is not a one-size-fits-all endeavour but rather a dynamic process that necessitates adaptability to accommodate the unique characteristics and requirements of each task.

Modelling approaches also exhibit a wide range, spanning across traditional machine learning algorithms such as SVM, KNN, Naïve Bayes, and logistic regression, along with probabilistic models like Markov Models. Deep neural networks inspired by architectures like AlexNet have been employed, as well as customized structural models tailored to specific prediction tasks.

The evaluation of model performance relies on a suite of metrics, including F1-Score, accuracy, precision, recall, ROC AUC, and more. These metrics encompass a spectrum of measures, including but not limited to the F1-Score, which balances precision and recall, accuracy, which quantifies overall correctness, precision, highlighting the proportion of correctly predicted positive instances among all instances predicted as positive, recall, indicating the proportion of actual positive instances that were correctly predicted, and (ROC AUC), which evaluates the model's ability to distinguish between classes.

The dynamic nature of the web imposes adaptability in data acquisition, preprocessing, and modelling. There are various ways of achieving similar results and only continuous creation and assessment of the models can reveal the best state-of-art approach to predicting diverse, only growing web-related dependencies. It is specifically the tempo of the growth of the web along rising users' activity that drives the scientific research in the fields of website categorization, identification of malicious content and user navigation prediction.

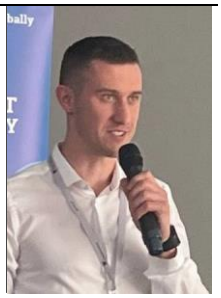
References

[1] Bozarth L., Budak C.: Lay it Out: Detecting Fake News Publishers through Website Structure Data, 2019 [http://doi.org/10.2139/ssrn.3419781].
[2] Cohen D. et al.: Website categorization via design attribute learning. Computers & Security 107, 2021, 102312 [http://doi.org/10.1016/j.cose.2021.102312].
[3] Delitzas A., Chatzidimitriou K. C., Symeonidis A. L.: Calista: A deep learning-based system for understanding and evaluating website aesthetics. International Journal of Human-Computer Studies 175, 2023, 103019.
[4] How many websites are there in the world? – A Daily Calculator [https://www.statsfind.com/how-many-websites-are-there-in-the-world-a-daily-calculator/] (available: 13.02.2024).
[5] Jindal H., Sardana N.: Web navigation prediction based on dynamic threshold heuristics. Journal of King Saud University–Computer and Information Sciences 34(6), 2022, Part A, 2820–2830 [http://doi.org/10.1016/j.jksuci.2020.03.004].
[6] Korkmaz M. et al.: Deep neural network based phishing classification on a high-risk URL dataset. International Conference on Soft Computing and Pattern Recognition. Springer International Publishing, Cham, 2020.
[7] Matošević G., Dobša J., Mladenić D.: Using Machine Learning for Web Page Classification in Search Engine Optimization. Future Internet. 13(9), 2021.
[8] Nandanwar A., Choudhary J.: Web Page Categorization based on Images as Multimedia Visual Feature using Deep Convolution Neural Network, 2020, 619–625.
[9] Opara C., Chen Y., Wei B.: Look before you leap: Detecting phishing web pages by exploiting raw URL and HTML characteristics. Expert Systems with Applications 236, 2024, 21183 [http://doi.org/10.1016/j.eswa.2023.121183].
[10] Sakar C. O. et al.: Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and LSTM recurrent neural networks. Neural Comput & Applic 31, 2019, 6893–6908.
[11] Shaffi S. S., Muthulakshmi I.: Search Engine Optimization by using Machine Learning for Web Page Classification. International Conference on Augmented Intelligence and Sustainable Systems – ICAISS, 2022, 342–349.

M.Sc. Eng. Patryk Mauer
e-mail: patryk.mauer@student.po.edu.pl

In 2022 he started his doctoral studies at the Opole University of Technology in the field of technical informatics and telecommunications.
Research interests: business process automation, web scraping, artificial intelligence.

https://orcid.org/0000-0003-4173-0424



http://doi.org/10.35784/iapgos.5628 *received: 30.11.2023 | revised: 13.02.2024 | accepted: 12.03.2024 | available online: 31.03.2024*

FORMATION OF HIGHLY SPECIALIZED CHATBOTS FOR ADVANCED SEARCH

Andrii Yarovyi, Dmytro Kudriavtsev
Vinnytsia National Technical University, Department for Computer Science, Vinnytsia, Ukraine

Abstract. *In this research, the formation of highly specialized chatbots was presented. The influence of multi-threading subject areas search was noted. The use of related subject areas in chatbot text analysing was defined. The advantages of using multiple related subject areas are noted using the example of an intelligent chatbot.*

Keywords: text-processing, intelligent data analysis, chatbot, advanced search

TWORZENIE WYSOCE WYSPECJALIZOWANYCH CHATBOTÓW DO ZAAWANSOWANEGO WYSZUKIWANIA

Streszczenie. *W tym badaniu przedstawiono tworzenie wysoce wyspecjalizowanych chatbotów. Zwrócono uwagę na wpływ wielowątkowego wyszukiwania obszarów tematycznych. Zdefiniowano wykorzystanie powiązanych obszarów tematycznych w analizie tekstu chatbota. Na przykładzie inteligentnego chatbota odnotowano zalety korzystania z wielu powiązanych obszarów tematycznych.*

Słowa kluczowe: przetwarzanie tekstu, inteligentna analiza danych, chatbot, zaawansowane wyszukiwanie

Introduction

In recent years, chatbots have gained considerable popularity among software tools using artificial intelligence technologies. Among the possibilities of their application, the most popular is the directory function, which consists in finding information upon request and forming the most correct answer. At the same time, the source of information can be both static and dynamic, according to needs and tasks [16]. During previous research in recent years, the main attention was paid to the multi-subject areas of the source, and the possibility of using the Internet as the largest source of information for searching for relevant information for the chatbot user [18]. Considering the spread of popular chatbots ChatGPT from OpenAI and Bard from Google, the research direction confirms the relevance of all previous efforts, and further development is possible due to increasing speed from a technical point of view and improving algorithms for searching and filtering data from a subject point of view [15]. Using the results of recent research on the use of machine learning algorithms and deep recurrent neural networks, it is worth noting the possibility of arbitrary horizontal scaling of the number of subject areas [6, 18]. But despite this, it is the need for expert knowledge that plays a key role in further research in the field of application of intelligent information technologies in chatbots, thanks to the levelling of the amount and volume of data operated by such chatbots as Bard and ChatGPT [5]. Applying machine learning technologies and deep neural networks as artificial intelligence technologies, it is possible to achieve a significant advantage in narrow-profile subject areas of technical, scientific, or commercial orientation. Using search engine databases as data sources, significant progress has been made, based on the current use of chatbots such as ChatGPT, Bard [2, 14]. But with an increase in the horizontal scaling of subject areas, the possibilities for vertical scaling are significantly reduced, which leads to a superficial level of information search and, accordingly, the results of information analysis and filtering [16]. Based on this statement, there is a need to research the use of a limited number of subject areas with the use of similar machine learning methods and artificial intelligence technologies to determine the feasibility of limiting the chatbot data source to achieve a more in-depth analysis of information and provide more correct search information. During further research, it is necessary to investigate the impact of the general application of intelligent search algorithms and data filtering of all knowledge bases available during the research in comparison with the selective application to a limited number of chatbot knowledge bases. As a result of the research, the qualitative impact of limiting the amount of information on the quality of its assimilation will be revealed, and the priority directions of further research in the field of commercialization and expert use of chatbots in solving industry tasks of a search nature will be determined.

1. Using related subject areas

During the research of the influence of the number of subject areas on the quality of information analysis when using the same software and technological capabilities, was found that the number of subject areas used during information analysis significantly depends on the size of the data source, as well as the distribution of data between subject areas [17]. For an ideal case, it is necessary that there is an even distribution of information between all subject areas included in the knowledge base of the chatbot. Thus, adding a new subject area to the chatbot knowledge base will require the addition of a significant amount of expert knowledge to perform an even distribution of data in the knowledge base. It is also worth noting that during past research it was noted that related subject areas form more relationships when using the dictionary data structure [18]. In this regard, an assumption was made that increasing the number of related subject areas for the chatbot knowledge base increases the quality component of the advanced search for relevant information due to a larger amount of data for filtering. To confirm this assumption, it was necessary to find the necessary data and configure the knowledge base of the chatbot by applying machine learning methods and recurrent neural networks.

2. Experiments

Data sets of the Kaggle platform were used as a data source. The total number of subject areas that were used for testing was 20, among which there were related subject areas and unrelated subjects, with uniform distribution of data, as well as uneven distribution during different experiments [11]. For the training sample, 60% of the entire available volume of data, which amounted to more than 10 million terms, was used. This capacity was selected as maximum that was found from Kaggle source that will be acceptable to use in the research. A total of 372 experiments were performed in which the total number of subject areas, the number of related subject areas, and the distribution of data between subject areas were varied. The tests based on average text length, number of subject areas, and using related and non-related subject areas in test groups. In average 28–34 tests per group were using all five different number of subject areas (4, 6, 10, 14, 20) and each test has his own pair with using non-related subject areas. Totally was created 12 test groups. Portion of results of tests are presented on the figures 1–5. The main characteristics that were operated by tests are text analysis

accuracy, velocity, and data distribution coefficient. Based on these result parameters can be defined influence comparison between using related and non-related subject areas. Also, from these experiments will be checked influence of number of subject areas on the accuracy of text analysis accuracy. According to the first two figures, was found that the influence of the uniformity of data distribution between the subject areas of the chatbot knowledge base decreases.

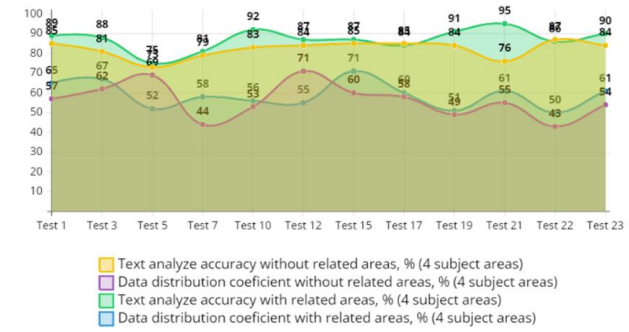


Fig. 1. Accuracy and data distribution comparison for 4 subject areas

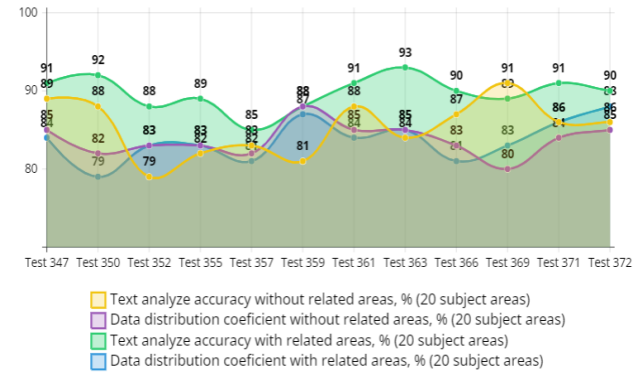


Fig. 2. Accuracy and data distribution comparison for 20 subject areas

Figure 3 corresponds to the higher average accuracy of text analyzing by keywords set up to 1–7 percent if related subject areas were used in experiments. This based mostly on bigger union kernel of terms, related to both subject areas. From technical aspects of this comparison was used improved semantic text analysis with programming module, created in previous research [18].

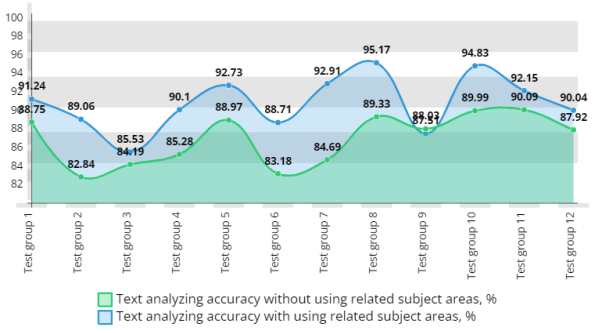


Fig. 3. Text analyzing accuracy comparison between using and non-using related subject areas

As a result of the testing, the adequacy of the test results increases (figure 3), and the speed of processing input information does not change significantly (figures 4, 5) when using most related subject areas. In each test group was selected 28–34 tests. Each test consists of 10–15 text messages from different topics, including subject areas content and context that are not related to any of existing subject area. Test group is balanced by data preparation process which includes precheck of well-known terms from subject areas in text messages from 10 up to 65 percent from whole list of words in messages. In this regard, it is expedient to state that the narrowly focused knowledge base of the chatbot increases the quality of providing information to the user, without reducing the speed of information analysis by the chatbot. As a use of these statements in further research, it is worth examining the field of commercialization, namely the formation of subject groups and criteria that will allow the use of highly specialized chatbots for the analysis of financial transactions, news, natural and man-made phenomena, etc. At the same time, the structure of the chatbot knowledge base, its organization and means of analysis and filtering of chatbot data must be clearly defined to obtain the best result.

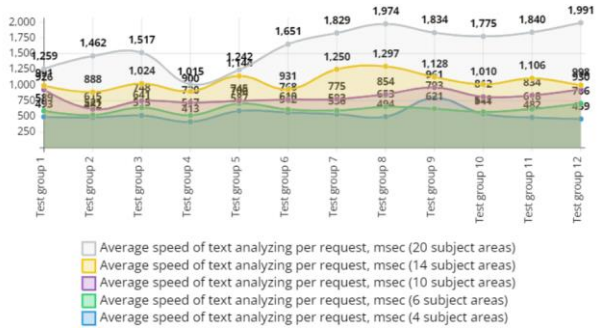


Fig. 4. Velocity of text analyzing per request with using related subject areas

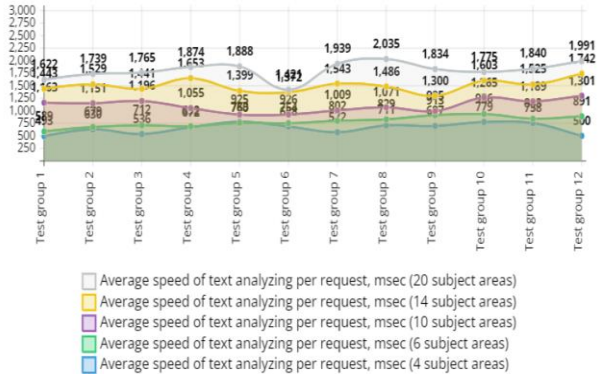


Fig. 5. Velocity of text analyzing per request without using related subject areas

3. Client-server chatbot architecture

For the actual application of a chatbot, it is necessary to develop a software architecture and use tools for its software implementation. Observing modern chatbots such as ChatGPT or Bard, a web interface using a client-server architecture is the best solution, since the key feature of use is the possibility of cross-platform support of the web interface with most modern devices and data protocols, as well as the remoteness of operational capabilities from the user interface with cloud services [10].

In addition, during previous research, the client-server architecture fully met all the requirements, namely the minimization of data transport time, the possibility of setting additional parameters, as well as the time of initial setup of the infrastructure for performing experiments [16]. In this regard, the main attention from a technological point of view is focused on the server architecture of the intelligent chatbot. Considering the latest research in the field of intelligent chatbots, let's consider the server architecture in more detail. It was based on an example from a previous research experiment with the use of additional metrics to determine the level of relatedness of subject areas, as well as an extended analysis of the distribution of data in the general and selective states [17].

Server implementation can be diverse in terms of system load, number of users, as well as the list of available hardware and software solutions. Since the key role of this article is to compare the impact of the quality of user query analysis when using related and unrelated subject areas, as well as when using data from narrow-profile subject areas, the focus is on data

and tools for their analysis. Accordingly, the server architecture is shown in figure 6.

Web-client will be separated project that can be scalable via using RabbitMQ or another message broker technology, Service bus is a hub of requests for chatbot which operate in parallel between user sessions. It responds to the destruction of client message to terms and then groups them into blocks with 3–7 terms in each block. By this operation, messages can be sent to all subject areas databases and results will have better accuracy to pass threshold level [17]. All services will be connected to the Service bus, including operational modules, database repository, logging service as presented in figure 6.

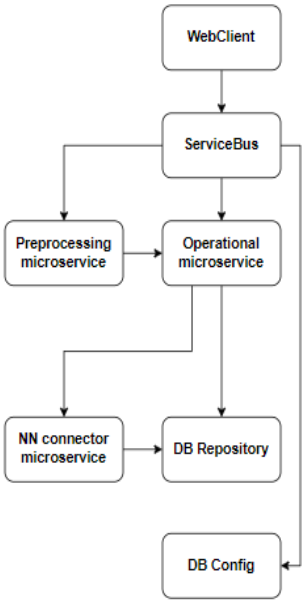


Fig. 6. Chatbot client-server architecture

As a server, ASP.NET Core Web API will be used, which has all the functionality to interact with all system components, including databases and neural network API interface. For client-server architecture will be used RESTful API.

For convenient observation of the results of the experiments, a metrics viewer program was created, which shows chronological graphs of all flows, intermediate calculations, and basic records of usage history, which can be exported to a text file or an Excel data table.

An N-number of input streams corresponding to the number of subject areas is allocated to the main-stream of user requests, and the adjacency metric of the streams is also calculated, and coefficients of affinity of the data from the user request and the streams are formed, forming a matrix of input signals with the dimension, where the additional line is the coefficients of the user's input request. Similar calculations were found in sphere of security for cryptosystems, which also use adjacency metrics for preventing side channels attacks [9]. Subsequently, this matrix is sent to a recurrent neural network, which returns a vector of values for each of the streams, which determines the level of contiguity of the subject area with the user's request [4]. As a result of the search in each of the subject areas, the streams form a list of the most relevant data in the form of a list of terms with a coefficient of adjacency to the user's request. After that, the multiplication of the term matrix of all subject areas with the vector of adjacency coefficients of the subject areas with the user's request is performed. After this operation, it remains only to screen out the terms with final adjacency coefficients to the user's request, by entering the adequacy threshold coefficient, which is described in more detail in the previous article. The final stage is the preparation of the result to a presentable appearance through the use of auxiliary syntactic structures and stylistic analysis of the user's message. In this way, it is advisable to form an improved semantic core that strengthens the context and makes.

In accordance with the chosen client-server architecture, the programming languages C# 7.0 were chosen for the server part, Typescript for the client part, which was implemented on the Angular 16 framework, YAML for combining and configuring services with each other thanks to Docker and Docker Compose, Python 3.10 for working with neural network and neural network learning [13]. Regarding technologies, only a basic list of technologies used in this client-server architecture is given: ASP.NET Core Web API, Angular 16, WPF, Entity Framework Core, SignalR, TensorFlow [7], Docker [12], MongoDB, Azure Pipelines. Visual Studio Code and PyCharm [1, 3, 8] were used as platforms for implementation.

4. Conclusion

After the program implementation, the experimental part of the solution was performed, the results of which are shown in figures 1–5, which revealed the positive impact of the use of related subject areas. Given the small amount of test data compared to the amount of data of search networks, it is difficult to clearly determine the influence of the affinity of subject areas, but more than 81% of all experiments indicate an increase in the accuracy of the response to a user's request in experiments using related subject areas by an average of 10–15% for most terms that passed the threshold than in the experiments where most of the subject areas were unrelated, as shown in comparison between table 1 (without using related subject areas) and table 2 (with using related subject areas).

Table 1. Formalized presentation of experimental results without using related subject areas

N_{SA}	4	6	10	14	20
N_{RSA}	1	2	2	3	7
N_{TE}	10	23	56	71	26
N_{TR}	1	3	5	7	10
C_{DDSA}	89%	24%	56%	72%	85%
N_{ATE}	23.39	24.16	17.328	16.24	18.25
T	0.427	0.529	0.831	0.828	0.991
C_{AA}	0.632	0.524	0.473	0.307	0.420

Table 2. Formalized presentation of experimental results with using related subject areas

N_{SA}	4	6	10	14	20
N_{RSA}	3	5	6	7	7
N_{TE}	10	23	56	71	26
N_{TR}	1	3	5	7	10
C_{DDSA}	63%	30%	53%	74%	85%
N_{ATE}	21.375	24.8	17.136	16.1	17.99
T	0.483	0.571	0.701	0.943	1.034
C_{AA}	0.794	0.678	0.572	0.415	0.515

Where columns are: N_{SA} – number of subject areas that used in experiments; N_{RSA} – number of related subject areas; N_{TE} – number of tests; N_{TR} – number of terms that used in tests; C_{DDSA} – coefficient of data distribution subject areas, which calculates as

$$AVG\left(\frac{ABS(K-N)}{K}\right) \tag{1}$$

where K – is average number of terms that used for one subject area in test, N – is number of terms that used for specific subject area; N_{ATE} – average terms in query; T – average processing time per user's request, sec; C_{AA} – avg. coefficient value of adjacency of terms which calculated as division of number of terms for primary subject area to all terms that found in user request.

Considering the results of the experiments, it is advisable to draw a conclusion about the relevance of the application of related subject areas for highly specialized industries that require the advanced search for specialized information in connection with the improvement of the quality of information search by a chatbot. As for the field of application, everything depends on the needs of the user, and the size of the knowledge

base can be arbitrary, because the main requirement is relatedness of the subject areas.

The most important and troublesome task in this case is the selection of expert data for the formation of the knowledge base of the chatbot.

For further research, it would be advisable to increase the scope of the subject areas, as well as the amount of data for analysis and training. The main sources of data used during a series of experiments were Kaggle data sets, which are quite popular among research in the field of machine learning and artificial intelligence technologies, but in the future, it is necessary to form data samples using search networks and their power, due to the lack of data limit in similar systems. One of the growth direction perspectives of the research could be cooperation with the ChatGPT development team from the OpenAI company.

During the research, an increase in the level of analysis of the user's request was noted when applying related subject areas in the chatbot knowledge base. At the same time, the dependence of the total volume of data in the knowledge base on the number of related subject areas has not been established, as well as a significant dependence on the speed of processing the user's request when the total volume of data increases, which confirms the assumption about the priority of data relatedness over uniform distribution. In the course of further research, more extensive testing will be conducted based on the use of formatted data from search networks.

References

[1] Agarwal S., Rahul P., Neetu: New Text Detection Technique Using Machine Learning Architecture. Dwivedi S., Singh S., Tiwari M., Shrivastava A. (eds): Flexible Electronics for Electric Vehicles. Lecture Notes in Electrical Engineering 863. Springer, Singapore, 2023 [https://doi.org/10.1007/978-981-19-0588-9_1].

[2] Arkoudas K.: ChatGPT is no Stochastic Parrot. But it also Claims that 1 is Greater than 1. Philos. Technol. 36(54), 2023 [https://doi.org/10.1007/s13347-023-00619-6A].

[3] Cao Y., Xu G., Gao Y., Song C.: Application of natural language processing technology based on TensorFlow framework in text mining and discovery algorithm. IET Communications 17, 2022 [https://doi.org/10.1049/cmu2.12534].

[4] Chen W. et al.: Improved Recurrent Neural Networks for Text Classification and Dynamic Sylvester Equation Solving. Neural Process Lett 55, 2023, 8755–8784 [https://doi.org/10.1007/s11063-023-11176-6Raj].

[5] Greco C. M., Tagarelli A.: Bringing order into the realm of Transformer-based language models for artificial intelligence and law. Artif Intell Law 2023) [https://doi.org/10.1007/s10506-023-09374-7].

[6] Henrickson L., Meroño-Peñuela A.: Prompting meaning: a hermeneutic approach to optimising prompt engineering with ChatGPT. AI & Soc 2023 [https://doi.org/10.1007/s00146-023-01752-8].

[7] Joseph J. F. J., Nonsiri S., Monsakul A.: Keras and TensorFlow: A Hands-On Experience. Prakash K. B., Kannan R., Alexander S., Kanagachidambaresan G. R. (eds): Advanced Deep Learning for Engineers and Scientists. EAI/Springer Innovations in Communication and Computing. Springer, Cham. 2021 [https://doi.org/10.1007/978-3-030-66519-7_4].

[8] Karchi R. P., Hatture S. M., Tushar T. S., Prathibha B. N.: AI-Enabled Sustainable Development: An Intelligent Interactive Quotes Chatbot System Utilizing IoT and ML. Whig P., Silva N., Elngar A. A., Aneja N., Sharma P. (eds): Sustainable Development through Machine Learning, AI and IoT. ICSD 2023. Communications in Computer and Information Science 1939. Springer, Cham. [https://doi.org/10.1007/978-3-031-47055-4_17].

[9] Kvyetnyy R., Ivanchuk Y., Yarovyi A., Horobets Y.: Algorithm for Increasing the Stability Level of Cryptosystems. Selected Papers of the VIII International Scientific Conference “Information Technology and Implementation" – IT&I-2021, 293–301.

[10] Meyer J. G. et al.: ChatGPT and large language models in academia: opportunities and challenges. BioData Mining 16(20), 2023 [https://doi.org/10.1186/s13040-023-00339-9].

[11] Mondal B.: Best 25 Datasets for NLP Projects. Kaggle [https://www.kaggle.com/discussions/general/150720] (avaible 13.05.2020).

[12] Pallis George, Trihinas D., Tryfonos A., Dikaiakos M.: DevOps as a Service: Pushing the Boundaries of Microservice Adoption. IEEE Internet Computing 22, 2018, 65–71 [https://doi.org/10.1109/MIC.2018.032501519].

[13] Raj A., Jasmine K.: Building Microservices with Docker Compose. The International Journal of Analytical and Experimental Modal Analysis XIII, 2021, 1215.

[14] Siad S. M.: The Promise and Perils of Google's Bard for Scientific Research. AI. 2023 [https://doi.org/10.17613/yb4n-mc79].

[15] Thapa S., Adhikari S.: ChatGPT, Bard, and Large Language Models for Biomedical Research: Opportunities and Pitfalls. Ann Biomed Eng 51, 2023, 2647–2651 [https://doi.org/10.1007/s10439-023-03284-0].

[16] Yarovyi A. et al.: Information technology in creating intelligent chatbots. Proc. SPIE 11176, 2019, 1117627 [https://doi.org/10.1117/12.2537415].

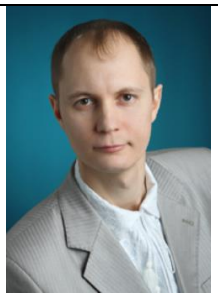
[17] Yarovyi A., Kudriavtsev D.: Dictionary data structure for a text analysis task using cross-references. IEEE 17th International Conference on Computer Sciences and Information Technologies – CSIT, 2022, 61–64 [https://doi.org/10.1109/CSIT56902.2022.10000460].

[18] Yarovyi A., Kudriavtsev D.: Method of multi-purpose text analysis based on a combination of knowledge bases for intelligent chatbot. CEUR Workshop Proceedings 2870, 2021, 1238–1248.

Prof. Andrii Yarovyi
e-mail: a.yarovyy@vntu.edu.ua

Head of Department for Computer Science of Vinnytsia National Technical University (Ukraine). Author of more than 100 technical articles (29 Scopus indexed articles), 5 monographs, 2 patents.
Scientific research related to computer science, intelligent information technologies, image processing, parallel computing.

https://orcid.org/0000-0002-6668-2425



M.Sc. Dmytro Kudriavtsev
e-mail: dmytro_k@vntu.edu.ua

Assistant at Computer Science Department from 2021. Author of more than 20 scientific publications, 7 Scopus indexed articles.
Scientific research related to artificial intelligence, chatbots, object-oriented programming, software development.

https://orcid.org/0000-0001-7116-7869

http://doi.org/10.35784/iapgos.5720 received: 25.12.2023 | revised: 18.01.2024 | accepted: 11.02.2024 | available online: 31.03.2024

METHOD FOR CALCULATING THE INFORMATION SECURITY INDICATOR IN SOCIAL MEDIA WITH CONSIDERATION OF THE PATH DURATION BETWEEN CLIENTS

Volodymyr Akhramovych[1], Yuriy Pepa[2], Anton Zahynei[1], Vadym Akhramovych[3], Taras Dzyuba[4], Ihor Danylov[1]

[1]State University of Information and Communication Technologies, Department of Information and Cybersecurity Systems, Kyiv, Ukraine, [2]State University of Information and Communication Technologies, Department of Robotics and Technical Systems, Kyiv, Ukraine, [3]National Academy of Statistics, Accounting and Auditing, Computing Center, Kyiv, Ukraine, [4]State University of Information and Communication Technologies, Department of Information and Cybersecurity Management, Kyiv, Ukraine

Abstract. *Heterogeneous differential dependencies of the information security indicator (HDISI) in social media (SM) were analyzed, taking into account the duration of the path between clients (UDPC). The resilience of the information security indicator system (RSIIS) was also determined. The HDISI in SM was developed based on UDPC conditions. It uses modern methods and techniques, including a non-specific method. The conditions of a fixed precondition were formed according to the time grid. This dependency provides a comprehensive explanation of how the previous transformation is replaced by the elapsed period. SM is a set of clients and their types of communication. Clients can be individuals, populations, settlements, or countries. Communication is understood as more than just the transmission and receipt of information. It also includes interaction, the exchange of knowledge and expertise, and discussion. Under the angle of mathematics, the HDISI model based on non-homogeneous differential equations (NDE) was analyzed and its transcendental study was done. The transcendental study of nonlinear HDISI models in SM showed that the characteristics of UDPC significantly affect the information security indicator (ISI) - up to one hundred percent. Phase diagrams (PDs) of ISI were studied, which indicate the highest ISI even at the maximum parameters of malicious actions. For the first time, the analysis of designed HDISI structures was carried out and numerical criteria between the capabilities of UDPC and the measures of ISI, as well as the highest ISI, were obtained, which shows the scientific content of this article.*

Keywords: social media, mathematical models, information security indicator

METODA OBLICZANIA WSKAŹNIKA BEZPIECZEŃSTWA INFORMACJI W MEDIACH SPOŁECZNOŚCIOWYCH Z UWZGLĘDNIENIEM DŁUGOŚCI ŚCIEŻKI MIĘDZY KLIENTAMI

Streszczenie. *Przeanalizowano niejednorodne zależności różnicowe wskaźnika bezpieczeństwa informacji (HDISI) w mediach społecznościowych (SM), biorąc pod uwagę długość ścieżki między klientami (UDPC). Określono również odporność systemu wskaźników bezpieczeństwa informacji (RSIIS). HDISI w SM został opracowany w oparciu o warunki UDPC. Wykorzystuje on nowoczesne metody i techniki, w tym metodę niespecyficzną. Warunki stałego warunku wstępnego zostały utworzone zgodnie z siatką czasową. Ta zależność zapewnia kompleksowe wyjaśnienie, w jaki sposób poprzednia transformacja jest zastępowana przez upływający okres. SM to zbiór klientów i ich rodzajów komunikacji. Klientami mogą być jednostki, populacje, osady lub kraje. Komunikacja jest rozumiana jako coś więcej niż tylko przekazywanie i odbieranie informacji. Obejmuje również interakcję, wymianę wiedzy i doświadczenia oraz dyskusję. Pod kątem matematyki przeanalizowano model HDISI oparty na niejednorodnych równaniach różniczkowych (NDE) i przeprowadzono jego transcendentalne badanie. Transcendentalne badanie nieliniowych modeli HDISI w SM wykazało, że charakterystyka UDPC znacząco wpływa na wskaźnik bezpieczeństwa informacji (ISI) – nawet do stu procent. Zbadano diagramy fazowe (PDs) ISI, które wskazują na najwyższy ISI nawet przy maksymalnych parametrach złośliwych działań. Po raz pierwszy przeprowadzono analizę zaprojektowanych struktur HDISI i uzyskano kryteria liczbowe między możliwościami UDPC a miarami ISI, a także najwyższym ISI, co pokazuje naukową treść tego artykułu.*

Słowa kluczowe: media społecznościowe, modele, wskaźnik bezpieczeństwa informacji

Introduction

Subject of research: To analyze the information security indicator (ISI) based on the HDISI.

In the future, the characteristics from table 1 will be used in the dependencies.

HDISI is deployed as recommended when the system characteristics correspond to reality, then attention is paid to its non-approximate class, and the identified interdependencies give a complete understanding of the replacement of the previous state from that time for this time. The variation of the ISI value is detected by prototypical NDE. The position of the system's values and its characteristics are interdependent, which makes it possible to solve NDE with the existing characteristics.

The closest in meaning, works [1, 2, 5, 8, 9]. In work [2], the analysis of the impact, taking into account the duration of the path between clients and other characteristics of the SM on the ISI was performed. It was concluded that the pattern is nonlinear. In the works [1, 5, 8, 9], the mathematical pattern of the ISI in the SM is considered and analyzed, and the characteristics are also taken into account: client interaction, media clustering parameters, information dissemination in media, client relationships and their impact on the ISI pattern.

In the works [3, 4, 6, 7, 10–18], the digital relationships between the characteristics of clients and the media itself and the characteristics of ISI have not been analyzed. This is a key flaw in the cited works.

Table 1. Characteristics that are in use

Characteristic	Description of a characteristic
Z	ISI metric
R_z	characteristic that shows numerical characteristics of the RSIIS
U_v	characteristic that shows the speed of numerical information conversions that are transmitted
U_k	characteristic that shows the speed of numerical information conversions between the information size and its extraction
n_i	unmetrical value of the vertices that are present in the media over time t
P	information volume in the SM
U_{d1}	characteristic that shows the activity of the ISI before information extraction
U_{d2}	characteristic that shows the activity of taking into account the path length between clients on the ISI
S , Q , P_0 , C_0	linear gradients
E	density of harmful entities
W	unsusceptible system rank
δ	prudence indicator of the spread of harmful entities
Characteristic	Description of a characteristic
φ	prudence indicator of harmful entity removal in interaction with the media's immune system
λ_0	interspecies interference indicator of harmful entities
ϑ	growth rate of the immune system
ρ	natural decomposition rate indicator of the immune system
ς	compelling elasticity of immune system increase upon interaction with harmful entities
ξ	indicator of elasticity of immune system removal upon interaction with harmful entities
A	amplitude of oscillations

The justified survey project:

- survey of the quantitative correspondence between the characteristics of the path duration between UDPC and the characteristics of ISI;
- survey of the resilience of the ISI system in the SM from potential impacts in the analysis of phase diagrams.

The addition of adequate "at your fingertips" information, activities and methods for analyzing the quantitative effect of the characteristics of UDPC on the characteristics of ISI to information security workers in the SM is seen as an applied benefit.

The study of the amplitudes of the oscillations of ISI and phase diagrams shows the existing threats and their strength in real time. This will provide the ability for information security workers to make decisions based on the characteristics of ISI in real time.

1. Literature survey and problem statement

In [1], the effects of the path duration between clients and other components of social media characteristics on the information security indicator were analyzed. The model status is nonlinear. The disadvantage is that a nonlinear model for calculating the information security indicator (NMISI) in social media taking into account the path duration between clients has not been developed and investigated.

A general shortcoming of the works [1, 3–18] is that the dependence of NMISI in social media, taking into account UDPC, has not been developed and investigated.

In [5], the effects of ISI were analyzed taking into account the characteristics of client interaction

In [8], the effects of ISI were analyzed taking into account the characteristics of media clustering and the amount of information in social media.

In [9], the effects of ISI were analyzed taking into account the characteristics of propagation and the amount of information in social media

In [1], the effects of ISI were analyzed taking into account the characteristics of client relationship parameters and the amount of information in social media.

In [15], it is pointed out that the storage and exchange of large amounts of information poses privacy problems for the clients of these websites. To prevent these problems, it is necessary to provide a strict privacy policy, data protection mechanisms, reliable and built-in programs that help protect the privacy of clients by limiting the people who have access to the client's personal information. A program is proposed that offers social media users a trust architecture based on reputation. It creates and tracks social reputation, finds circles of communication, and helps users easily, meaningfully, and automatically group their friends to protect their privacy. This system provides grouping of clients through an automated system into different social circles by analyzing the client's social connections and depending on what common information or application they share that other users should not have access to.

In [3], a conceptual approach to the analysis of online social media is developed. The issues of social media management are considered.

In [6, 14], it is pointed out that analysts note that the main causes of incidents in Internet resources are related to human factor, mass hacking of IoT devices and cloud services. This problem is especially exacerbated by the strengthening of the digital humanistic nature of education, the growing role of social media in human life in general. Therefore, the issue of protecting personal information is constantly growing. To solve this problem, we propose a methodology for evaluating the dependence of personal data protection on the volume of information in the system and trust in social media.

In [11], programs are modelled as finite automates and the required profile attributes of the client are used as the conditions that govern the execution of the program. The problems of generalizing the minimum attributes are formulated, and a solution is proposed that reflects the shortest path problem in order to find the minimum set of generalization attributes required to access the program services. The feasibility of this approach is evaluated under the conditions of verifying the implementation of the concept and conducting customer research.

In [4], it is pointed out that the protection of location privacy of clients still remains an open task for social media service providers. It has been shown that hiding the real person and choosing a pseudonym does not guarantee the protection of the client's privacy, since privacy can be violated only by analyzing location data. This is indeed a big problem, since other personal information of clients can be disclosed by analyzing their location data (for example, home address, health status, interests, etc.). In this regard, a distance-based privacy protection mechanism (DBLP2) is developed, a customized location privacy protection approach that is uniquely designed for social media clients. It uses the concept of social distance to generalize client location data before publishing it on social media. The level of generalization is determined taking into account the social distance between clients. Secondly, cryptographic methods for protecting location privacy in geolocation services (LBS) and social networks are considered.

In [7], computer viruses are studied in the form of theory and experiments, as well as security. An epidemiological model of virus spread and cleaning.

In [10], it is reported that if A tells B, and B tells C (and A does not tell C), then clients A and C are at a distance of two. How many clients are at different distances from each client can be important for understanding the differences between clients in the constraints and opportunities they have as a result of their position. That is, can actor A reach client B in several ways? Sometimes multiple links can indicate a stronger connection between two clients than one link.

In [17], the distance between clients in media can be an important macro-characteristic of the media as a whole. Where distances are large, it may take a long time for information to spread among the population. It may also be that some clients are completely unaware of it and are under the influence of others – even if they are technically available, the costs may be too high to conduct the exchange. The variability between clients in the distances they have from other clients can be the basis for differentiation and even stratification. Those clients who are closer to others may have more power than those who are more distant.

In [13], it is noted that one of the most acute problems of social media is the disclosure and unauthorized access to information, data and communication between clients, which is a kind of violation of their privacy, which can be done by social media providers, especially unauthorized clients. One way to protect privacy is to use encryption.

In [16, 18], it is proposed, in particular, to use the distance between two clients to predict whether they are friends. It is first shown that distance is a useful indicator for dividing into friends and strangers. Taking into account the popularity of the location together with the distance, the difference between friends and strangers becomes even greater. It is further shown that distance can be used for effective link prediction. A machine learning classifier is used to further improve the prediction efficiency.

In the work [12], presents nonlinear models, their characteristics, and methods of determination. The results of calculations are presented.

2. Concept and objectives of analysis

The purpose of the analysis is to improve the methodology of ISI in SM, taking into account the characteristics of the UDPC.

The most closely related works are [1, 2, 5, 8, 9], which analyze the impact of parameters such as the size of social media development, user interactions and relationships, network clustering, the spread and volume of information, and their actions on ISI. In [2] analyzed the impact of the duration of the path between clients and other components of the characteristics of social media on the information security indicator. It was shown that the model is nonlinear. The error is that a nonlinear ISI model has not been developed or analyzed. This article investigates the indicated problem.

The scientific article explores the relationship between the characteristics of UDPC and the characteristics of ISI in the analysis and rationalization of ISI in SM.

The main research perspectives are:

- investigating the numerical correspondence between the characteristics of UDPC and the characteristics of ISI;
- investigating the RSIIS in SM from potential impacts based on PD.

Motivations and methods of testing

The paper is devoted to the definition and analysis of the HDISI in SM, with the processing of its characteristics and characteristics of the UDPC. ISI is calculated using indefinite cognitive imitation. The development and investigation were carried out on the basis of scientific thought, the plan of non-precise dependencies. This provided an opportunity to understand and nurture almost unfamiliar technological actions.

To master the HDISI in SM, NDE systems that depict ISI were developed.

The following were adopted: methods for obtaining the solution of NDE (cut-out technique, group search for the corresponding native dependence, etc.); finding characteristics in the MATLAB program. RSIIS from influences on ISI was carried out through the study of NDE and the electronic circuit developed in MATLAB/MULTISIM.

3. Proposed methodology

This methodology is a continuation of the developments outlined in [1, 5, 8, 9], which allowed us to develop these models in the direction of creating a methodology for determining quantitative indicators of information protection in social media from the influence of the distance between users and other media parameters.

3.1. Features ISI in SM considering characteristics UDPC

The graphical dependence of the differential [2] of the UDPC (Fig. 1), the differential of the ISI (Fig. 2) [2] are presented.

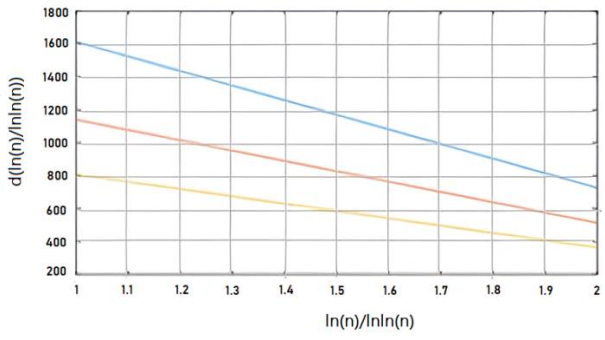


Fig. 1. Differential of the path duration function between clients

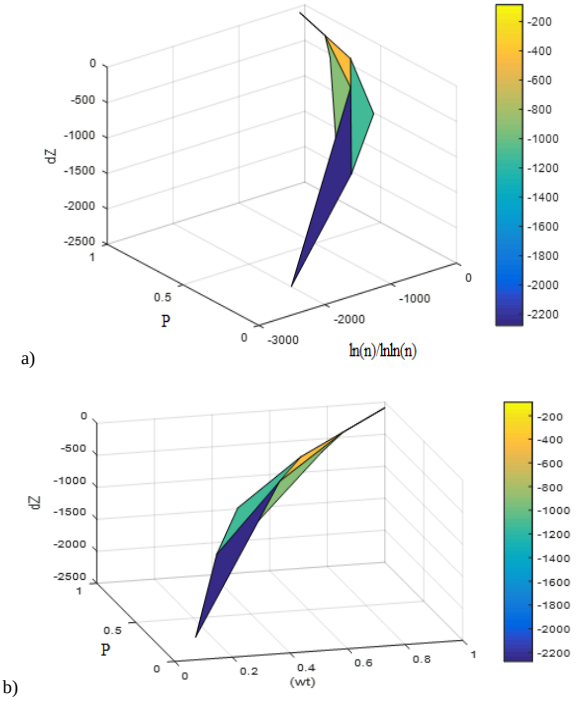


Fig. 2. Differential ISI: a) $dZ = f(P, \ln(n)/\ln\ln(n))$; b) $dZ = f(P, wt)$

To start, we will use a system of equations [1, 5, 8, 9].

The illustrated relationships show that the derivative of ISI is negative. Therefore, according to the Lyapunov theorem, the protection system is reliable and stable.

The ISI nonlinearity is negligible, which suggests that we can solve the dependencies using a successive approximation method.

$$P = P_1 + P_2 + P_3 + ...$$

$$Z = Z_1 + Z_2 + Z_3 + ...$$

We believe that when

$$dP = 0\ ,\ \frac{dP}{dt} = 0 \text{ and } dZ = 0\ ,\ \frac{dZ}{dt} = 0$$

$$P = P_0 \sin \omega t\ ,\ Z = R_0 \sin \omega t\ .$$

The equation can be expressed as follows:

$$\begin{cases} \dfrac{dP}{dt} = R_z Z + P\left(U_v + U_k\right) - S_2\left(P_0^2 \sin^2 \omega t\right) - \\ -S_3\left(P_0^3 \sin^3 \omega t\right) - ... \\ \dfrac{dZ}{dt} = \dfrac{\ln\ln n - n}{n(\ln\ln n)^2} - P\left(U_{d2} + U_{d1}\right) - Q_0\left(C_0^2 \sin^2 \omega t\right) - \\ -Q_0\left(C_0^3 \sin^3 \omega t\right) - ... \end{cases} \quad (1)$$

Let's generalize the equation:

$$\begin{cases} \dfrac{dP}{dt} = \alpha Z + \beta_1 P - \sum_{k=2}^{\infty} S_k P_0^k \sin^k \omega t \\ \dfrac{dZ}{dt} = \beta_2 P + \gamma - \sum_{k=2}^{\infty} Q_k C_0^k \sin^k \omega t \end{cases} \quad (2)$$

where

$$\alpha = R_z\ ,\ \beta_1 = U_v + U_k\ ,\ \beta_2 = -\left(U_{d2} + U_{d1}\right)\ ,\ \gamma = \frac{\ln\ln n - n}{n(\ln\ln n)^2}$$

We will use the process of elimination:

$$\frac{dZ}{dt} = \beta_2 P + \gamma - \sum_{k=2}^{\infty} Q_k C_0^k \sin^k \omega t \Rightarrow$$

$$\Rightarrow I = \frac{1}{\beta_2}\left(\frac{dZ}{dt} - \gamma + \sum_{k=2}^{\infty} Q_k C_0^k \sin^k \omega t\right) \Rightarrow$$

$$\Rightarrow \frac{dP}{dt} = \frac{1}{\beta_2}\left(\frac{d^2 Z}{dt^2} + \frac{1}{\omega}\sum_{k=2}^{\infty}\left(k Q_k C_0^k \sin^{k-1} \omega t \cos \omega t\right)\right) \quad (3)$$

We substitute (3) into the first equation (2):

$$\frac{1}{\beta_2}\left(\frac{d^2Z}{dt^2}+\frac{1}{\omega}\sum_{k=2}^{\infty}\left(kQ_kC_0^k\sin^{k-1}\omega t\cos\omega t\right)\right)=$$
$$\alpha Z+\frac{\beta_1}{\beta_2}\left(\frac{dZ}{dt}-\gamma+\sum_{k=2}^{\infty}Q_kC_0^k\sin^k\omega t\right)-$$
$$-\sum_{k=2}^{\infty}Q_kC_0^k\sin^k\omega t, \tag{4}$$

we will get:

$$\frac{d^2Z}{dt^2}-\beta_1\frac{dZ}{dt}-\alpha\beta_2Z=$$
$$=-\frac{1}{\omega}\sum_{k=2}^{\infty}\left(kQ_kC_0^k\sin^{k-1}\omega t\cos\omega t\right)-$$
$$-\beta_2\sum_{k=2}^{\infty}S_kP_0^k\sin^k\omega t \tag{5}$$

The general form of a homogeneous equation is:

$$Z''-\beta_1Z'-\alpha\beta_2Z=0 \tag{6}$$

eigenvalue equation: $\lambda^2-\beta_1\lambda-\alpha\beta_2=0$

This equation has a positive discriminant:

$$D=\beta_1^2+4\alpha\beta_2>0\Rightarrow\lambda_{1,2}=\frac{\beta_1\pm\sqrt{\beta_1^2+4\alpha\beta_2}}{2} \tag{7}$$

We will obtain:

$$Z_{od}(t)=c_1e^{\frac{\beta_1+\sqrt{\beta_1^2+4\alpha\beta_2}}{2}}+c_2e^{\frac{\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}}$$

common solution of homogeneous equation.

The calculation of the general inhomogeneous equation will be performed by the method of separation of arbitrary constants, where $c_1'(t)$, $c_2'(t)$ is established as:

$$\begin{cases}c_1'(t)e^{\frac{\beta_1+\sqrt{\beta_1^2+4\alpha\beta_2}}{2}}+c_2'(t)e^{\frac{\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}}=0\\ c_1'(t)\dfrac{\beta_1+\sqrt{\beta_1^2+4\alpha\beta_2}}{2}e^{\frac{\beta_1+\sqrt{\beta_1^2+4\alpha\beta_2}}{2}}+\\ +c_2'(t)\dfrac{\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}e^{\frac{\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}}=N(t)\end{cases} \tag{8}$$

Thereafter:

$$N(t)=-\frac{1}{\omega}\sum_{k=2}^{\infty}\left(kQ_kP_0^k\sin^{k-1}\omega t\cos\omega t\right)-$$
$$-\beta_1\gamma-\beta_1\sum_{k=2}^{\infty}Q_kP_0^k\sin^k\omega t$$
$$Z(s)=\int_{t_0}^{t}\left(N(s)-e^{\frac{-\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}s}\frac{e^{\frac{-\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}s}}{\sqrt{\beta_1^2+4\alpha\beta_2}}\right)ds \tag{9}$$

From equations (8, 9) we will expand:

$$c_1'(t)e^{\frac{\beta_1+\sqrt{\beta_1^2+4\alpha\beta_2}}{2}}=-c_2'(t)e^{\frac{\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}}\Rightarrow$$
$$\Rightarrow c_2'(t)e^{\frac{\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}t}\left(-\frac{\beta_1+\sqrt{\beta_1^2+4\alpha\beta_2}}{2}+\frac{\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}\right)= \tag{10}$$
$$=N(t)$$

or:

$$c_2'(t)e^{\frac{\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}t}\sqrt{\beta_1^2+4\alpha\beta_2}=-N(t) \tag{11}$$

Now we will expose:

$$c_2(t)=-\frac{1}{\sqrt{\beta_1^2+4\alpha\beta_2}}\int N(t)e^{\frac{-\beta_1+\sqrt{\beta_1^2+4\alpha\beta_2}}{2}t}dt \tag{12}$$

$$c_1(t)=\frac{1}{\sqrt{\beta_1^2+4\alpha\beta_2}}\int N(t)e^{\frac{-\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}t}dt \tag{13}$$

By carrying out an analysis of equations (7–9) together, we will obtain:

$$Z(s)=\int_{t_0}^{t}\left(N(s)-e^{\frac{-\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}s}\frac{e^{\frac{-\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}s}}{\sqrt{\beta_1^2+4\alpha\beta_2}}\right)ds-$$
$$-\int_{t_0}^{t}\left(N(s)-e^{\frac{-\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}s}\frac{e^{\frac{-\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}s}}{\sqrt{s\beta_1^2+4\alpha\beta_2}}\right)ds \tag{14}$$

To confirm the assessment taking into account the UDPC on the ISI, we will conduct modeling. Graphical dependencies of the ISI are presented in Fig. 3.

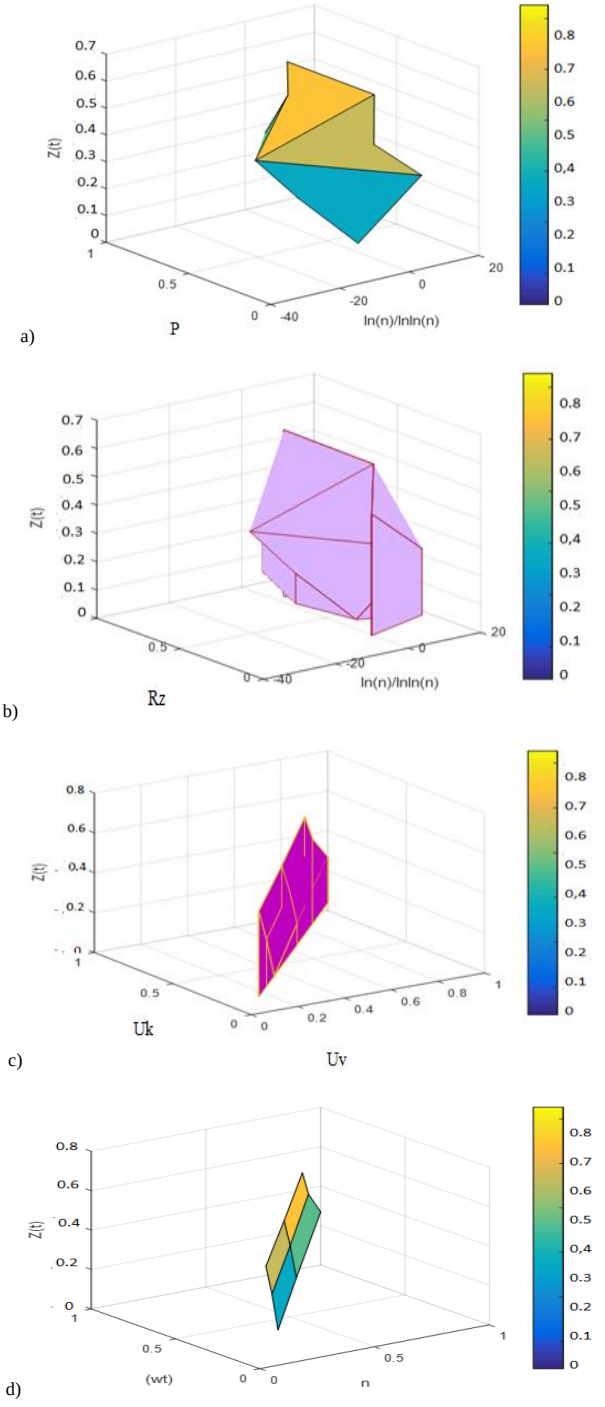


Fig 3. Dependences ISI: a) $Z(t)=f(P,\ln(n)/\ln\ln(n))$, *b)* $Z(t)=f(R_z,\ln(n)/\ln\ln(n))$, *c)* $Z(t)=f(U_k,U_v)$, *d)* $Z(t)=f(wt,n)$

3.2. Formation PD ISI

Output dependency similar to [4, 7, 10, 11, 15]:

$$\frac{d^2Z}{dt^2} - \beta_1 \frac{dZ}{dt} - \alpha\beta_2 Z =$$

$$= -\frac{1}{\omega}\sum_{k=2}^{\infty}\left(kS_k P_0^k \sin^{k-1}\omega t \cos\omega t\right) -$$

$$-\beta_1\left(\frac{\ln\ln n - n}{n(\ln\ln n)^2}\right) + \beta_1\sum_{k=2}^{\infty} S_k P_0^k \sin^k \omega t -$$

$$-\beta_2\sum_{k=2}^{\infty} Q_k C_0^k \sin^k \omega t \qquad (15)$$

The illustrated phase diagrams are elliptical and closed, which indicates that the ISI system is reliable and stable.

Electronic circuit components identification for Fig. 5.

The analysis of the dependence was performed in the Multisim system. An electronic circuit was developed (Fig. 6). The results of the system's operation are presented in Fig. 4.

Thanks to the developed phase diagram, it was possible to model the behavior of the social media information protection system depending on various parameters and media loads. The graphical results are shown in Figs. 4b, c.

The results obtained in Figs. 4b, c are closed oscillating curves resembling a closed ellipse, indicating that the social media information protection system is resistant to various types of influences in the operating range of parameters. Fig. 4a shows the oscillations of the protection system without external influences.

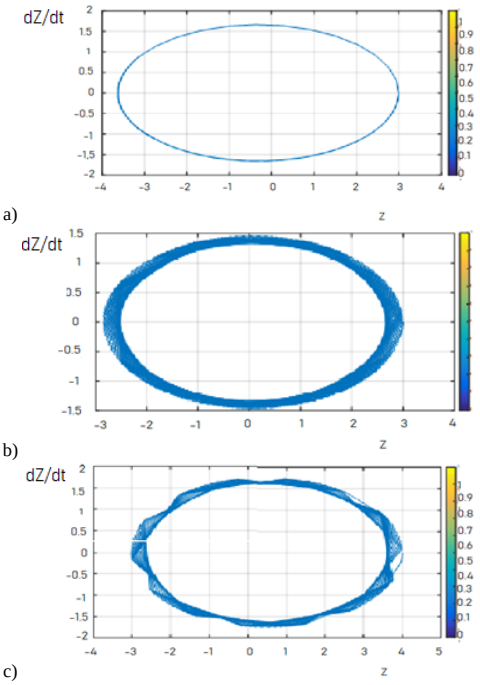


Fig. 4. Phase diagrams of ISI: a) in the absence of influences; b) when the influence is half its maximum value; c) at its maximum value

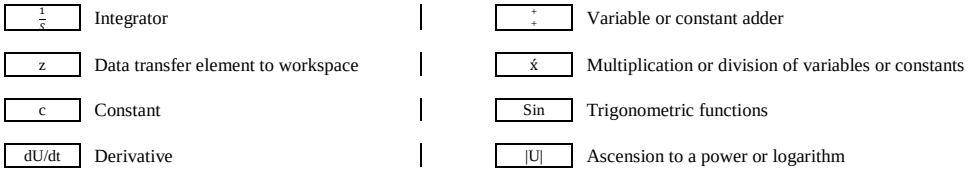


Fig. 5. Electronic circuit components

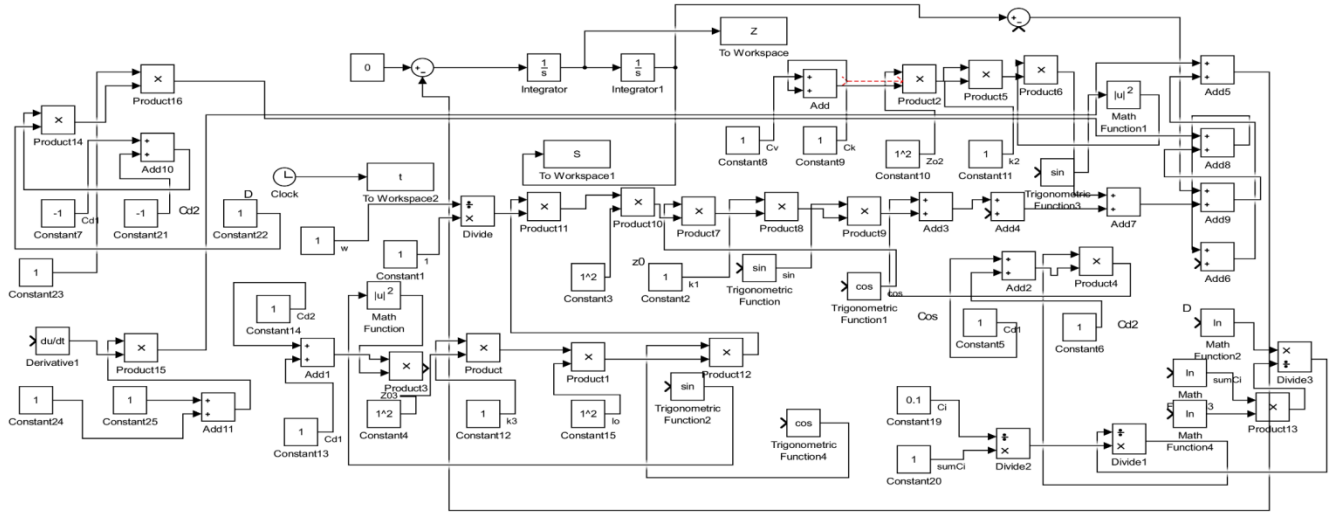


Fig. 6. Phase diagram circuit in Multisim

3.3. Load simulation for ISI

We are developing a system to simulate the impact on the system and the immune resistance of the system. We assume that the dynamics of harmful actions behaves in a logistic pattern. The growth of harmful infection depends on its initial state, weakening caused, including, by immune influence, and the effect of density, while the mutation of the immune response depends on its initial state, natural reduction, activation that leads to the increase of the response, and damage from harmful actions. The comparative characteristic of the injured subject depends on the density of the harmful object and its natural reduction. We will reproduce such a dynamic system with the support of a further system of differential dependencies:

$$\begin{cases} \dfrac{dE}{dt} = \delta E - \varphi WE - \lambda_0 E^2 \\ \dfrac{dW}{dt} = \vartheta - \rho W + \varsigma WP - \xi\varphi WE \end{cases} \qquad (16)$$

This is a predator-prey system. Graphical dependencies of the amplitude of influence according to equation (16) are presented in Figs. 7, 8.

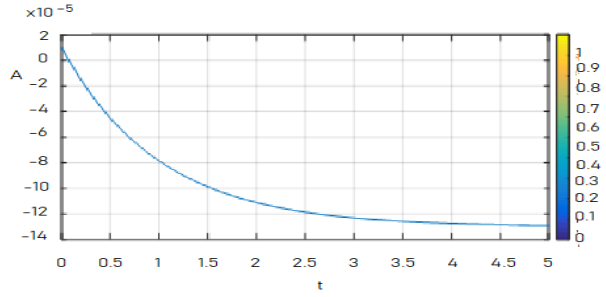


Fig. 7. The smallest amplitude of the force of action of a destructive object. All characteristics (16) are equal to 0.1 relative units

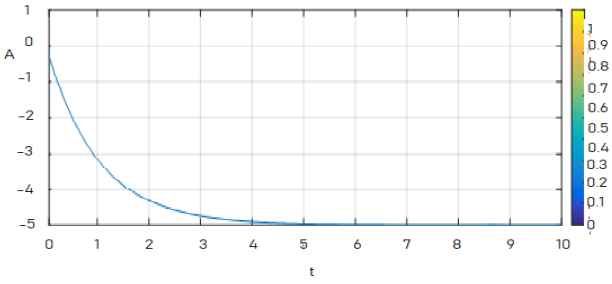


Fig. 8. The greatest amplitude of the force of action of a destructive object. All characteristics (16) are equal to 1.0 relative units

3.4. Exchange of expressed considerations from the study of ISI in SM taking into account UDPC

The analysis of the HDISI using the probability characteristics and the characteristics of the UDPC provided the opportunity to obtain numerical evidence of extremely important isolated characteristics of SM which include the effect of UDPC on the ISI (1, 2, 14) (Figs. 2, 3). A reasoned assessment of the fundamental approaches, introduced scientific thoughts, models that have been evaluated by simulation modeling of the ISI taking into account the effects of influences, approved the methodology.

The analysis of the RSIIS (14) (Fig. 2) and the PD (Fig. 4) shows the reliability and stability of the ISI.

The method allows obtaining quantitative characteristics of the ISI from the characteristics of the SM and UDPC. To date, there is no method that allows obtaining such an achievement.

It is satisfactory that the ISI withstands the maximum loads, including those to UDPC.

Adding to the information security workers in the SM adequate "at your fingertips" information, activities and methods for analyzing the quantitative effect of UDPC characteristics on ISI characteristics, this is seen as an applied benefit.

The study of the amplitudes of the ISI oscillations and the phase diagrams shows the existing threats in real time, their strength. This will provide the opportunity for information security workers to make decisions in real time based on the characteristics of the ISI.

The prospect in further scientific work lies in the identification of other SM parameters and the identification of their impact on the ISI.

4. Numerical expression

The illustration of the ISI differential in Fig. 2 indicates that it has negative values in the entire operating range, which corresponds to Lyapunov's theorem, which means that the RSIIS is ensured. This is confirmed by further studies (Fig. 4).

The values of the ISI characteristics lie within the range from 0 to 1, which indicates a strong influence of the UDPC characteristics (Figs. 3 a, 3 b).

Closed curves without bifurcation points (Fig. 4) at different values of influences indicate ISI stability indicators.

5. Summary

1) The analysis of the model, which is reproduced by the HDISI, shows that quantitative results of the activity of SM characteristics and characteristics of UDPC on ISI in SM and their understanding were obtained. Characteristics of the UDPC action on ISI are in the range from zero to one hundred percent, which made it possible to further conduct research on the RSIIS.
2) The acquisition of RSIIS in SM at different indicators of the action of harmful elements on ISI was performed according to the electronic circuit, in the program MULTISIM. The study of illustrations with oscillations of ISI and phase diagrams confirms RSIIS at different influences of harmful elements.

As a result of the analysis, it can be recognized that the study of the impact of the UDPC on the ISI is accurate. Further scientific work, from our point of view, consists in exploring and using other unique characteristics of SM to establish their impact on ISI.

The study of the amplitudes of ISI oscillations and PD shows the existing threats and their strength in real time. This will provide the opportunity for information security workers to make decisions based on ISI characteristics in real time.

References

[1] Akhramovych V. et al.: Devising a Procedure to Determine the Level of Informational Space Security in Social Networks Considering Interrelations Among Users. Eastern-European Journal of Enterprise Technologies 1 (9(115)), 2022, 63–74.
[2] Akhramovych V.: Method of Calculation of Protection of Information from the Average Length of the Road Between Users in Social Networks. Collection Information Technology and Security 10(2), 2022, 153–164.
[3] Gubanov D., Chkhartishvili A.: A Conceptual Approach to the Analysis of Online Social Networks. Large-Scale Systems Control 45, 2013, 222–236.
[4] Hanneman R., Riddle M.: Introduction to Social Network Methods [https://faculty.ucr.edu/~hanneman/nettext/C7_Connection.html#distance].
[5] Khrashchevskyi R. et al.: Method of Calculating Information Protection from Mutual Influence of Users in Social Networks. International Journal of Computer Network and Information Security – IJCNIS 15(5), 2023, 27–40 [https://doi.org/10.5815/ijcnis.2023.05.03].
[6] Nosouhi M. R.: Location Privacy Protection in Social Networks. Ph.D. thesis. University of Technology Sydney, Sydney 2020 [http://hdl.handle.net/10453/143934].
[7] Rohloff K.: Stochastic Behavior of Random Constant Scanning Worms. The 14th ICCCN, 2005, 339–344.
[8] Savchenko V. et al.: Method of Calculation of Information Protection from Clusterization Ratio in Social Networks. Proceedings of the 3rd International Conference on Information Security and Information Technologies – ISecIT 2021, 24–31.
[9] Savchenko V. et al.: Methodology for Calculating Information Protection from Parameters of its Distribution in Social Networks. 3rd International Conference on Advanced Trends in Information Theory – ATIT, 99–105.
[10] Seyyed M. S. et al.: Privacy Protection Scheme for Mobile Social Network. Journal of King Saud University – Computer and Information Sciences 34(7), 2022, 4062–4074.
[11] Shehab M. et al.: Access Control for Online Social Networks Third Party Applications. Computers & Security 31(8), 2012, 897–911.
[12] Trubetskoy D. I.: Introduction to Synergetics. Chaos and structures. Editorial, Moscow 2004.
[13] Yating L. et al.: Intelligent Privacy Protection of End User in Long Distance Education. Mobile Networks and Applications 27, 2022, 1162–1173.
[14] Yevseiev S. et al.: Modeling the Protection of Personal Data from Trust and the Amount of Information on Social Networks. Eureka Physics and Engineering 13(1), 2021, 24–31.
[15] Yüksel A. S., Halim Zaim A.: A Reputation-based Privacy Management System for Social Networking Sites. Turkish Journal of Electrical Engineering and Computer Sciences 21(3), 2015, 766–784.
[16] Zan Y., Wu J., Li P., Yu Q.: SICR Rumor Spreading Model in Complex Networks: Counterattack and Self-resistance. Physics A: Statistical Mechanics and its Applications 405, 2014, 159–170.
[17] Zhang Y., Pang J.: Distance and Friendship: A Distance-based Model for Link Prediction in Social Networks [https://satoss.uni.lu/members/jun/papers/APWeb15.pdf].
[18] Zhang Y., Zhu J.: Stability Analysis of I2S2R Rumor Spreading Model in Complex Networks. Physics A: Statistical Mechanics and its Applications 503, 2018, 862–881.

Prof. Volodymyr Akhramovych
e-mail: 12z@ukr.net

Doctor of Technical Science, Professor, Associate Professor, Department of Information and Cybersecurity Systems of the State University of Information and Communication Technologies, Kyiv, Ukraine.
Over 200 scientific papers have been published, 13 of it are the USSR copyright certificates, 14 patents of Russia. The sole author of four textbooks, co-authored with two textbooks and three workshops.
He has been working in the field of higher education for more than thirty years.

https://orcid.org/0000-0002-6174-5300

Vadym Akhramovych
e-mail: 12zstzi@gmail.com

Head of the Computing Center of the National Academy of Statistics, Accounting and Auditing, Kyiv, Ukraine.

https://orcid.org/0009-0003-2787-8745

Ph.D. Yuriy Pepa
e-mail: yurka14@ukr.net

Candidate of technical sciences, Head of the Department of Robotics and Technical Systems, State University of Information and Communication Technologies, Kyiv, Ukraine.

https://orcid.org/0000-0003-2073-1364

Ph.D. Taras Dzyuba
e-mail: iwartar@gmail.com

Ph.D. in Technical Science, associate professor of the Department of Information and Cybersecurity Management, State University of Information and Communication Technologies, Kyiv, Ukraine.
He has been working in the field of higher education more than twenty years.

https://orcid.org/0000-0001-6607-2507

Anton Zahynei
e-mail: Antonio.com237@gmail.com

Postgraduate student, senior lecturer. Department of Information and Cybersecurity Systems of the State University of Information and Communication Technologies, Kyiv, Ukraine. Interested in: cloud computing, fog computing, comprehensive information security systems.

https://orcid.org/0000-0002-0303-8501

Ihor Danylov
e-mail: danylovihor@gmail.com

Postgraduate student of Department of Information and Cybersecurity Systems of the State University of Information and Communication Technologies, Kyiv, Ukraine.

https://orcid.org/0009-0000-1426-6414

http://doi.org/10.35784/iapgos.5498 *received: 23.10.2023 | revised: 18.02.2023 | accepted: 14.03.2024 | available online: 31.03.2024*

CORRESPONDENCE MATCHING IN 3D MODELS FOR 3D HAND FITTING

Maksym Tymkovych[1], Oleg Avrunin[1], Karina Selivanova[1], Alona Kolomiiets[2], Taras Bednarchyk[3], Saule Smailova[4]

[1]Kharkiv National University of Radio Electronics, Kharkiv, Ukraine, [2]Vinnytsia National Technical University, Vinnytsia, Ukraine, [3]Vinnytsia Pyrohov National Medical University, Vinnytsia, Ukraine, [4]D.Serikbayev East Kazakhstan State Technical University, Ust-Kamenogorsk, Kazakhstan

Abstract. *Upper limb prosthetic is an area of medical research and development that aims to restore functionality and improve the quality of life of people affected by the loss of one or both upper limbs. The development and implementation of 3D scanning tools and analysis of 3D scanning data requires the use of specialized analysis methods that ensure the achievement of the required indicators. It should take into account the impact of the model resolution on the result. This paper is devoted to the analysis of finding matches between a point cloud of a hand model and another point cloud using Gromov-Wasserstein distance. For analysis, a subset of the MANO dataset was employed, containing a substantial volume of data and serving as a representative sample of the human population. The results obtained indicate the possibility of using this approach in the processing and analysis of three-dimensional data, which serves as one of the stages of designing individualized prostheses.*

Keywords: health care, medical technology, physical rehabilitation, 3D modeling, correspondence matching

DOPASOWANIE ZGODNOŚCI W MODELACH 3D DLA DOPASOWANIA DŁONI 3D

Streszczenie*. Protetyka kończyn górnych to dziedzina badań i rozwoju medycznego mająca na celu przywrócenie funkcjonalności i poprawę jakości życia osób dotkniętych utratą jednej lub obu kończyn górnych. Opracowanie i wdrożenie narzędzi do skanowania 3D oraz analiza danych pochodzących ze skanowania 3D wymaga zastosowania specjalistycznych metod analizy, które zapewnią osiągnięcie wymaganych wskaźników. Należy przy tym uwzględnić wpływ rozdzielczości modelu na uzyskany wynik. Niniejszy artykuł poświęcony jest analizie znajdowania dopasowań między chmurą punktów modelu dłoni a inną chmurą punktów przy użyciu odległości Gromova-Wassersteina. Do analizy wykorzystano podzbiór zbioru danych MANO, który zawiera znaczną ilość danych i służy jako reprezentatywna próbka populacji ludzkiej. Uzyskane wyniki wskazują na możliwość wykorzystania tego podejścia w przetwarzaniu i analizie danych trójwymiarowych, które służą jako jeden z etapów projektowania zindywidualizowanych protez*.

Słowa kluczowe: opieka zdrowotna, technologia medyczna, rehabilitacja fizyczna, modelowanie 3D, dopasowanie

Introduction

The wide-scale aggression against Ukraine has resulted in injuries to a significant number of individuals across different age groups, professions, and social statuses. This has already led to a shortage of appropriate resources in both the domestic and international markets. The development of accessible and cost-effective solutions enabling the modeling, simulation, fabrication, and testing of prosthetic devices for a wide audience is essential. In this context, the mathematical foundations, theoretical advancements, and practical research must be presented to facilitate their implementation. It should be noted that in the case of prosthetics for children, the prosthetic process usually involves multiple stages due to the healthy growth of children, necessitating prosthetic updates [1, 7, 9].

Furthermore, the market demands not only a certain quantity of specialized materials and products but also specialized devices (such as scanners) that enable the scanning and digitization of the hand, thereby facilitating the personalization of the prosthesis [8, 20]. Moreover, in the current landscape, a multitude of scanners exists, encompassing diverse pricing tiers and being founded upon various underlying technologies. Notably, scanners employing techniques such as photogrammetry, Time of Flight, and others emerge, each imbued with its unique capabilities [6, 14, 16]. These advancements facilitate precise measurements and digitization of anatomical structures, thereby fostering the development of highly customized and tailored prosthetic solutions. The availability of a wide spectrum of scanning technologies empowers researchers and practitioners alike to select the most suitable and effective approach for capturing intricate anatomical details. Photogrammetry, for instance, leverages the principles of triangulation from multiple images, enabling accurate three-dimensional reconstructions [9]. On the other hand, Time of Flight scanners utilize the measurement of light travel time to create detailed depth maps, ensuring high precision [5, 27]. Such technological diversity is a testament to the rapid progress in the field, opening doors for innovative solutions and personalized prosthetic designs.

However, measurements based on the polygonal model data require monotonous work from operators, which is not always convenient. Moreover, contemporary medical solution development practices focus on automating both the analysis and the automatic modeling (with varying degrees of interaction), offering numerous advantages [2, 11, 12]. Thus, there is a need for scientific research to address these challenges and explore the benefits of automating the analysis [21, 22, 24] and modeling processes in the field of prosthetics for efficient and personalized solutions [18, 22, 26].

In the realm of polygonal model analysis, a crucial undertaking revolves around the delineation of anatomical landmarks – distinctive elements corresponding to specific anatomical structures. In the context of the human upper limb, fingers stand out as prominent and recognizable landmarks. Consequently, the simplest constituent of a polygonal model manifests as the distal segment of a finger, which aptly encapsulates the digit's 'culmination.' Thus, the fundamental objective entails establishing correlations among the five anatomical markers of a reference model – representing the fingers' extremities – and the corresponding five markers within a target model. It is noteworthy to highlight that a principal challenge stems from the irregularity inherent in the input structure, coupled with the absence of a well-defined regular grid-like pattern. This intricacy substantially complicates the applicability of solutions rooted in Convolutional Neural Networks (CNNs). The irregular and often unstructured nature of anatomical models poses a unique hurdle, necessitating innovative approaches that transcend the conventional paradigms of neural network utilization. This foundational alignment not only serves as a cornerstone for subsequent analyses but also paves the way for a myriad of insights. Such alignments facilitate advanced examinations, spanning the segmentation of complete fingers, delineation of distinct anatomical segments, and extraction of specific anatomical characteristics. Consequently, the primary thrust of this work materialized as an exploration of methodologies to uncover correspondences between digitized hand models, coupled with a rigorous assessment of their efficacy.

By embarking on this journey, we aim to harness the potential of these correspondences as a potent tool for various applications, ranging from personalized prosthetic design to ergonomic studies and beyond. Through a comprehensive exploration of these inter-model relationships, we endeavor to unlock novel avenues for enhancing the precision, utility, and impact of anatomical modeling in both research and practical applications [10, 13, 25].

1. Materials and methods

The foundation of this study was established using the MANO dataset [15], comprising three-dimensional polygonal hand models, as depicted in figure 1.

a b c d

Fig. 1. Visualization of initial hand scans in the zero position: a) models "on top of each other", b) 01_01r.ply, c) 06_01r.ply, d) 09_01r.ply

Specifically selected for this investigation, the dataset features a diverse range of hand scans, encompassing various poses. Furthermore, it is pertinent to note that exclusively scans of the right hand were employed in this study. This deliberate selection serves to streamline the complexity of the task by focusing on a singular hand configuration, enhancing the interpretability of results and facilitating a more targeted analysis For the initial phase of research, data corresponding to an open-palm pose were utilized, effectively considered as the 'neutral' position. Moreover, a subset of the 3D models was chosen, as outlined in table 1.

Each polygonal model underwent a simplification process, transforming them into point clouds while preserving key metrics, as detailed in table 2. MeshLab [3] was employed for this simplification procedure, utilizing the Quadric Collapse Edge Decimation function [4]. This technique facilitates the reduction of computational complexity by intelligently collapsing edges, resulting in streamlined point cloud structures.

Consequently, the resultant three-dimensional point clouds (Fig. 2) exhibited the quantifiable attributes outlined in table 2. These parameters serve to expedite calculations, enabling quicker processing times and the judicious utilization of system memory resources. This streamlined data representation proves essential for the subsequent algorithmic phases, enhancing computational efficiency and scalability while laying the groundwork for more advanced analyses.

Additionally, each such model underwent further annotation, encompassing the identification of vertices corresponding to the distal segment of each finger, essentially the fingertip. In this process, due consideration was given to the inherent variability within the dataset, accounting for nuances and potential ambiguities. Notably, the annotation process revealed a nuanced characteristic: the correspondence between specific finger tips and vertices is not always a one-to-one relationship, but rather a discernible grouping.

a b c d

Fig. 2. Visualization of optimized hand point clouds in the zero position: a) models "on top of each other", b) 01_01r.ply, c) 06_01r.ply, d) 09_01r.ply

Table 1. Quantitative indicators of the initial dataset

File Name	Vertices count	Faces count	File Name	Vertices count	Faces count
01_01r.ply	39003	77592	34_01r.ply	42745	85083
06_01r.ply	37398	74338	35_01r.ply	36082	71574
09_01r.ply	48553	96714	36_01r.ply	43123	85771
15_01r.ply	43919	87393	39_01r.ply	43456	86501
18_01r.ply	49572	98530	40_01r.ply	51491	102430
27_01r.ply	36166	72027	41_01r.ply	32189	64069
30_01r.ply	39955	79447	43_01r.ply	45013	89534
32_01r.ply	43597	86500	50_01r.ply	41846	83340

Table 2. Quantitative indicators of processed point clouds

File Name	Vertices count	File Name	Vertices count	File Name	Vertices count
01_01r.ply	2019	30_01r.ply	2014	40_01r.ply	2020
06_01r.ply	2018	32_01r.ply	2021	41_01r.ply	2019
09_01r.ply	2018	34_01r.ply	2028	43_01r.ply	2018
15_01r.ply	2017	35_01r.ply	2025	50_01r.ply	2021
18_01r.ply	2019	36_01r.ply	2018		
27_01r.ply	2018	39_01r.ply	2020		

Furthermore, in tandem with vertex selection, the annotation also involved a subjective assignment of radii, delineating the region associated with the fingertip. This iterative and subjective process aimed to capture the spatial extent accurately. This meticulous annotation methodology yielded the dataset detailed in Table 3, encapsulating the crucial spatial attributes of the distal finger segments across the selected models.

Figure 3 illustrates two plots: the distribution of radii for each of the fingers, as well as the vertex count distribution for each of the fingers.

An instance of annotated distal fingertip regions from a specific subset of the right-hand dataset is presented in figure 4.

As evident from the illustrations, the hand coordinate systems exist in diverse planes and lack a shared orientation. Hand dimensions, finger orientations, and deformations exhibit non-uniform variations. Consequently, the task of aligning corresponding markers between one point cloud and another is inherently intricate.

Table 3. Annotated fingertip data

File Name	Thumb Vertex Count	Thumb Radius, (Rel. Units)	Index Vertex Count	Index Radius, (Rel. Units)	Middle Vertex Count	Middle Radius, (Rel. Units)	Ring Vertex Count	Ring Radius, (Rel. Units)	Little Vertex Count	Little Radius, (Rel. Units)
01_01r.ply	11	0.006533	5	0.004197	7	0.005311	8	0.005311	4	0.00344
09_01r.ply	12	0.007458	13	0.007417	12	0.007417	7	0.005268	12	0.005398
39_01r.ply	17	0.007024	8	0.005022	9	0.006026	9	0.006886	9	0.00506
27_01r.ply	18	0.007211	18	0.007293	10	0.006537	13	0.006179	6	0.004006
43_01r.ply	15	0.007978	10	0.006492	8	0.005536	9	0.006407	5	0.005263
40_01r.ply	13	0.007393	13	0.007139	8	0.006091	8	0.006153	13	0.004701
50_01r.ply	13	0.006932	8	0.004814	9	0.005315	11	0.005976	12	0.006107
36_01r.ply	19	0.007367	12	0.007282	12	0.006313	5	0.004762	6	0.004587
30_01r.ply	20	0.006142	21	0.007532	12	0.006327	12	0.005476	7	0.004163
18_01r.ply	23	0.007729	8	0.005478	7	0.005707	9	0.006218	8	0.005185
35_01r.ply	16	0.007708	15	0.007209	10	0.005076	21	0.007485	6	0.004445
32_01r.ply	15	0.006418	14	0.007742	8	0.00537	7	0.005344	6	0.005175
41_01r.ply	12	0.005582	16	0.006202	13	0.006425	12	0.005168	8	0.003938
06_01r.ply	18	0.00699	12	0.006861	11	0.005738	10	0.005358	8	0.004345
15_01r.ply	32	0.008502	12	0.006458	9	0.006162	9	0.00611	6	0.004455
34_01r.ply	27	0.008376	14	0.007543	13	0.007705	8	0.005127	3	0.004349

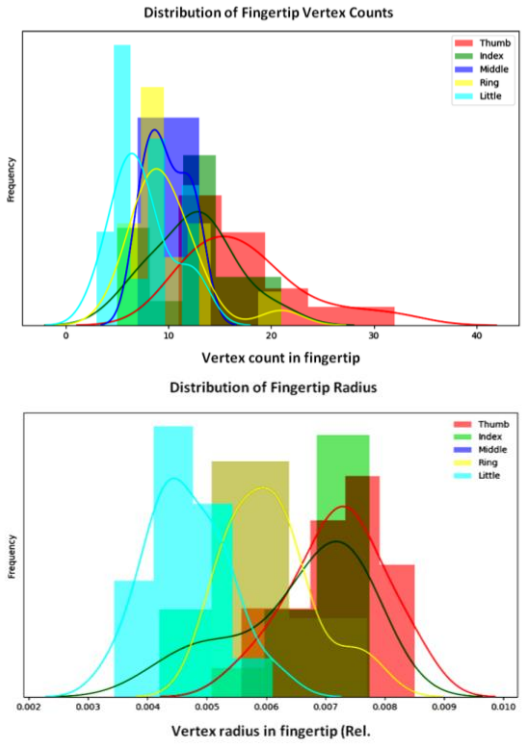
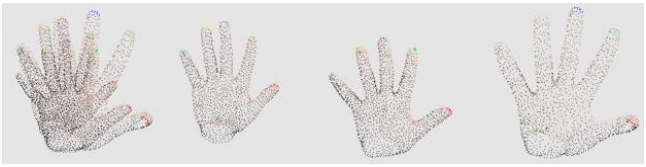


Fig. 3. Comparison of Fingertip Radius and Vertex Distribution



a b c d

Fig. 4. Example of annotated point clouds of a 3D hand model: a) models "on top of each other", b) 01_01r.ply, c) 06_01r.ply, d) 09_01r.ply

Thus, the fundamental 3D hand model can be generalized as a point cloud representation ($X_1 \in \mathbb{R}^3$):

$$X_1 = \{(x_{11}, y_{11}, z_{11}), (x_{12}, y_{12}, z_{12}), ..., (x_{1m}, y_{1m}, z_{1m})\} \quad (1)$$

where x_{11}, y_{11}, z_{11}, x_{1m}, y_{1m}, x_{1m} are coordinates of corresponding point of point cloud.

Similarly, the target 3D hand model is described by a point cloud representation ($X_2 \in \mathbb{R}^3$):

$$X_2 = \{(x_{21}, y_{21}, z_{21}), (x_{22}, y_{22}, z_{22}), ..., (x_{2m}, y_{2m}, z_{2m})\} \quad (2)$$

For each finger tip i, let A_{1i}, and A_{2j} be anatomical landmarks represented as subsets of points in their respective point clouds.

$$A_{1i}, X_1, \text{ for } i = 1, 2, ..., 5 \quad (3)$$

and:

$$A_{2j}, X_2, \text{ for } j = 1, 2, ..., 5 \quad (4)$$

The goal is to find a mapping function $f : A_{1}i \rightarrow A_{2}j$, where f establishes correspondence between anatomical landmarks:

$$f: A_{1i} \rightarrow A_{2j}, \text{ for } j \in \{2, 3, 4, 5\} \quad (5)$$

Moreover, it should be noted that to evaluate the performance of the function *f*, a quality assessment criterion must be chosen. Given the substantial complexity of such a function and its inherent ambiguity, we will employ the following loss function:

$$Loss_{logic}(A_{1i}, B_{1ik}) = [A_{1i} \cap B_{1ik} \neq \emptyset] \quad (6)$$

where A_{1i} is the set corresponding to annotated ground-truth data for anatomical area *i*, B_{1ik} is the predicted set for area *i* using the mapping function *f* from point cloud *k*.

Furthermore, we will utilize the Intersection over Union (IoU) evaluation criterion (7), which quantifies the extent of vertex overlap between two sets:

$$Loss_{IoU}(A_{1i}, B_{1ik}) = A \cap B \,/\, A \cup B \quad (7)$$

As a function for finding the "closest" vertex from one point cloud to another, we will employ the Gromov-Wasserstein distance over distance matrices obtained from these point clouds [4, 13]:

$$GW(P,Q) = \inf_{\pi \in \Pi(X,Y)} \sum_{(x,y) \in X \times Y} c(x,y)\pi(x,y) \quad (8)$$

where *GW(P,Q)* represents the Gromov-Wasserstein distance between distributions *P* and *Q*; *X,Y* are spaces on which the distributions *P* and *Q* defined; *Π(X,Y)* is the set of all possible transport plans between distributions *P* and *Q*; *c(x,y)* is the metric or distance between objects *x* and *y*; π(x,y) is an element of the transport plan indicating how much mass is transported from object *x* to object *y*.

2. Results and discussion

Thus, for each annotated set of vertexes in the point cloud, a search for correspondences was performed on another point cloud, resulting in corresponding $Loss_{IoU}$ metrics for each finger tip (Fig. 5).

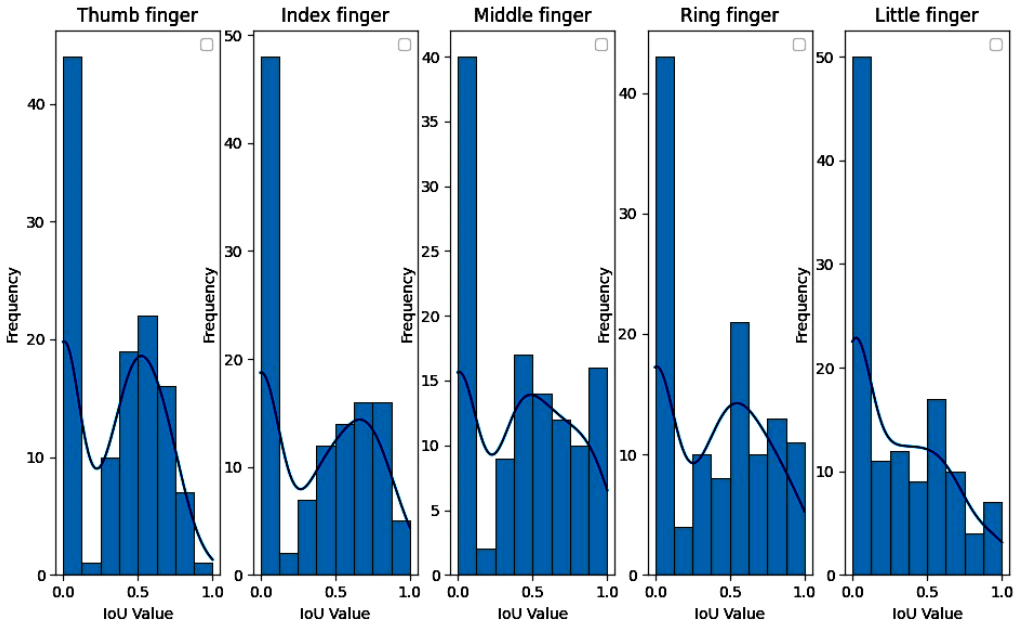


Fig. 5. Visualization of Loss IoU distribution for different fingers

The graph indicates that achieving a 100% match of annotated points is not feasible, although we consider a match of at least one vertex from the set to be sufficient. The number of matches can be increased by enlarging the radius of annotations, although this may not always be acceptable. Additionally, the distributions for all fingers appear similar, implying a consistent error pattern in both erroneously and correctly annotated sets.

An example of annotated sets of anatomical landmarks for fingertips, along with the corresponding sets obtained using equation (8), is illustrated in Fig. 6.

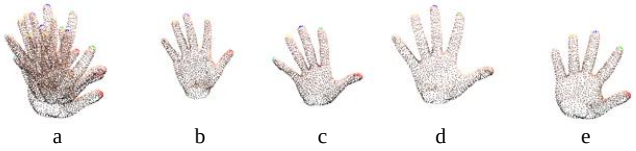


Fig. 6. Visualization of accurate matches in 3D Hand Model Point Clouds: a) models "on top of each other", b) ground truth 01_01r.ply, c) predicted 06_01r.ply from 01_01r.ply, d) 09_01r.ply predicted from 01_01r.ply, e) 15_01r.ply predicted from 01_01r.ply

As evident from the figure, the annotations correspond to anatomical landmarks, indicating the viability of the proposed approach in automating the search for anatomical landmarks. Additionally, the high level of stability of this approach is indicated by quantitative metrics, namely the $Loss_{logic}$, which is presented in the form of a diagram in Fig. 7.

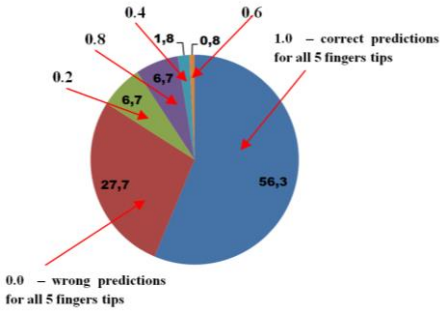


Fig. 7. Visualization of distribution of $Loss_{logic}$ in the test dataset

It should be noted that there exists a certain percentage of "inaccuracies" in the predictions. Such a level of errors can be considered acceptable for the initial stage of research and development. Furthermore, a certain level of errors is attributed to the low point cloud density and small annotation radius.

Furthermore, as evident from Fig. 8, these inaccuracies are partly attributed to the presence of certain "noisy elements" in the data, indicating the potential utilization of this method during pre-processing.

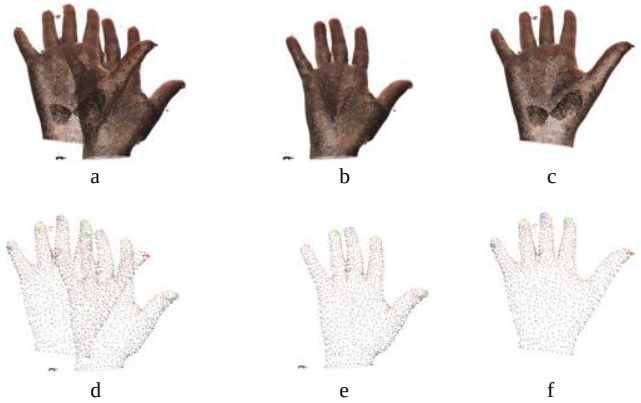


Fig. 8. Visualization of inaccurate matches in 3D Hand Model Point Clouds: a) original models "on top of each other", b) original reference model 34_01r.ply with "noise island", c) original target model 30_01r.ply, d) annotated models "on top of each other", e) ground truth annotation of reference model 34_01r.ply, f) inaccurate annotation prediction for model 30_01r.ply

3. Conclusions

The conducted research has thoroughly examined the quantitative indicators pertaining to the potential utilization of the Gromov-Wasserstein distance function for the determination of corresponding segments within a model (point clouds). These indicators attest to the promising nature of the method for applications in the development of automated prosthetic modeling tools. Furthermore, it is essential for future investigations to explore the impact of point cloud density as well as the variability in hand poses. Many patients are unable to maintain their hand in a specific pose due to various injuries, necessitating a matching function for anatomical markers that remains effective irrespective of the chosen hand pose. Such approaches can be used for image processing in a variety of medical applications, including disaster medicine and microscopic image analysis [19, 23]. The prospects of this work encompass the practical implementation of the developed method in the realm of modeling and designing personalized prosthetic devices. As such, this method holds potential to significantly enhance the accuracy and efficiency of prosthetic modeling processes, thus contributing to improved patient care and rehabilitation outcomes.

References

[1] Avrunin O. G. et al.: Application of 3D printing technologies in building patient-specific training systems for computing planning in rhinology. Proceedings of the International Scientific Internet Conference on Computer Graphics and Image Processing and 48th International Scientific and Practical Conference on Application of Lasers in Medicine and Biology, 2019, 1 [https://doi.org/10.1201/9780429057618-1].
[2] Boleneus G. J. et al.: Top-down design enables flexible design of prosthetic forearms and hands. ASEE Annual Conference and Exposition, Conference Proceedings, 2019.
[3] Cignoni P. et al.: MeshLab: an Open-Source Mesh Processing Tool. Sixth Eurographics Italian Chapter Conference, 2008, 129–136.
[4] Garland M., Heckbert P. S.: Simplifying surfaces with color and texture using quadric error metrics. Proceedings Visualization 98, 2000.
[5] Guidi G., Gonizzi S., Micoli L.: 3D capturing performances of low-cost range sensors for mass-market applications. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 41, 2016, 33–40 [https:// doi.org/10.5194/isprsarchives-XLI-B5-33-2016].
[6] Kim Y. et al.: Dynamic elasticity measurement for prosthetic socket design. International Conference on Rehabilitation Robotics – ICORR, London, UK, 2017, 1281–1286 [https://doi.org/10.1109/ICORR.2017.8009425].
[7] Neri P. et al.: Semi-automatic Point Clouds Registration for Upper Limb Anatomy. International Joint Conference on Mechanics, Design Engineering and Advanced Manufacturing – JCM, 2023, 733–742 [https://doi.org/10.1007/978-3-031-15928-2_64].
[8] Neri P. et al.: 3D scanning of Upper Limb anatomy by a depth-camera-based system. International Journal on Interactive Design and Manufacturing, 2023 [https://10.1007/s12008-023-01248-1].
[9] Olsen J. et al.: 3D-Printing and Upper-Limb Prosthetic Sockets: Promises and Pitfalls. IEEE Transactions on Neural Systems and Rehabilitation Engineering 29, 2021, 527–535 [https://doi.org/10.1109/tnsre.2021.3057984].
[10] Pavlov S. V. et al.: Photoplethysmohrafic technologies of the cardiovascular control. Universum-Vinnitsa, Vinnitsa, 2007.
[11] Pavlov S. V. et al.: A simulation model of distribution of optical radiation in biological tissues. Visnyk VNTU 3, 2011, 191–195.
[12] Pavlov S. V. et al.: Laser photoplethysmography in integrated evaluation of collateral circulation of lower extremities. Proc. SPIE 8698, 2012, 869808.
[13] Peyré G., Cuturi M., Solomon J.: Gromov-Wasserstein averaging of kernel and distance matrices. International Conference on Machine Learning (ICML), 2016.
[14] Román-Casares A. M., García-Gómez O., Guerado E.: Prosthetic Limb Design and Function: Latest Innovations and Functional Results. Current Trauma Reports 4(4), 2018, 256–262 [https://doi.org/10.1007/s40719-018-0150-2].
[15] Romero J., Tzionas D., Black M. J.: Embodied hands: Modeling and capturing hands and bodies together. ACM Transactions on Graphics 36(6), 2017, 245 [https://doi.org/10.1145/3130800.3130883].
[16] Ryniewicz A. et al.: The use of laser scanning in the procedures replacing lower limbs with prosthesis. Measurement 112, 2017, 9–15.
[17] Selivanova K. G. et al.: 3D visualization of human body internal structures surface during stereo-endoscopic operations using computer vision techniques. Przeglad Elektrotechniczny 9, 2021, 30–33 [https://doi.org/10.15199/48.2021.09.06].
[18] Serkova V. et al.: Medical expert system for assessment of coronary heart disease destabilization based on the analysis of the level of soluble vascular adhesion molecules. Proc. SPIE 10445, 2017, 104453O.
[19] Sokol Y. et al.: Using medical imaging in disaster medicine. Proceedings of IEEE 4th International Conference on Intelligent Energy and Power Systems – IEPS 2020, 287–290 [https://doi.org/10.1109/IEPS51250.2020.9263175].
[20] Tymkovych M. et al.: Ice crystals microscopic images segmentation based on active contours. 2019 IEEE 39th International Conference

on Electronics and Nanotechnology – ELNANO 2019, 493–496 [https://doi.org/10.1109/ELNANO.2019.8783332].
[21] Tymkovych M. et al.: Detection of Chest Deviation During Breathing Using a Depth Camera. Proceedings of IEEE 8th International Conference on Problems of Infocommunications, Science and Technology – PIC S and T, 85 [https://doi.org/10.1109/PICST54195.2021.9772111].
[22] Tymkovych M. et al.: Application of SOFA Framework for Physics-Based Simulation of Deformable Human Anatomy of Nasal Cavity. Proceedings of IFMBE, 2021, 112 [https://doi.org/10.1007/978-3-030-64610-3_14].
[23] Tymkovych M. et al.: Application of Artificial Neural Networks for Analysis of Ice Recrystallization Process for Cryopreservation. Proceedings of IFMBE, 2021, 102 [https://doi.org/10.1007/978-3-030-64610-3_13].
[24] Wojcik W. et al.: ECTL application for carbon monoxide measurements. Proc. of SPIE 5958, 2005, 595837.
[25] Xu H. et al.: Gromov-wasserstein learning for graph matching and node embedding. International Conference on Machine Learning – ICML, 2019.
[26] Zabolotna N. et al.: Diagnostic efficiency of Mueller-matrix polarization reconstruction system of the phase structure of liver tissue, Proc. SPIE 9816, 2015, 98161E [https://doi.org/10.1117/12.2229018].
[27] Zanuttigh P. et al.: Time-of-Flight and Structured Light Depth Cameras Technology and Applications. Springer, 2016.

Ph.D. Maksym Tymkovych
e-mail: maksym.tymkovych@nure.ua

Ph.D., senior lecturer of the Department of Biomedical Engineering, Kharkiv National University of Radio Electronics, Ukraine.
Research interests: 3D visualization, biomedical image analysis, medical robotics, deep learning, artificial intelligence, computer-assisted surgery.

https://orcid.org/0000-0001-5613-1104

Prof. Oleg Avrunin
e-mail: oleh.avrunin@nure.ua

Doctor of Technical Sciences, professor, Head of Biomedical Engineering Department, Kharkiv National University of Radio Electronics, Ukraine. Scientific supervisor of research work on research of theoretical and technical principles of diagnostics, assessment and correction of medical and social human conditions. Invited Professor in Gottfried Wilhelm Leibniz Universität Hannover (Germany) and Harbin Engineering University (China).

https://orcid.org/0000-0002-6312-687X

Ph.D. Karina Selivanova
e-mail: karina.selivanova@nure.ua

Ph.D., associate professor of the Department of Biomedical Engineering, Kharkiv National University of Radio Electronics, Ukraine.
Research interests: methods of automated processing and analysis of biomedical data, study of individual characteristics of fine motor skills of hands and psychomotor skills, development of tests, complexes for psychoneurology, virtual simulators. Have more than 60 scientific publications and 1 patent of Ukraine. Participation in the international project DAAD No. 54364768 (2013), (Hannover, Germany).

https://orcid.org/0000-0003-1002-0761

D.Sc. Alona Kolomiiets
e-mail: alona.kolomiets.vnt@gmail.com

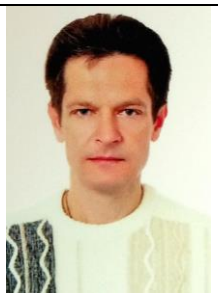
Doctor of Pedagogical Sciences, associate professor, professor of Department of Higher Mathematics, Vinnytsia National Technical University.
Research interests: professional pedagogy, the fundamentalization of the mathematical training of future bachelors of technical specialties, mathematical modeling, methods of statistical analysis of experimental data.

http://orcid.org/0000-0002-7665-6247

Ph.D. Taras Bednarchyk
e-mail: taras2010.2017@gmail.com

Vinnytsia Pyrohov National Medical University, Ph.D. (Philosophy), associate professor.
Research interest: medical ethics, bioethics, philosophy of science, methodology of scientific research, social psychology.

https://orcid.org/0000-0003-2336-4635



Ph.D. Saule Smailova
e-mail: Saule_Smailova@mail.ru

Saule Smailova is currently a lecturer at the School of Digital Technologies and Artificial Intelligence D.Serikbayev East Kazakhstan University, Ust-Kamenogorsk, Kazakhstan. She is a co-author over 60 papers in journals, book chapters, and conference proceedings. Member of Expert Group in the Computer Science specialization of IQAA.
Her professional interests are teaching, artificial intelligence, software engineering, data processing.

https://orcid.org/0000-0002-8411-3584

http://doi.org/10.35784/iapgos.5775 *received: 08.01.2024 | revised: 29.01.2024 | accepted: 07.03.2024 | available online: 31.03.2024*

GENETIC ALGORITHM-BASED DECISION TREE OPTIMIZATION FOR DETECTION OF DEMENTIA THROUGH MRI ANALYSIS

Govada Anuradha, Harini Davu, Muthyalanaidu Karri
Velagapudi Ramakrishna Siddhartha Engineering College, Department of Computer Science and Engineering, Kanuru, Vijayawada, India

Abstract. *Dementia is a devastating neurological disorder that affects millions of people globally, causing progressive decline in cognitive function and daily living activities. Early and precise detection of dementia is critical for optimal dementia therapy and management however, the diagnosis of dementia is often challenging due to the complexity of the disease and the wide range of symptoms that patients may exhibit. Machine learning approaches are becoming progressively more prevalent in the realm of image processing, particularly for disease prediction. These algorithms can learn to recognize distinctive characteristics and patterns that are suggestive of specific diseases by analyzing images from multiple medical imaging modalities. This paper aims to develop and optimize a decision tree algorithm for dementia detection using the OASIS dataset, which comprises a large collection of MRI images and associated clinical data. This approach involves using a genetic algorithm to optimize the decision tree model for maximum accuracy and effectiveness. The ultimate goal of the paper is to develop an effective, non-invasive diagnostic tool for early and accurate detection of dementia. The GA-based decision tree, as proposed, exhibits strong performance compared to alternative models, boasting an impressive accuracy rate of 96.67% according to experimental results.*

Keywords: dementia, genetic algorithm, decision tree

OPTYMALIZACJA DRZEWA DECYZYJNEGO OPARTA NA ALGORYTMIE GENETYCZNYM DO WYKRYWANIA DEMENCJI POPRZEZ ANALIZĘ MRI

Streszczenie. *Demencja jest wyniszczającym zaburzeniem neurologicznym, które dotyka miliony ludzi na całym świecie, powodując postępujący spadek funkcji poznawczych i codziennych czynności życiowych. Wczesne i precyzyjne wykrywanie demencji ma kluczowe znaczenie dla optymalnej terapii i zarządzania demencją, jednak diagnoza demencji jest często trudna ze względu na złożoność choroby i szeroki zakres objawów, które mogą wykazywać pacjenci. Podejścia oparte na uczeniu maszynowym stają się coraz bardziej powszechne w dziedzinie przetwarzania obrazu, szczególnie w zakresie przewidywania chorób. Algorytmy te mogą nauczyć się rozpoznawać charakterystyczne cechy i wzorce, które sugerują określone choroby, analizując obrazy z wielu modalności obrazowania medycznego. Niniejszy artykuł ma na celu opracowanie i optymalizację algorytmu drzewa decyzyjnego do wykrywania demencji przy użyciu zbioru danych OASIS, który obejmuje duży zbiór obrazów MRI i powiązanych danych klinicznych. Podejście to obejmuje wykorzystanie algorytmu genetycznego do optymalizacji modelu drzewa decyzyjnego w celu uzyskania maksymalnej dokładności i skuteczności. Ostatecznym celem artykułu jest opracowanie skutecznego, nieinwazyjnego narzędzia diagnostycznego do wczesnego i dokładnego wykrywania demencji. Zaproponowane drzewo decyzyjne oparte na GA wykazuje wysoką wydajność w porównaniu z alternatywnymi modelami, szczycąc się imponującym współczynnikiem dokładności wynoszącym 96,67% zgodnie z wynikami eksperymentalnymi.*

Słowa kluczowe: demencja, algorytm genetyczny, drzewo decyzyjne

Introduction

Dementia represents a progressive neurodegenerative state characterized by a slow decline in cognitive functions, affecting memory, reasoning, behavior, and the ability to carry out everyday tasks. The World Health Organization (WHO) estimates that approximately 50 million individuals globally are presently grappling with dementia, and this number is projected to triple by the year 2050 [39]. As the population continues to age, dementia has emerged as a prominent public health issue, presenting difficulties in both the timely identification and the development of effective treatments. The timely identification of dementia is pivotal in facilitating prompt interventions, tailored treatment, and enhancing patient outcomes. Over the past few years, medical imaging methods have surfaced as potent instruments for detecting dementia. Specifically, Magnetic Resonance Imaging (MRI) delivers intricate perspectives on the structural and functional modifications in the brain that are linked to dementia. By analyzing MRI data, we can identify specific biomarkers and patterns that aid in the detection and classification of dementia cases.

Similar to the research conducted by Javeed, A. et al. of [22] which demonstrates the potential of machine learning techniques for improving the accuracy of dementia diagnosis through MRI analysis, our paper seeks to leverage genetic algorithms to optimize decision tree models for more accurate and efficient detection of dementia, addressing the limitations of traditional diagnostic methods. Nowroozpoor, A. et al. [32] highlights the challenge of detecting cognitive disorders in a clinical setting can reduce the effects of dementia. So our paper aims to develop an optimized decision tree model using genetic algorithms to improve the accuracy of dementia detection through MRI analysis.

The goal is to enhance the accuracy and reliability of dementia diagnosis by leveraging advanced machine learning techniques. By optimizing the decision tree model using a Genetic Algorithm, we aim to overcome the limitations of traditional decision tree approaches and improve the accuracy of dementia detection. Traditional decision tree models often struggle to accurately classify dementia cases due to the intricate patterns and high-dimensional nature of MRI data. To address this challenge, we propose the utilization of a Genetic Algorithm, which draws inspiration from natural evolution, to fine-tune the decision tree's hyperparameters. The optimization process is designed to identify the best set of hyperparameters that enhances the model's effectiveness in detecting dementia.

1. Literature review

Garcia-Gutierrez et al. [19] have introduced GA-MADRID, a Python-based framework for Alzheimer's Disease and Frontotemporal Dementia diagnosis. This framework integrates data processing, feature engineering, predictive modeling, and multi-class decision-making modules. It employs Genetic Algorithms for feature selection and machine learning algorithm optimization, along with evolutionary grammars and Bayesian networks for decision-making. This work holds promise for advancing the field of medical diagnosis in neurodegenerative diseases.

In Basheer et al. study [8], the primary focus was on dementia prediction using longitudinal MRI data from the OASIS dataset. A modified capsule network, known for its efficacy in image classification, was employed. Before model development, comprehensive exploratory data analysis was carried out to identify crucial features, assess data correlations and density, and hierarchically examine various factors in the dataset. Singular Value Decomposition (SVD) matrix decomposition was utilized for feature extraction, reducing dimensionality while preserving essential information. The resulting model exhibited exceptional performance, achieving an impressive accuracy rate of 92.39%,

surpassing other contemporary machine learning models. These findings hold significant promise for early dementia detection and intervention, potentially improving patient outcomes through timely treatments.

Within the realm of cognitive impairment research, Liang et al. [24] delved into the innovative application of Amazon Alexa as a Voice-Assistant System (VAS) for capturing voice commands from older adults, aged 65 and above, encompassing both individuals with healthy cognitive function and those with Mild Cognitive Impairment (MCI). The main aim of this study was to create machine learning models for classification and then evaluate their performance in comparison to cognitive assessment scores. Using a range of machine learning algorithms such as neural networks, support vector machines, decision trees, and random forests, Liang's research investigated the use of VAS data in conjunction with advanced computational methods.

In the work conducted by Perovnik et al. [33], an extensive review was carried out to identify recent English-language publications within the past five years, drawing upon databases such as PubMed and IEEE Xplore. The review primarily focused on research papers investigating the integration of Artificial Intelligence (AI) into computerized cognitive tests. Notably, the findings from these studies revealed substantial improvements in discrimination sensitivity, increasing by 4%, and specificity, by 3%, when AI-based cognitive tests were employed. Of particular significance was the integration of acoustic and linguistic features within speech, conversation, and language tests, which yielded an impressive accuracy rate of 94%. Additionally, the application of deep learning techniques in the analysis of brain scans demonstrated remarkable accuracy, approaching 92%. These results underscore the transformative potential of AI-based methodologies in enhancing the accuracy and efficiency of cognitive assessment, carrying significant implications for the early detection and intervention in cognitive impairments.

Diaz-Alvarez et al. [13] introduced an innovative methodology for precise diagnoses of Alzheimer's disease (AD), Fronto Temporal Dementia (FTD), and related disorders utilizing FDG-PET imaging. This approach entailed a comprehensive comparative analysis of genetic algorithms and principal component analysis (PCA) techniques, each tailored with K-Nearest Neighbor and Bayes Networks. To validate their effectiveness, the researchers employed K-fold cross-validation and external validation approaches. The study unequivocally affirmed the potency of genetic algorithms in accurately diagnosing AD, FTD, and related disorders based on FDG-PET imaging data. These findings underscore the significant potential of genetic algorithms as a powerful tool in neuroimaging analysis, offering promise for the precise classification and diagnosis of neurodegenerative disorders, thus advancing the field of diagnostic accuracy and precision.

Mirzaei et al. [29] conducted a thorough review of journal articles published since 2016, focusing on the application of machine learning techniques for Alzheimer's Dementia (AD) detection. This review critically analyzed various approaches, including Support Vector Machine, Random Forest, Convolutional Neural Network, and K-means, employed in these studies. While the review highlighted the absence of a single best approach for AD diagnosis due to the disease's complexity, it noted the potential of deep learning techniques, particularly Convolutional Neural Networks, in improving accuracy and effectiveness. These advanced techniques allow for intricate pattern and feature extraction from imaging data, offering the promise of more accurate and reliable AD diagnosis. The review emphasized the importance of continued research and development in the field of machine learning for AD diagnosis to advance our understanding and enhance patient care.

Javeed et al. [22] conducted a systematic literature review (SLR) that spanned the years 2011 to 2022, focusing on machine learning (ML)-based automated diagnostic systems designed for dementia prediction. Throughout this rigorous review, the selected research articles underwent critical analysis and discussion to evaluate the effectiveness of ML-driven automated diagnostic systems in predicting dementia. The review's specific emphasis was placed on comparing the performance of various data modalities, including images, clinical features, and voice data, when integrated with ML-based models for dementia prediction.

An alternative method was developed by Angelillo et al. [5] using an Automated Measurement Tool (AMT) implemented on a digitizing tablet that incorporated an electronic pen. This novel approach enabled the acquisition of additional measures beyond traditional paper-based test observations. Handwriting measures were obtained through the digitizing tablet, which served as input for machine learning algorithms aimed at automating disease detection. The ensemble approach demonstrated notable performance with an area under the curve (AUC) of 87.30% and a sensitivity rate of 86.11%. The outcomes underscore the efficacy of the AMT-based approach and underscore the potential of incorporating digitizing tablets and machine learning algorithms for the automated detection of the specified disease.

Varoquaux et al. research [38] critically examines challenges in computer analysis of medical images. Focusing on data limitations and research incentives, the paper discusses potential biases at every step. Drawing evidence from the literature and data challenges, the study offers insights into the pervasive nature of biases in the field. Despite challenges, the paper highlights ongoing efforts to counteract these issues and provides recommendations for future improvement. This work contributes to a nuanced understanding of obstacles and potential solutions in advancing computer analysis for medical imaging and patient health.

Nori et al. [31] conducted research focusing on predicting incident mild cognitive impairment, Alzheimer's Disease, and related dementias using structured data from administrative and electronic health records. The study, encompassing a cohort of patients and controls, employed gradient boosting machine, lightGBM, for model fitting. The 2-year model demonstrated a sensitivity of 47% and an impressive area-under-the-curve of 87%. Even in the 3-year model, the sensitivity remained notable at 24%. While this dropped to 15% in year 8, the model maintained a commendable AUC of 72%. The study suggests the model's potential as a valuable tool for screening patients in trial recruitment and management. Nori, V. S.'s methodology showcases the application of machine learning in predicting dementia-related conditions from diverse healthcare data sources.

Bansal et al. [7] contributes to the emerging field of applying machine learning to neurodegenerative disorders, with a particular focus on dementia. The study presents a comparative analysis of four machine learning algorithms – J48, Naïve Bayes, Random Forest, and Multilayer Perceptron – utilizing CFSSubsetEval for attribute reduction. The results highlight J48 as the most effective algorithm for dementia detection among the tested methods. This comparative assessment underscores the potential of machine learning techniques, particularly J48, in the early detection of dementia – a crucial step in addressing this global health issue. The last page should be filled at least 50%.

In the study by So et al. [36], a two-layer model inspired by dementia support centers is proposed for early dementia diagnosis. Utilizing data from dementia screenings at the Gangbuk-Gu center in the Republic of Korea, the model incorporates patient information, MMSE-KC dementia screening results, and CERAD-K for precise testing. The two-stage classification involves initial categorization of MMSE-KC data into normal and abnormal, followed by the use of CERAD-K data for dementia and mild cognitive impairment classification. Comparative analysis with various algorithms, including Naive Bayes, Random Forest, and Multilayer Perceptron (MLP), shows that MLP achieves the highest F-measure values for normal cases, while SVM exhibits the highest values for mild cognitive impairment and dementia. The proposed early diagnosis model is positioned as a time and cost-effective solution, simplifying the dementia diagnostic process.

Al-Badarneh et al. [3], presents a hybrid approach for the classification of Magnetic Resonance Imaging (MRI) brain images. The study utilizes decision tree (DT) and genetic algorithms (GA) to construct a binary classifier categorizing MRIs as normal or abnormal. The methodology comprises three stages: texture features extraction, features reduction through principal component analysis, and MRI classification using the proposed DT/GA approach. The classifier is evaluated on a benchmark MRI dataset of 710 brain images from Harvard Medical School, demonstrating significant accuracy improvements. Results indicate an accuracy of 96.31% with the decision tree and a notable improvement to 98.55% with the proposed DT/GA approach. This approach showcases the efficacy of combining DT and GA for accurate classification of MRI brain images.

Miled et al. [27], explores the transformation of medical notes into dementia risk classifiers, emphasizing their potential in predicting the risk of developing chronic diseases. The study utilizes TF-ICF for keyword selection, employing two encodings for summarizing the notes: the average vector embedding using BERT or Clinical BERT, and aggregation based on UMLS concepts. Misspellings of keywords are considered for enhanced predictive performance. A neural network is applied to the first encoding, while a gradient boosted trees model is used for the second encoding. Results indicate successful identification of dementia risk one year ahead with an AUC of 75%, though challenges in generalizability to other healthcare institutions are noted. The study underscores the promise of medical notes for risk prediction models, considering factors such as note length.

Mirheidari B. [28] addresses the challenge of dementia detection by proposing an automatic classification system using an intelligent virtual agent (IVA). The study focuses on conversation analysis (CA) during interviews between patients and neurologists, aiming to distinguish between progressive neurodegenerative memory disorder (ND) and non-progressive functional memory disorders (FMD). The manual CA process is resource-intensive, prompting the introduction of an IVA-led approach. Analyzing neurologist-patient and IVA-patient conversations, the study achieves impressive ND/FMD classification rates of 90.0% and 90.9%, respectively. Mirheidari B.'s work underscores the potential of leveraging IVA technology to automate and enhance the efficiency of dementia detection through conversational analysis, providing valuable insights into distinguishing features between human and IVA-led interactions.

Zhao et al. [40] introduced a voice recognition-based digital cognitive screener (DCS) for community-based dementia screening. This proof-of-concept study assesses the reliability, validity, and feasibility of the DCS among Chinese older adults. Participants underwent demographic, clinical, and DCS assessments. Dementia and mild cognitive impairment (MCI) diagnoses were based on the Montreal Cognitive Assessment. The DCS exhibited promising results, showing good internal consistency, test-retest reliability, and inter-rater reliability. Receiver operating characteristic analyses indicated high discriminative validity, with the DCS demonstrating excellent sensitivity (100%) and good specificity (80%) for dementia detection. The findings suggest that the DCS is a reliable and valid tool for digital dementia screening in a community setting, warranting further investigation in larger-scale screenings.

Gorji et al. [20] proposes an effective method for early Alzheimer's disease (AD) diagnosis, crucial for increased eligibility in clinical trials and better planning for patients. Focusing on mild cognitive impairment (MCI) as a key AD risk factor, the study introduces a novel diagnostic approach utilizing pseudo Zernike moments (PZMs) on structural MRI data. PZMs extract discriminative information from AD, MCI, and healthy control (HC) groups. Artificial neural networks, specifically pattern recognition and learning vector quantization (LVQ) networks, classify the extracted information. With 500 MRIs from the ADNI database, the method achieves impressive results, notably 94.88% accuracy, 94.18% sensitivity, and 95.55% specificity for AD vs. MCI, and 95.59% accuracy, 95.89% sensitivity, and 95.34% specificity for MCI vs. HC. This approach showcases significant potential for accurate early AD diagnosis using advanced image analysis and neural networks.

2. Proposed methodology

The methodological flowchart of our suggested model is depicted in Fig. 1. In our study, we utilize the OASIS Longitudinal dataset for dementia detection through MRI analysis. Data preprocessing involves mean imputation for missing values and one-hot encoding for categorical variables. Principal Component Analysis (PCA) is employed for feature extraction, mitigating high-dimensionality challenges. Our unique approach involves optimizing Decision Tree hyperparameters using Genetic Algorithm (GA). The GA-tuned Decision Tree is then used for model training, resulting in improved performance.

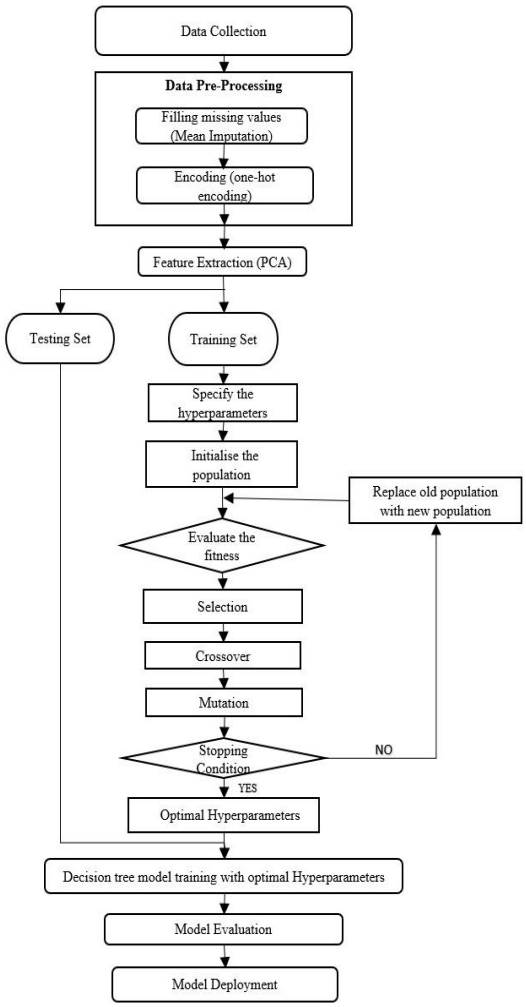


Fig. 1. Proposed system flow diagram

2.1. Data collection

Our data collection revolves around the Oasis dataset [30], encompassing longitudinal MRI and PET images accumulated over a 15-year period from clinical patients. The dataset encompasses 150 subjects aged between 60 and 96, each having undergone multiple scans, resulting in a total of 373 imaging sessions. Each subject has access to 3 or 4 T1-weighted MRI scans. The cohort includes both right-handed men and women. Within this group of 150 subjects, 72 consistently remained nondemented, 64 remained persistently demented, and 14 underwent a change in diagnostic status during the study. In summary, our dataset provides a longitudinal MRI perspective on 150 subjects, capturing diverse diagnostic trajectories. Table 1 outlines diagnoses, including counts for demented, non-demented, and converted individuals, along with gender distribution.

Table 1. Class-labels of Oasis dataset

Diagnosis row labels	Number of individuals	Male	Female
Demented	146	86	60
Non-Demented	190	61	129
Converted	37	37	34

2.2. Data pre-processing

Within our study, we encountered null values within the Oasis dataset during data pre-processing. To ensure data quality, we employed a simple algorithm that systematically identified and replaced these null values with the respective column's mean value. This meticulous approach resulted in a clean dataset devoid of null values, setting the stage for reliable and noise-free analysis as we delve into dementia prediction through MRI analysis.

Data encoding stands as a crucial step in the data pre-processing phase. In our paper, the one-hot encoding technique plays a pivotal role in effectively utilizing the OASIS Longitudinal dataset. It serves as a vital step in the data pre-processing phase, where categorical variables within the dataset are transformed into a numerical format compatible with machine learning models. Through one-hot encoding, each category within a categorical feature is systematically converted into binary variables, ensuring that our dataset is primed for seamless integration with machine learning algorithms. This transformation enhances data compatibility, simplifies computations, and facilitates the subsequent analysis as we embark on dementia prediction through MRI analysis.

2.3. Feature extraction using PCA

In this module, we leverage Principal Component Analysis (PCA) to distill essential information from our pre-processed MRI data. PCA stands as a potent technique for dimensionality reduction, effectively addressing the challenges posed by high-dimensional feature spaces. By transforming our original data into a lower-dimensional representation, PCA retains maximum information while shedding redundant or less informative features. This dimensionality reduction not only enhances computational efficiency but also aids in subsequent optimization and model training phases. It allows us to identify influential features, improving the accuracy and reliability of our dementia detection model by focusing on the most discriminative aspects of the MRI data. The formula [13] for the principal component is characterized as linear combinations of the fundamental variables, which is given by eq. (1).

$$PCA_n = b_{n1} * y_1 + b_{n2} * y_2 + \cdots + b_{nr} * y_r \quad (1)$$

where PCA_n represents the n^{th} principal component, b_{nj} denotes the weights or coefficients associated with the j^{th} original variable y_j, and y_j represents the j^{th} original variable. The formula illustrates how each principal component is formed as a linear combination of the original variables with specific weights.

2.4. Optimization of decision tree hyperparameters using Genetic Algorithm

Our approach involves the optimization of Decision Tree hyperparameters using the Genetic Algorithm (GA) module, a pivotal component in our MRI-based dementia detection framework. Decision Trees, a foundational machine learning technique, rely heavily on hyperparameter settings to determine their structure and predictive performance. These hyperparameters encompass critical factors such as maximum depth, minimum samples required for splits, maximum leaf nodes, and maximum features, each exerting a substantial influence on model accuracy and generalization capabilities. The task of identifying the ideal hyperparameter configuration can be formidable, with traditional methods proving computationally expensive or inadequate is described in Algorithm 1. GA offers an efficient alternative, systematically exploring hyperparameter combinations, uncovering intricate associations, and elevating the Decision Tree's accuracy and reliability in dementia detection is given by eq. (2).

$$Population_{new} = Mutation\Big(Crossover\big(Selection(Population_{old})\big)\Big) \quad (2)$$

Algorithm 1: Hyper parameter tuning using Genetic Algorithm

```
Begin
    Population ← Create the initial population by employing
     random selection.
    Loop: For each iteration
    Begin
        Calculate fitness values for each individual in the population
        in the array.

        p1, p2 ←  Choose parents through bit tournament selection.
        o1, o2 ← single-point crossover between selected parents
        m_o1 ←  offspring1 should be subjected to bit-flip mutation
        m_o2 ←  offspring2 should be subjected to bit-flip mutation
       replace the least suited individuals with their mutant progeny

    End

   Hyper_parameters ← select the most fit people from population

End
```

2.5. Model training with optimized Decision Tree

In this module, the optimized Decision Tree obtained from the previous step is utilized for model training. The optimized Decision Tree refers to the Decision Tree that has undergone hyperparameter tuning using the Genetic Algorithm, resulting in improved performance. Algorithm 2 describes the process of model training.

Algorithm 2 serves as a comprehensive guide, encapsulating the procedural intricacies entailed in this model training endeavor. This extensive process extends beyond the initial optimization through the Genetic Algorithm, encompassing the subsequent training phase where the Decision Tree engages with the dataset. In this phase, the Decision Tree dynamically learns and adapts based on the refined parameters achieved through the optimization process. The nuanced interplay between the Genetic Algorithm and the Decision Tree stands as a sophisticated approach aimed at enhancing the model's discernment and classification accuracy. This holistic methodology underscores a meticulous and iterative training framework, ensuring that the model refines its understanding of intricate patterns within the data for more precise and reliable decision-making.

By harnessing the capabilities of the optimized Decision Tree, the overarching objective is to attain an elevated level of accuracy and reliability in the subsequent stages of the project. This module holds paramount significance as a critical step in the evolution of the model. The synergistic integration of advanced optimization techniques and model training principles establishes the groundwork for robust and effective decision-making in the classification of dementia stages. The refined Decision Tree, a product of meticulous tuning, acts as a pivotal component in shaping the model's ability to discern intricate patterns within the data. This strategic approach ensures a more precise and reliable classification process, marking a pivotal advancement in the overall effectiveness of the model for dementia diagnosis.

Algorithm 2: GA-Decision Tree Model ()

Begin

ID3(Instances, Desired_feature, Attributes)
Input: Instances - Training examples
Desired_feature - Attribute for prediction
Attributes - List of attributes for evaluation
Output: Decision tree accurately classifying the provided examples.

1. Create a new tree node called root.
2. If all instances are positive:
   2.1 Set the root node label to '+'.
   2.2 Return the root.
3. If all instances are negative:
   3.1 Set the root node label to '-'.
   3.2 Return the root.
4. If Attributes is empty:
   4.1 Set the root node label to the most frequent value of Desired_feature in Instances.
   4.2 Return the root.
5. Otherwise:
   5.1 Select the best attribute (A) from Attributes that most effectively classifies the instances.
   5.2 Set the decision attribute for the root node as A.
   5.3 For each potential value (vi) of attribute A:
      5.3.1 Create a new tree branch beneath the root, associated with the condition A=vi.
      5.3.2 Consider Examples_vi as the subset of instances within Instances having the value vi for attribute.
      5.3.3 If Examples_vi is empty:
         5.3.3.1 Create a leaf node beneath the branch with the label set to the most prevalent value.
      5.3.4 Else:
         5.3.4.1 Add the subtree ID3(Examples_vi, Desired_feature, Attributes - {A}) beneath the branch.
6. Return the root.

End

2.6. Performance evaluation

In our research, a comprehensive evaluation and comparison are undertaken, focusing on two distinct models: the GA-Decision Tree model and the standalone Decision Tree model. This meticulous analysis aims to gauge the efficacy of these classification models in the context of dementia detection through MRI analysis. Key performance metrics, including precision, recall, F1-score, and accuracy, serve as fundamental benchmarks in this assessment. The detailed results of this comparative analysis are meticulously presented in table 2, offering nuanced insights into the individual and relative performance levels of both models. This thorough examination provides a robust foundation for understanding the strengths and limitations of each approach, contributing valuable knowledge to the ongoing discourse in the field of dementia diagnosmodel and the standalone Decision Tree model. To assess the efficacy of these classification models, we utilize fundamental performance metrics such as precision, recall, F1-score, and accuracy. The results of this comparative analysis are presented in Table 2, providing valuable insights into the respective performance levels of both models in the context of dementia detection through MRI analysis.

Table 2. Performance Comparison of GA-Decision Tree and Decision Tree Models

Model	Accuracy, %	Precision, %	Recall, %
Decision tree	86.66	50.00	42.85
GA-Decision tree	96.67	97.46	96.25

3. Results and discussion

Within the context of our investigation, we conducted a comparative assessment of two models: the standalone Decision Tree model and the GA-Decision Tree model. The Decision Tree model exhibited an accuracy rate of 86.66%, demonstrating its competence in dementia detection through MRI analysis. In contrast, the GA-Decision Tree model showcased a remarkable accuracy rate of 96.67%, signifying a significant performance enhancement. This substantial improvement in accuracy observed in the GA-Decision Tree model underscores the efficacy of employing Genetic Algorithms (GAs) for hyperparameter tuning within the decision tree framework. The GA-Decision Tree model, by optimizing hyperparameters, elevates its capacity to capture intricate patterns, resulting in more precise dementia prediction. The integration of the Genetic Algorithm (GA) not only refines hyperparameter settings but also significantly enhances the model's ability to accurately identify essential diagnostic patterns. This synergy between GA optimization and the Decision Tree architecture represents a substantial improvement in the model's overall performance, emphasizing its potential as a robust tool for nuanced dementia prediction through the meticulous extraction and utilization of crucial features within the MRI data.

3.1. Experimental results

In this section, we present the experimental findings of our study, placing specific emphasis on evaluating the performance of both the Decision Tree and GA-based Decision Tree models in the domain of dementia detection using MRI data. To ensure a comprehensive assessment of the models' effectiveness, we utilized Receiver Operating Characteristic - Area Under the Curve (ROC-AUC) analysis, a widely acknowledged metric for binary classification tasks. The ROC curve provides a visual representation of the true positive rate (sensitivity) plotted against the false positive rate (1-specificity) across various classification thresholds. This approach enables a nuanced exploration of the models' performance characteristics, offering valuable insights into their capacity for accurate dementia classification.

Figure 2 vividly displays the Receiver Operating Characteristic (ROC) curve corresponding to the Decision Tree model in our study. Calculating the Area Under the Curve (AUC) for this model yields a noteworthy value of 0.87, attesting to its commendable discriminative power. The ROC curve visually portrays the model's proficiency in effectively distinguishing between dementia and non-dementia cases, with higher AUC values indicative of a heightened level of diagnostic accuracy. This outcome underscores the robust performance of the Decision Tree model in delineating nuanced patterns within the MRI data and substantiates its potential as an effective tool for dementia detection.

Figure 3 depicts the ROC curve for the Decision Tree model optimized with Genetic Algorithm in our investigation. The noteworthy AUC of 0.97 emphasizes the model's strong discriminative ability. This high AUC underscores the proficiency of the GA-based Decision Tree model in accurately identifying dementia cases using MRI data.

Figure 4 illustrates the confusion matrix of the Decision Tree model when applied to testing data. This matrix provides valuable information about the model's accuracy in classification. The model demonstrates a higher proportion of correct predictions than incorrect ones. Specifically, it effectively identifies the presence of the condition and accurately recognizes its absence.

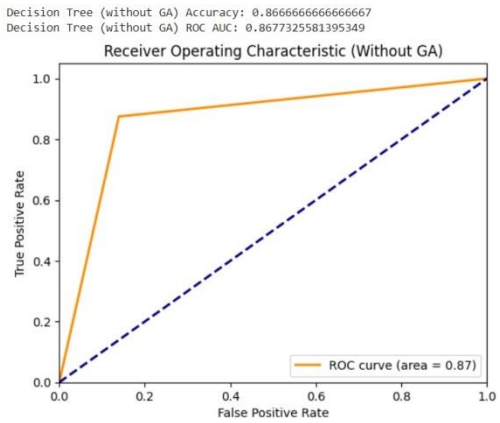


Fig. 2. ROC curve for the Decision Tree model

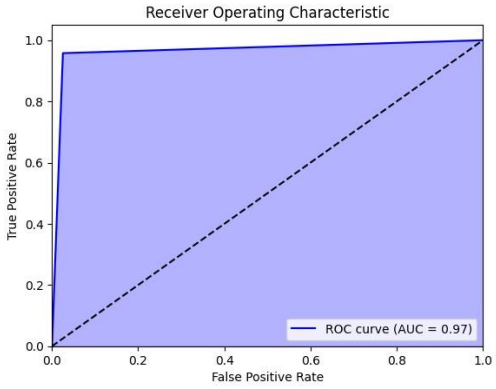


Fig. 3. ROC curve of GA-Decision tree model

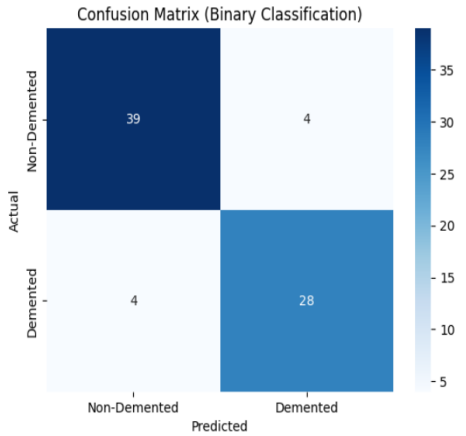


Fig. 4. Confusion Matrix of Decision tree model

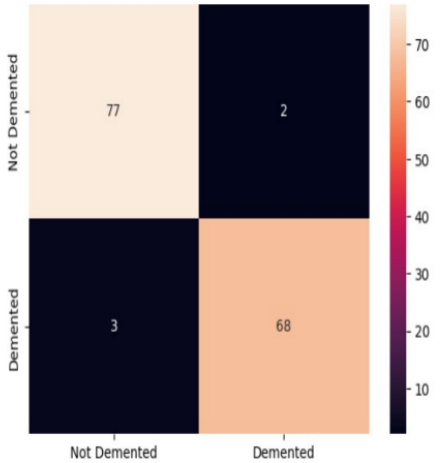


Fig. 5. Confusion Matrix of GA-Decision tree model

Figure 5 displays the confusion matrix for the GA-Decision Tree model, showcasing a significant presence of accurate positive and negative predictions. This suggests adeptness in accurately predicting both positive cases and non-diseased samples. Furthermore, the visualization emphasizes the equilibrium in addressing both facets of the classification task. Additionally, the matrix accentuates the skill in minimizing erroneous positive and negative predictions, enhancing its overall dependability in discerning between positive and healthy samples.

3.2. Discussion

Table 2 provides a comprehensive comparison between our proposed GA-Decision Tree model and previous work in the field. Dementia prediction through MRI analysis has been explored using various machine learning approaches. The table summarizes the performance of both our model and previously implemented methods, offering insights into the advancements achieved through the integration of Genetic Algorithms for hyperparameter tuning. This comparison stands as a useful reference for researchers aiming to assess their approaches in the current landscape of this domain.

Table 3. Comparative analysis of prior approaches and our proposed model

Methods	Accuracy, %	Specificity, %	Recall, %
Deep Neural Networks [5]	92.39	79	74
Neural networks, SVM, Random Forest [6]	68	77	71
Deep Learning [7]	92	-	91.67
KNN, Bayes Networks [8]	86.7	71	86
SVM, Random Forest, CNN, K-means [9]	85	68	84
Decision tree	81.08	90.0	42.85
GA-Decision tree	96.67	97.14	96.25

4. Conclusion

This research work presents a comprehensive comparative analysis between the Decision Tree and GA-Decision Tree models for the task of dementia detection using MRI data. The results demonstrate the superior performance of the GA-Decision Tree model, which achieved remarkable accuracy at 96.67%, precision at 97.46%, recall at 96.25%, and an impressive F1-score of 96.85%. These findings underscore the potency of Genetic Algorithms (GAs) in optimizing hyperparameters and elevating the overall predictive capabilities of the model. Notably, our experiments stress the critical role of parameter fine-tuning, encompassing learning rate and epochs, to unlock the full potential of the model. The GA-Decision Tree's ability to converge swiftly during training, courtesy of its adaptive learning rate mechanism, has significantly contributed to its superior predictive accuracy within a reduced number of iterations. This study reveals the promising prospects of incorporating GAs into the Decision Tree framework, marking a pivotal step toward advancing dementia detection through MRI analysis. The potential implications include early diagnosis and the capacity to revolutionize patient care, representing a significant advancement in imaging and diagnostic practices.

References

[1] Abdollahi J., Nouri-Moghaddam B.: Hybrid stacked ensemble combined with genetic algorithms for diabetes prediction. Iran Journal of Computer Science 5(3), 2022, 205–220.
[2] Adeli H., Ghosh-Dastidar S., Dadmehr N.: Alzheimer's disease and models of computation: Imaging, classification, and neural models. Journal of Alzheimer's Disease 7(3), 2005, 187–199.
[3] Al-Badarneh A., Najadat H., Alraziqi A.: Brain Images Classifier: A Hybrid Approach Using Decision Trees and Genetic Algorithms. JINT 7(2), 2016.
[4] Aminizadeh S. et al.: The applications of machine learning techniques in medical data processing based on distributed computing and the Internet of Things. Computer methods and programs in biomedicine 241, 2023, 107745.
[5] Angelillo M. T. et al.: Attentional pattern classification for automatic dementia detection. IEEE Access 7, 2019, 57706–57716.
[6] Azad R. et al.: Medical image segmentation on MRI images with missing modalities: a review [http://arxiv.org/abs/2203.06217].

[7] Bansal D. et al.: Comparative analysis of various machine learning algorithms for detecting dementia. Procedia computer science 132, 2018, 1497–1502.
[8] Basheer S., Bhatia S., Sakri S. B.: Computational modeling of dementia prediction using deep neural network: analysis on OASIS dataset. IEEE access 9, 2021, 42449–42462.
[9] Biswal A.: What Is Principal Component Analysis? Simplilearn.com [www.simplilearn.com/tutorials/machine-learning-tutorial/principal-component-analysis] (avaible 7.11.2023).
[10] Bukhari S. N. H., Webber J., Mehbodniya A.: Decision tree based ensemble machine learning model for the prediction of Zika virus T-cell epitopes as potential vaccine candidates. Scientific Reports 12(1), 2022, 7810.
[11] Deng W. et al.: An enhanced fast non-dominated solution sorting genetic algorithm for multi-objective problems. Information Sciences 585, 2022, 441–453.
[12] Dhiman G. et al.: A novel machine-learning-based hybrid CNN model for tumor identification in medical image processing. Sustainability 14(3), 2022, 1447.
[13] Díaz-Álvarez J. et al.: Genetic algorithms for optimized diagnosis of Alzheimer's disease and Frontotemporal dementia using Fluorodeoxyglucose positron emission tomography imaging. Frontiers in aging neuroscience 13, 2022, 983.
[14] Drouka A. et al.: Dietary and nutrient patterns and brain MRI biomarkers in dementia-free adults. Nutrients 14(11), 2022, 2345.
[15] Elhazmi A. et al.: Machine learning decision tree algorithm role for predicting mortality in critically ill adult COVID-19 patients admitted to the ICU. Journal of infection and public health 15(7), 2022, 826–834.
[16] Elyan E. et al.: Computer vision and machine learning for medical image analysis: recent advances, challenges, and way forward. Artificial Intelligence Surgery 2, 2022.
[17] Emam M. M., Houssein E. H., Ghoniem R. M.: A modified reptile search algorithm for global optimization and image segmentation: Case study brain MRI images. Computers in Biology and Medicine 152, 2023, 106404.
[18] Fang L., Wang X.: Brain tumor segmentation based on the dual-path network of multi-modal MRI images. Pattern Recognition 124, 2022, 108434.
[19] García-Gutierrez F. et al.: GA-MADRID: Design and validation of a machine learning tool for the diagnosis of Alzheimer's disease and frontotemporal dementia using genetic algorithms. Medical & Biological Engineering & Computing 60(9), 2022, 2737–2756.
[20] Gorji H. T., Haddadnia J.: A novel method for early diagnosis of Alzheimer's disease based on pseudo Zernike moment from structural MRI. Neuroscience 305, 2015, 361–371.
[21] Haug C. J., Drazen J. M.: Artificial intelligence and machine learning in clinical medicine. New England Journal of Medicine 388(13), 2023, 1201–1208.
[22] Javeed A. et al.: Machine Learning for Dementia Prediction: A Systematic Review and Future Research Directions. Journal of Medical Systems 47(1), 2023, 17.
[23] Leocadi M. et al.: Awareness impairment in Alzheimer's disease and fronto-temporal dementia: a systematic MRI review. Journal of Neurology 270(4), 2023, 1880–1907.
[24] Liang X. et al.: Evaluating voice-assistant commands for dementia detection. Computer Speech & Language 72, 2022, 101297.
[25] Li R. et al.: Applications of artificial intelligence to aid early detection of dementia: a scoping review on current capabilities and future directions. Journal of biomedical informatics 127, 2022, 104030.
[26] Liu H. et al.: NeuroCrossover: An intelligent genetic locus selection scheme for genetic algorithm using reinforcement learning. Applied Soft Computing 146, 2023, 110680.
[27] Miled Z. B. et al.: Feature engineering from medical notes: A case study of dementia detection. Heliyon 9(3), 2023.
[28] Mirheidari B. et al.: Dementia detection using automatic analysis of conversations. Computer Speech & Language 53, 2019, 65–79.
[29] Mirzaei G., Adeli H.: Machine learning techniques for diagnosis of Alzheimer disease, mild cognitive disorder, and other types of dementia. Biomedical Signal Processing and Control 72, 2022, 103293.
[30] Marcus D. S. et al.: Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults. Journal of Cognitive Neuroscience 22(12), 2010, 2677–2684.
[31] Nori V. S. et al.: Machine learning models to predict onset of dementia: a label learning approach. Alzheimer's & Dementia: Translational Research & Clinical Interventions 5, 2019, 918–925.
[32] Nowroozpoor A. et al.: Detecting cognitive impairment and dementia in the emergency department: a scoping review. Journal of the American Medical Directors Association 23(8), 2022, 1314–1315.
[33] Perovnik M. et al.: Automated differential diagnosis of dementia syndromes using FDG PET and machine learning. Frontiers in Aging Neuroscience 14, 2022, 1005731.
[34] Ramos D. et al.: Using decision tree to select forecasting algorithms in distinct electricity consumption context of an office building. Energy Reports 8, 2022, 417–422.
[35] Shehab M. et al.: Machine learning in medical applications: A review of state-of-the-art methods. Computers in Biology and Medicine 145, 2022, 105458.
[36] So A. et al.: Early diagnosis of dementia from clinical data by machine learning techniques. Applied Sciences 7(7), 2017, 651.
[37] Squires M. et al.: A novel genetic algorithm based system for the scheduling of medical treatments. Expert Systems with Applications 195, 2022, 116464.
[38] Varoquaux G. Cheplygina V.: Machine learning for medical imaging: methodological failures and recommendations for the future. NPJ Digital Medicine 5(1), 2022, 48.
[39] World Health Organization: WHO and World Health Organization: WHO. Dementia [www.who.int/news-room/fact-sheets/detail/dementia/?gclid=CjwKCAiA-P-rBhBEEiwAQEXhHzn09E8kOoRAOoS8xNltu1svCep3MzGBPB363AJ20n3XF3v9M9C9axoCS7QQAvD_BwE] (avaible 15.03.2023).
[40] Zhao X. et al.: A voice recognition-based digital cognitive screener for dementia detection in the community: Development and validation study. Frontiers in Psychiatry 13, 2022, 899729.

Prof. Govada Anuradha
e-mail: ganuradha@vrsiddhartha.ac.in

Dr. G. Anuradha, received her Ph.D. from Andhra University, Visakhapatnam, Andhra Pradesh. She is working as an Associate professor in Department of Computer Science and Engineering, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, Krishna district. She published 25 papers in national and international journals and more than 10 papers published in national and international conferences. She has guided several projects at M. Tech. level. Her research interests are artificial intelligence and machine learning.

https://orcid.org/0000-0002-0999-0376

Eng. Harini Davu
e-mail: davuharini@gmail.com

Harini Davu is currently pursuing B.Tech. degree in Computer Science and Engineering at Velagapudi Ramakrishna Siddhartha Engineering College. She is interested in doing projects in the machine learning domain. She contributed to the project "Crow detection in peanut field using Raspberry Pi".

https://orcid.org/0009-0008-6187-1797

Eng. Muthyalanaidu Karri
e-mail: karrimuthyalanaidu3802@gmail.com

Muthyalanaidu Karri is currently pursuing B.Tech. degree in Computer Science and Engineering at Velagapudi Ramakrishna Siddhartha Engineering College. He contributed to the project "Morning Start: An Android Application for Milk Delivery".

https://orcid.org/0009-0006-5850-3761

http://doi.org/10.35784/iapgos.5585 *received: 19.11.2023 | revised: 22.02.2024 | accepted: 05.03.2024 | available online: 31.03.2024*

MEDICAL FUZZY-EXPERT SYSTEM FOR PREDICTION OF ENGRAFTMENT DEGREE OF DENTAL IMPLANTS IN PATIENTS WITH CHRONIC LIVER DISEASE

Vitaliy Polishchuk[1], Sergii Pavlov[2], Sergii Polishchuk[1], Sergii Shuvalov[1], Andriy Dalishchuk[1], Natalia Sachaniuk-Kavets'ka[2], Kuralay Mukhsina[3], Abilkaiyr Nazerke[4]

[1]National Pirogov Memorial Medical University, Vinnytsia, Ukraine, [2]Vinnytsia National Technical University, Vinnytsia, Ukraine, [3]Institute of Information and Computing Technologies of the CS MES RK. Almaty, Kazkhstan, [4]Al-Farabi Kazakh National University, Almaty, Kazkhstan

Abstract. *The paper presents an information technology for assessing the degree of engraftment of dental implants in the event of a pathology violation through the use of fuzzy sets, which allows using this method for medical diagnostic tasks. Main scientific results: developed algorithms and mathematical models that formalize the process supporting diagnostic decisions based on fuzzy logic; developed mathematical models of membership functions that formalize the presentation of qualitative and qualitative informational features based on the rules of fuzzy logic, which can be used in information expert systems when assessing the degree of engraftment of dental implants in case of disease with pathological diseases.*

Keywords: medical expert systems, fuzzy logic, patient safety, dental implants, chronic liver pathology

MEDYCZNY ROZMYTY SYSTEM EKSPERCKI DO PRZEWIDYWANIA STOPNIA WSZCZEPIENIA IMPLANTÓW DENTYSTYCZNYCH U PACJENTÓW Z PRZEWLEKŁĄ CHOROBĄ WĄTROBY

Streszczenie. *W artykule przedstawiono technologię informacyjną do oceny stopnia wszczepienia implantów stomatologicznych u pacjentów z przewlekłą chorobą wątroby za pomocą zbiorów rozmytych, co pozwala na zastosowanie tej metody do medycznych zadań diagnostycznych. Główne wyniki naukowe: opracowano algorytmy i modele matematyczne formalizujące proces wspomagania podejmowania decyzji diagnostycznych w oparciu o logikę rozmytą; opracowano matematyczne modele funkcji przynależności formalizujące reprezentację ilościowych i jakościowych cech informacyjnych opartych na regułach logiki rozmytej, które mogą być wykorzystane w informatycznych systemach ekspertowych do oceny stopnia wszczepienia implantów stomatologicznych u pacjentów z przewlekłą chorobą wątroby.*

Słowa kluczowe: medyczny system ekspertowy, logika rozmyta, bezpieczeństwo pacjentów, implanty stomatologiczne, przewlekła choroba wątroby

Introduction

Recent years have shown an increase in the number of patients with tooth loss against the background of concomitant pathologies, which can be important when choosing treatment tactics. It is the loss of teeth that necessitates the search for new methods and means of treatment and restoration of the continuity of the dental row. At the moment, in the practice of dentists, more and more are inclined to restore defects of the dentition with the help of dental implants [1–3]. Dental implants can be used with complete and partial edentia, which is important for practical dentistry. When carrying out dental implantation, you need to pay attention to risk factors that can negatively affect the osseointegration of dental implants. There are many multifactorial factors that affect dental regeneration. Among them, we can single out cigarette smoking, poor oral hygiene, concomitant diseases of the liver, pancreas, and osteoporosis. All this creates conditions that negatively affect the implantation of dental implants. When dental implantation is performed, a dental analysis of the functioning of the body is necessary, because any risk factor can affect the progress of dental implant implantation due to the quality of the bone system of the jaw bones. The mineral density of the bone system of the jaw bones directly affects the success of the dental implant surgery and is directly related to the remodeling of the bone system [4, 6, 13]. The very mechanism of bone tissue remodeling takes place in several interrelated phases. At the same time, the processes of osteogenesis and osteoresorption take place. And depending on the state of the body, the first or second process prevails, which affects the occurrence of complications. It should be noted that these two processes can be both physiological and pathological, both local and general. Implantation of dental implants may depend on the conditions under which surgical intervention is performed. Bleeding, pain during surgery and in the postoperative period, inflammation in the area of the surgical wound, separation of sutures often occur [1, 4, 15]. All this can be eliminated in the presence of information about the general state of the body's functioning, which we learn from a detailed history collection and examination of the patient in the pre-operative period [6, 13, 16].

To evaluate and predict the degree of engraftment of dental implants in patients with chronic liver pathology, the authors propose the use of medical expert systems based on fuzzy sets. At the same time, the task of formalizing medical data by creating formalized and standardized outpatient charts, medical histories, and medical knowledge bases is relevant. Therefore, the relevance of the work is the development of expert medical systems, which will make it possible to accelerate the collection and analysis of medical information. In addition to the direct formalization of medical data, it is promising to evaluate their informativeness and develop mathematical methods and models for the synthesis of computer diagnosis [7, 11, 17].

1. Method

All our actions in the preoperative period should be aimed at timely identification and elimination of risk factors and the occurrence of complications during dental implantation. Timely correction of factors that contribute to the occurrence of complications and rejection of dental complications will preserve the functioning of the dental implant for many years, especially if there is a background accompanying pathology, in particular, the hepatobiliary system. After all, the hepatobiliary system is a system that has a significant impact on osteoregenerative and osteoplastic processes, taking into account the function of the liver and bile. Thanks to the function of the liver, the exchange of proteins, fats, carbohydrates, as well as vitamins and hormones takes place [2, 12]. Along with the metabolism, the metabolic and detoxification function of the liver should be noted. The gallbladder and the formation of bile, which in the body allows the gastrointestinal tract to function properly, are inextricably linked to the function of the liver. Bile takes part in the emulsification of fats and their assimilation, the absorption of proteins and carbohydrates, trace elements, including calcium in the distal parts of the intestines.

The mineral density of bone tissue can be analyzed based on densitometry indicators, as well as biochemical indicators. When analyzing densitometry indicators, it is advisable to measure the mineral density of spongy and compact matter. Alkaline phosphatase, a marker of bone matrix formation, and osteocalcin can be distinguished from biochemical indicators that most successfully characterize metabolic changes in bone tissue. The search of many researchers draws attention to the need for the most reliable prognosis of the implantation of dental implants depending on many risk factors [12, 15, 17].

The purpose of the study is to develop a mathematical model for predicting the implantation of dental implants in patients with chronic liver disease.

2. Peculiarities of using medical information systems

In medical information systems, data and knowledge processing is reduced to three main stages [13,15,16]:

1. Elements of information are placed in certain sections, which have the form of parameters and diagnoses [7, 14, 18].
2. Databases of collected data and theoretical knowledge are organized - their structure is formed, the order of information placement and the nature of the relationship between information elements are determined [19, 20, 25].
3. The most necessary information is selected, a decision is made, the knowledge base and database are edited.

In practice, both approaches are used at MIT, because during research, the obtained biomedical data are quite closely correlated and, thus, the final result of data processing is used for analysis, selection of treatment and rehabilitation methods, and prediction of long-term results. [8, 9, 21] .

When implementing information expert systems, the analysis of biomedical indicators is the basis for making a final diagnostic conclusion.

When creating an expert medical system, it is necessary to solve the following main problems: [18, 19, 24]:

- analysis and area of use of the system,
- synthesis regarding the construction of a logical scheme,
- formation and interpretation of nosological forms, which must be analyzed and based on statistical information containing the classification of symptoms, as well as the peculiarities of the state of a certain body system,
- recommendations regarding the optimality of biomedical information analysis technology,
- implementation of algorithmic software for assessing the level of pathology and determining recommendations for diagnostic and prognostic conclusions.

MIS design will be of high quality only if the study is conducted by an experienced diagnostician. Such research can be designed by a group of qualified experts in this field of diagnostics [23, 26].

3. Materials and methods

53 patients who had dental implants installed were examined in the dental clinic. The age of the patients ranged from 22 to 45 years. All patients underwent densitometric (research of the mineral density of the compact and spongy substance of the jaws), biochemical (alkaline phosphatase, bilirubin) and oral hygiene (Fedorov-Volodkin index). All patients were divided into 2 groups: main (without liver pathology) and comparison (with chronic liver disease). The expert database for determining the degree of engraftment of dental implants in patients with chronic liver pathology is presented in the form of a table 1.

Table 1. Expert database for determining the degree of engraftment of dental implants in patients with chronic liver pathology

Degree of engraftment	Mineral density of compact substance (HU)					Mineral density of spongy substance (HU)					Total hemoglobin level (μmol/l)					Alkaline phosphatase (units/l)					Fedorov-Volodkin Index				
	1250 ÷ 3000	850 ÷ 1249	350 ÷ 850	150 ÷ 350	0 ÷ 150	850 ÷ 3000	450 ÷ 850	250 ÷ 450	150 ÷ 250	0 ÷ 150	0 ÷ 10	10.1 ÷ 20.0	20.1 ÷ 30.0	30.1 ÷ 40.0	40.1 ÷ 100	35 ÷ 79	80 ÷ 119	120 ÷ 159	160 ÷ 199	200 ÷ 500	1.1 ÷ 1.5	1.6 ÷ 2.0	2.1 ÷ 2.5	2.6 ÷ 3.4	3.5 ÷ 5.0
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
High	+					+					+					+					+				
Sufficient		+					+					+					+					+			
Moderate			+					+					+					+					+		
Relative				+					+					+					+					+	
Low					+					+					+					+					+
One-way ANOVA & LSD test	p1-2=0.002																								

Table 2. Features of quantitative assessment for determining the degree of engraftment of dental implants in patients with chronic liver pathology

Factors / Degree of engraftment	Mineral density of compact substance (HU), (X1)	Mineral density of spongy substance (HU), (X2)	Total hemoglobin level (μmol/l), (X3)	Alkaline phosphatase (units/l), (X4)	Fedorov-Volodkin Index, (X5)
High $\mu^{I}(x_1x_2x_3x_4x_5)$	1250 ÷ 3000	850 ÷ 3000	0 ÷ 10	35 ÷ 79	1.1 ÷ 1.5
Sufficient $\mu^{II}(x_1x_2x_3x_4x_5)$	850 ÷ 1249	450 ÷ 850	10.1 ÷ 20.0	80 ÷ 119	1.6 ÷ 2.0
Moderate $\mu^{III}(x_1x_2x_3x_4x_5)$	350 ÷ 850	250 ÷ 450	20.1 ÷ 30.0	120 ÷ 159	2.1 ÷ 2.5
Relative $\mu^{IV}(x_1x_2x_3x_4x_5)$	150 ÷ 350	150 ÷ 250	30.1 ÷ 40.0	160 ÷ 199	2.6 ÷ 3.4
Low $\mu^{V}(x_1x_2x_3x_4x_5)$	0 ÷ 150	0 ÷ 150	40.1 ÷ 100	200 ÷ 500	3.5 ÷ 5.0
min/max	0 ÷ 3000	0 ÷ 3000	0 ÷ 100	35 ÷ 500	1.1 ÷ 5.0

Table 3. Knowledge base for fuzzy for diagnostics determining the degree of engraftment of dental implants in patients with chronic liver pathology

Clinical forms of the degree of engraftment	(X1)	(X2)	(X3)	(X4)	(X5)
$\mu^{I}(x_1x_2x_3x_4x_5)$	A	A	L	L	L
	A	HA	L	L	L
	A	H	L	L	L
	HA	HA	L	L	L
	HA	H	L	L	L
	H	H	L	L	L
$\mu^{II}(x_1x_2x_3x_4x_5)$	LA	LA	L	LA	LA
	A	LA	LA	LA	LA
$\mu^{III}(x_1x_2x_3x_4x_5)$	LA	L	LA	LA	LA
	LA	L	A	LA	LA
	LA	LA	LA	LA	LA
	LA	LA	A	LA	LA
$\mu^{IV}(x_1x_2x_3x_4x_5)$	L	L	LA	LA	A
	L	L	A	LA	A
$\mu^{V}(x_1x_2x_3x_4x_5)$	L	L	A	A	HA
	L	L	HA	HA	H
	L	L	HA	H	HA
	L	L	H	HA	HA
	L	L	H	H	H

According to the methodology for building decisive rules of the decision support subsystem based on fuzzy logic, we break down the range of changes of each of the informative parameters: densitometric (research of the mineral density of the compact and spongy substance of the jaws), biochemical (alkaline phosphatase, bilirubin) and oral hygiene (the index Fedorov-Volodkin) by 5 degrees of engraftment, corresponding to the qualitative fuzzy terms low (L), low average (LA), average (A), high average (HA), high (H) (table 2 and 3).

When setting the functions of belonging of informative features X1 ÷ X5 to fuzzy terms on the interval [0;1], the well-known function [17] is used, the graph of which is shown in Fig. 1.

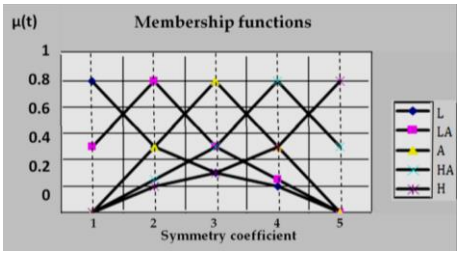


Fig. 1. Membership functions [17, 21, 29]

Based on the knowledge base developed by experts, presented in table 2 and table 3, and the above-mentioned membership functions of fuzzy terms [10, 17], models of decision-making support rules were developed for assessing the degree of engraftment of dental implants in patients with chronic liver pathology.

For degree of engraftment (High) $\mu^{I}(x_1x_2x_3x_4x_5)$

$$\mu^{d1}(X_1,X_2,X_3,X_4.X_5)=\mu^{A}(X_1)\cdot\mu^{A}(X_2)\cdot\mu^{L}(X_3)\cdot\mu^{L}(X_4)\cdot\mu^{L}(X_5)\cup$$
$$\mu^{A}(X_1)\cdot\mu^{HA}(X_2)\cdot\mu^{L}(X_3)\cdot\mu^{L}(X_4)\cdot\mu^{L}(X_5)$$
$$\cup\mu^{A}(X_1)\cdot\mu^{H}(X_2)\cdot\mu^{L}(X_3)\cdot\mu^{L}(X_4)\cdot\mu^{L}(X_5)\cup$$
$$\mu^{HA}(X_1)\cdot\mu^{HA}(X_2)\cdot\mu^{L}(X_3)\cdot\mu^{L}(X_4)\cdot\mu^{L}(X_5)\cup$$
$$\mu^{HA}(X_1)\cdot\mu^{H}(X_2)\cdot\mu^{L}(_3)\cdot\mu^{L}(X_4)\cdot\mu^{L}(X_5)\cup$$
$$\mu^{H}(X_1)\cdot\mu^{H}(X_2)\cdot\mu^{L}(X_3)\cdot\mu^{L}(X_4)\cdot\mu^{L}(X_5)$$

For degree of engraftment (Sufficient) $\mu^{II}(x_1x_2x_3x_4x_5)$

$$\mu^{d2}(X_1,X_2,X_3,X_4.X_5)=\mu^{LA}(X_1)\cdot\mu^{LA}(X_2)\cdot\mu^{L}(X_3)\cdot\mu^{LA}(X_4)\cdot\mu^{LA}(X_5)\cup$$
$$\mu^{A}(X_1)\cdot\mu^{LA}(X_2)\cdot\mu^{LA}(X_3)\cdot\mu^{LA}(X_4)\cdot\mu^{LA}(X_5)$$

For degree of engraftment (Moderate) $\mu^{III}(x_1x_2x_3x_4x_5)$

$$\mu^{d3}(X_1,X_2,X_3,X_4.X_5)=\mu^{LA}(X_1)\cdot\mu^{L}(X_2)\cdot\mu^{LA}(X_3)\cdot\mu^{LA}(X_4)\cdot\mu^{LA}(X_5)\cup$$
$$\mu^{LA}(X_1)\cdot\mu^{L}(X_2)\cdot\mu^{A}(X_3)\cdot\mu^{LA}(X_4)\cdot\mu^{LA}(X_5)$$
$$\cup\mu^{LA}(X_1)\cdot\mu^{LA}(X_2)\cdot\mu^{LA}(X_3)\cdot\mu^{LA}(X_4)\cdot\mu^{LA}(X_5)\cup$$
$$\mu^{LA}(X_1)\cdot\mu^{LA}(X_2)\cdot\mu^{A}(X_3)\cdot\mu^{LA}(X_4)\cdot\mu^{LA}(X_5)$$

For degree of engraftment (Relative) $\mu^{IV}(x_1x_2x_3x_4x_5)$

$$\mu^{d4}(X_1,X_2,X_3,X_4.X_5)=\mu^{L}(X_1)\cdot\mu^{L}(X_2)\cdot\mu^{LA}(X_3)\cdot\mu^{LA}(X_4)\cdot\mu^{A}(X_5)$$
$$\cup\mu^{L}(X_1)\cdot\mu^{L}(X_2)\cdot\mu^{A}(X_3)\cdot\mu^{LA}(X_4)\cdot\mu^{A}(X_5)$$

For degree of engraftment (Low) $\mu^{V}(x_1x_2x_3x_4x_5)$

$$\mu^{d5}(X_1,X_2,X_3,X_4.X_5)=\mu^{L}(X_1)\cdot\mu^{L}(X_2)\cdot\mu^{A}(X_3)\cdot\mu^{A}(X_4)\cdot\mu^{LA}(X_5)\cup$$
$$\mu^{L}(X_1)\cdot\mu^{L}(X_2)\cdot\mu^{HA}(X_3)\cdot\mu^{HA}(X_4)\cdot\mu^{H}(X_5)$$
$$\cup\mu^{L}(X_1)\cdot\mu^{L}(X_2)\cdot\mu^{HA}(X_3)\cdot\mu^{H}(X_4)\cdot\mu^{HA}(X_5)\cup$$
$$\mu^{L}(X_1)\cdot\mu^{L}(X_2)\cdot\mu^{H}(X_3)\cdot\mu^{HA}(X_4)\cdot\mu^{HA}(X_5)\cup$$
$$\mu^{L}(X_1)\cdot\mu^{L}(X_2)\cdot\mu^{H}(_3)\cdot\mu^{H}(X_4)\cdot\mu^{H}(X_5)$$

For the formalization of the indices, membership functions are given that correspond to the rules of fuzzy sets [17, 27, 28].

Therefore, logical equations for evaluating the degree of engraftment of dental implants in patients with chronic liver pathology will have the following form for factors (X1 – X5).

For factors X1:

$$\tilde{\mu}^{L}(X_1)=\begin{cases}1.0-0.00067x_1, x_1\in[0;750)\\0.667-0.0002x_1, x_1\in[750;3000]\end{cases}$$

$$\tilde{\mu}^{LA}(X_1)=\begin{cases}0.00067x_1+0.5, x_1\in[0;750)\\1.5-0.00067x_1, x_1\in[750;1500)\\1.0-0.0003x_1, x_1\in[1500;3000]\end{cases}$$

$$\tilde{\mu}^{A}(X_1)=\begin{cases}0.00067x_1, x_1\in[0;1500)\\2-0.00067x_1, x_1\in[1500;3000]\end{cases}$$

$$\tilde{\mu}^{HA}(X_1)=\begin{cases}0.0003x_1, x_1\in[0;1500)\\0.00067x_1-0.5, x_1\in[1500;2250)\\2.5-0.00067x_1, x_1\in[2250;3000]\end{cases}$$

$$\tilde{\mu}^{H}(X_1)=\begin{cases}0.0002x_1, x_1\in[0;2250)\\0.00067x_1-1, x_1\in[2250;3000]\end{cases}$$

For factors X2:

$$\tilde{\mu}^{L}(X_2)=\begin{cases}1.0-0.00067x_2, x_2\in[0;750)\\0.667-0.0002x_2, x_2\in[750;3000]\end{cases}$$

$$\tilde{\mu}^{LA}(X_2)=\begin{cases}0.00067x_2+0.5, x_2\in[0;750)\\1.5-0.00067x_2, x_2\in[750;1500)\\1.0-0.0003x_2, x_2\in[1500;3000]\end{cases}$$

$$\tilde{\mu}^{A}(X_2)=\begin{cases}0.00067x_2, x_2\in[0;1500)\\2-0.00067x_2, x_2\in[1500;3000]\end{cases}$$

$$\tilde{\mu}^{HA}(X_2)=\begin{cases}0.0003x_2, x_2\in[0;1500)\\0.00067x_2-0.5, x_2\in[1500;2250)\\2.5-0.00067x_2, x_2\in[2250;3000]\end{cases}$$

$$\tilde{\mu}^{H}(X_2)=\begin{cases}0.0002x_2, x_2\in[0;2250)\\0.00067x_2-1, x_2\in[2250;3000]\end{cases}$$

For factors X3:

$$\tilde{\mu}^{L}(X_3)=\begin{cases}1.0-0.02x_3, x_3\in[0;25)\\0.67-0.0067x_3, x_3\in[25;100]\end{cases}$$

$$\tilde{\mu}^{LA}(X_3)=\begin{cases}0.02x_3+0.5, x_3\in[0;25)\\1.5-0.02x_3, x_3\in[25;50)\\1.0-0.01x_3, x_3\in[50;100]\end{cases}$$

$$\tilde{\mu}^{A}(X_3)=\begin{cases}0.02x_3, x_3\in[0;50]\\2-0.02x_3, x_3\in[50;100]\end{cases}$$

$$\tilde{\mu}^{HA}(X_3)=\begin{cases}0.01x_3, x_3\in[0;50)\\0.02x_3-0.5, x_3\in[50;75)\\2.5-0.02x_3, x_3\in[75;100]\end{cases}$$

$$\tilde{\mu}^{H}(X_3)=\begin{cases}0.0067x_3, x_3\in[0;75)\\0.02x_3-1, x_3\in[75;100]\end{cases}$$

For factors X4:

$$\tilde{\mu}^{L}(X_4)=\begin{cases}1.15-0.0043x_4, x_4\in[35;151.25)\\0.72-0.014x_4, x_4\in[151.25;500]\end{cases}$$

$$\tilde{\mu}^{LA}(X_4)=\begin{cases}0.004x_4+0.349, x_4\in[35;151.25)\\1.65-0.0043x_4, x_4\in[151.25;267.5)\\1.07-0.002x_4, x_4\in[267.5;500]\end{cases}$$

$$\tilde{\mu}^{A}(X_4)=\begin{cases}0.0043x_4-0.15, x_4\in[35;267.5]\\2.15-0.0043x_4, x_4\in[267.5;500]\end{cases}$$

$$\tilde{\mu}^{HA}(X_4)=\begin{cases}0.002x_4-0.075, x_4\in[35;267.5)\\0.0043x_4-0.65, x_4\in[267.5;383.75)\\2.65-0.0043x_4, x_4\in[383.575;500]\end{cases}$$

$$\tilde{\mu}^{H}(X_4)=\begin{cases}0.001x_4-0.05, x_4\in[35;383.75)\\0.0043x_4-1.15, x_4\in[383.75;500]\end{cases}$$

For factors X5:

$$\tilde{\mu}^{L}(X_5)=\begin{cases}1.57-0.51x_5, x_5\in[1.1;2.07)\\0.85-0.17x_5, x_5\in[2.07;5.0]\end{cases}$$

$$\tilde{\mu}^{LA}(X_5)=\begin{cases}0.52x_5+0.067, x_5\in[1.1;2.07)\\2.06-0.51x_5, x_5\in[2.07;3.05)\\1.67-0.33x_5, x_5\in[3.05;5.0]\end{cases}$$

$$\tilde{\mu}^{A}(X_5)=\begin{cases}0.51x_5-0.56, x_5\in[1.1;3.05]\\3.3-0.67x_5, x_5\in[3.05;5.0]\end{cases}$$

$$\tilde{\mu}^{HA}(X_5)=\begin{cases}0.25x_5-0.28, x_5\in[1.1;3.05)\\0.51x_5-1.06, x_5\in[3.05;4.03)\\3.08-0.52x_5, x_5\in[4.03;5.0]\end{cases}$$

$$\tilde{\mu}^{H}(X_5)=\begin{cases}0.17x_5-0.19, x_5\in[1.1;4.03)\\0.52x_5-1.58, x_5\in[4.03;5.0]\end{cases}$$

Transition from the function $\tilde{\mu}^j(u)$ to required functions $\mu^j(x_i)$ is performed in the following way [17, 30, 31].

$$u_i = 4\frac{x_n - \underline{x_n}}{\overline{x_n} - \underline{x_n}}, \quad \tilde{\mu}^j(u_n) = \mu^j(x_n)$$

In the process of assessing the degree of engraftment of dental implants in patients with chronic disease pathology, the task of developing and configuring a neurofuzzy network becomes. To develop the configuration of the parameters of this network, recurrent components proposed by Professor O. V. Mr. Rothstein [15, 31]. The task of adjustment is carried out in the selection of such factors of membership functions $(b_i^{jp}(t), c_i^{jp}(t))$ and weights of fuzzy rules $w_{jp}(t)$, which ensure the minimum discrepancy between the models and the evaluation results.

$$\sum_{i=1}^{M}(F_y(\tilde{x}_1^l, \tilde{x}_2^l, ...\tilde{x}_{12}^l, W_i) - \tilde{y}_l)^2 = \min_{W_i}$$

where $\langle \tilde{X}_l, \tilde{y}_l \rangle, l = \overline{1, M}$ experimental research data; b is the maximum coordinate; c is the factor of compression and extension (Fig. 2).

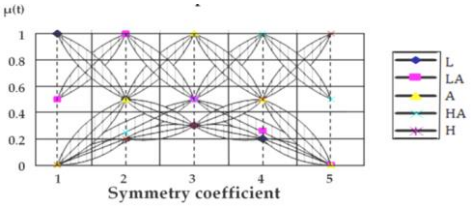


Fig. 2. Function setting procedure

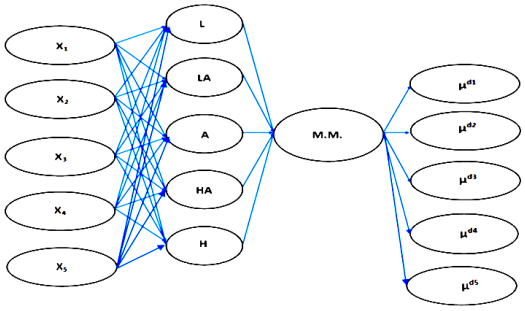


Fig. 3. Medical expert system for determining the degree of engraftment of dental implants in patients with chronic liver pathology

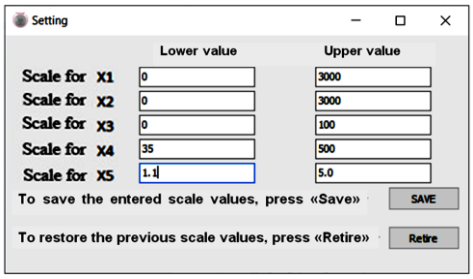


Fig. 4. Function of entering minimum and maximum values for each factor

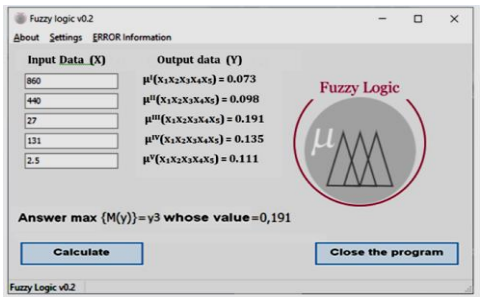


Fig. 5. Medical expert system for determining for determining the degree of engraftment of dental implants in patients with chronic liver pathology

The implementation scheme of the expert medical system in the form of a neural network for assessing the degree of implantation of dental implants in patients is shown in Fig. 3.

An interface was developed for the doctor, which allows to process the results of studies in a convenient way to assess the degree of implantation of dental implants in patients with chronic liver pathology [1, 14, 22].

The result of implementing a user interface is a software system that works like this (Fig. 4 and 5).

It should be noted that the reliability of supporting the correct decision-making of the medical expert system for determining the degree of engraftment of dental implants in patients with chronic liver pathology based on fuzzy sets for expert assessment was 95%.

4. Conclusions

The paper analyzes the main areas of application of mathematical methods in medical diagnostics, formulates the principles of diagnostics based on fuzzy logic. The basic structure of the MIS medical information system for assessing the degree of engraftment of dental implants in patients with chronic liver pathology was developed and the main recommendations for its design were put forward, namely: the selection and purpose of the system; selection of the structural scheme of the system; formation and analysis of the list of nosological forms, collection of statistically reliable information about the severity of symptoms; choosing a method of processing biomedical information; construction of an algorithm for solving the problems of evaluating biomedical information and forming diagnostic and prognostic conclusions.

Hardware and software were developed for the implementation of a convenient interface, which made it possible to formalize quantitative indicators in the form of informational signs when solving the problems of medical diagnosis of assessing the degree of engraftment of dental implants in patients with chronic liver pathology.

The results of the research and the approval of the expert medical system demonstrate the high reliability of the obtained results when assessing the degree of engraftment of dental implants in patients with chronic liver pathology.

The practical value of the work lies in the possibility of using an automated medical expert system to solve the problems of medical diagnosis based on fuzzy logic when assessing the degree of engraftment of dental implants in patients with chronic liver pathology.

Acknowledgments

This research has been is funded by the National Research Fund of Ukraine (Grant No. 2022.01/013).

We would like to thank Sergii Veremienko for reviewing the English language in this manuscript.

References

[1] Abdelhay N., Prasad S., Gibson M. P.: Guided versus non-guided dental implant placement: a systematic review and meta-analysis. BDJ Open. 7(1), 2021, 31.
[2] Arunyanak S. P. et al.: The effect of factors related to periodontal status toward peri-implantitis. Clin Oral Implants Res. 30(8), 2019, 791–799.
[3] Atieh M. A. et al.: Interventions for replacing missing teeth: alveolar ridge preservation techniques for dental implant site development. Cochrane Database of Systematic Reviews 4, 2021, CD010176.
[4] Bertolini M. M. et al.: Does traumatic occlusal forces lead to peri-implant bone loss? A systematic review. Braz Oral Res. 33(suppl 1), 2019, e069.
[5] Clinical guidelines for the management of pulpal diseases, approved by Decree 15 of the Council of Public Association: Russian Dental Association. 2018.
[6] Demkovich A. E., Yakymchuk M. M., Sverstyuk A. S.: Etiological risk factors for the occurrence of peri-implantitis. Clinical dentistry 2(31), 2020, 62–69.
[7] Guo Y. et al.: Influence of marginal bone resorption on two mini implant-retained mandibular overdenture: An in vitro study. J Adv Prosthodont. 13(1), 2021, 55–64.
[8] Katelyan O. V. et al.: Study of the peripheral blood circulation of an abdominal wall using optoelectronic plethysmograph. W.Wojcik et al. (eds): Information Technology in Medical Diagnostics II. CRC Press, Balkema book, Taylor & Francis Group, London, UK, 2019, 119–125.
[9] Kozlovska T. I. et al.: Device to determine the level of peripheral blood circulation and saturation. Proc. SPIE 10031, 2016, 100312Z.
[10] Nizhynska-Astapenko Z. et al.: Information medical fuzzy-expert systemfor the assessment of the diabetic ketoacidosis severity on the base of the blood gases indices. Proc. SPIE 12126, 2021.
[11] Pavlov S. V. et al.: Analysis of microcirculatory disorders in inflammatory processes in the maxillofacial region on based of optoelectronic methods. Przeglad Elektrotechniczny 93(5), 2017, 114–117.

[12] Pavlov S. V. et al.: Electro-optical system for the automated selection of dental implants according to their colour matching. Przeglad Elektrotechniczny 93(3), 2017, 121–124.
[13] Pelekhan B. L., Rozhko M. M.: Bone tissue resorption around intraosseous dental implants in patients with mandible edentulousness. Stomatological Bulletin 121(4), 2023, 55–62.
[14] Pelekhan B. et al.: Analytical Modeling of the Interaction of a Four Implant-Supported Overdenture with Bone Tissue. Materials 15(7), 2022, 2398.
[15] Polishchuk S. S., Skyba V. Ya., Davydenko I. S.: Histological changes of bone tissue in the perforation defect site of the rat mandibule when using hepatoprotector in odstructive hepatitis. World of medicine and biology 16(72), 2020, 193–198.
[16] Polishchuk V. S., Polishchuk S. S.: Peculiarities of the course after the operative course of patients after dental implantation on the background of the pathology of the hepatobilier system. Stomatological Bulletin 120(3), 2022, 51–56.
[17] Rotshtein A.: Design and Tuning of Fussy IF –THEN Vuly for Medical Didicol Diagnosis. H.-N.L Teodorescu, et al. (eds): Fuzzy and Neuro-Fuzzy Systems in Medicine. CRC-Press, 1998, 235–295.
[18] Sidor O. V.: The strategy for planning surgical stage dental implantation. Stomatological Bulletin 118(1), 2022, 50–55.
[19] Semenov Ye. I. et al.: Comparative characteristics of dental defects and the volume of implantological care in the young population of Ukraine. Stomatological Bulletin 119(2), 2022, 60–65.
[20] Serkova V. K. et al.: Medical expert system for assessment of coronary heart disease destabilization based on the analysis of the level of soluble vascular adhesion molecules. Proc. SPIE 10445, 2017, 104453O.
[21] Shkilniak L. et al.: Expert fuzzy systems for evaluation of intensity of reactive edema of soft tissues in patients with diabetes. Informtyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska – IAPGOS, 2022, 3, 59–63.
[22] Taubayev G. et al.: Machine learning algorithms and classificationof textures. Journal of Theoretical and Applied Information Technologythis 98(23), 2020, 3854–3866.
[23] Ushenko Yu. A., Sidor M. I., Bodnar G. B.: Mueller-matrix mapping of optically anisotropic fluorophores of biological tissues in the diagnosis of cancer. Quanrum Electron. 44(8), 2014, 785–790.
[24] Ushenko V. A., Gavrylyak M. S.: Azimuthally invariant Mueller-matrix mapping of biological tissue in differential diagnosis of mechanisms protein molecules networks anisotropy. Proc. SPIE 8812, 2016, 88120Y.
[25] Vasilevskyi O. et al.: Method of evaluating thelevel of confidence basedon metrological risks for determining the coverage factorin the concept of uncertainty. Proc. SPIE. 10808, 2018, 108082C.
[26] Vassilenko V. et al.: Automated features analysis of patients with spinal diseases using medical thermal images. Proc. SPIE 11456, 2020, 114560L.
[27] Wójcik W. et al. (eds): Information Technology in Medical Diagnostics. CRC Press, 2017.
[28] Wójcik W. et al. (eds): Information Technology in Medical Diagnostics II. Taylor & Francis Group. CRC Press, Balkema Book, London, 2019.
[29] Wójcik W. et al.: Medical Fuzzy-Expert System for Assessment of the Degree of Anatomical Lesion of Coronary Arteries. Int. J. Environ. Res. Public Health 20(2), 2023, 979 [https://doi.org/10.3390/ijerph20020979].
[30] Zabolotna N. I. et al.: Diagnostics of pathologically changed birefringent networks by means of phase Mueller matrix tomography. Proc. SPIE 8698, 2013, 86980C.
[31] Zabolotna N. I. et al.: System of polarization phasometry of polycrystalline blood plasma networks in mammary gland pathology diagnostics. Proc. SPIE 9613, 2015, 961311.

M.Sc. Vitaliy Polishchuk
e-mail: polischuk_vitaliy@ukr.net

Assistant of the Department of Surgical Dentistry and Maxillofacial Surgery, National Pirogov Memorial Medical University, Vinnytsia, Ukraine.
Scientific direction: maxillofacial surgery, medical diagnostics, engraftment of dental implants in patients with chronic liver pathology, therapeutic effect, regeneration.

https://orcid.org/0000-0001-7180-3650

M.S. Andriy Dalishchuk
e-mail: andriy.dalishchuk@gmail.com

Assistant of the Department of Surgical Dentistry and Maxillofacial Surgery, National Pirogov Memorial Medical University, Vinnytsia, Ukraine.
Scientific direction: maxillofacial surgery, medical diagnostics, engraftment of dental implants in patients with chronic liver pathology, therapeutic effect, regeneration.

https://orcid.org/0000-0002-5090-6172

Prof. Sergii Pavlov
e-mail: psv@vntu.edu.ua

Academician of International Applied Radioelectronic Science Academy, Professor of Biomedical Engineering and Optic-Electronic Systems Department, Vinnytsia National Technical University
Scientific direction: biomedical information optoelectronic and laser technologies for diagnostics and physiotherapy influence.
Deals with issues of improving the distribution of optical radiation theory in biological objects, particularly through the use of electro-optical systems, and the development of intelligent biomedical optoelectronic diagnostic systems and standardized methods for reliably determining the main hemo-dynamic cardiovascular system of comprehensive into account scattering effects.

https://orcid.org/0000-0002-0051-5560

Ph.D. Eng. Natalia Sachaniuk-Kavets'ka
e-mail: skn1901@gmail.com

In 2003 she received a Candidate of Sciences degree (Ph.D.) at the State Research Institute of Information Infrastructure of the State Committee for Communications and Informatization of Ukraine and the National Academy of Sciences of Ukraine (Lviv) with a specialty in automated control systems and advanced information technologies. Currently, she is an associate professor at the Department of Higher Mathematics of the Vinnytsia National Technical University. Scientific research includes issues related to mathematical modeling of information processing in a logical-temporal environment, as well as the possibility of using mathematical models of image processing for information protection.

https://orcid.org/0000-0001-6405-1331

Prof. Sergii Polishchuk
e-mail: polischuk07@ukr.net

Professor of the Department of Surgical Dentistry and Maxillofacial Surgery, National Pirogov Memorial Medical University, Vinnytsia, Ukraine.
Scientific direction: maxillofacial surgery, medical diagnostics, engraftment of dental implants in patients with chronic liver pathology, therapeutic effect, regeneration.

https://orcid.org/0000-0002-8635-9932

Ph.D. Kuralay Mukhsina
e-mail: kuka_ai@mail.ru

Ph.D., senior researcher of the Institute of Information and Computing Technologies of the CS MES RK. Graduated from the Faculty of Physics and Mathematics of the Abai Almaty State University. Master of NARXOZ University. She defended her dissertation for the degree of Doctor of Philosophy in the specialty 6D07400 – Computer engineering and software on the topic „Development of a system for analyzing multilingual text information based on machine learning".

http://orcid.org/0000-0002-8627-1949

Prof. Sergii Shuvalov
e-mail: surgeon.shuvalov@gmail.com

Professor, Doctor of Medicine, Head of the Department of Surgical Dentistry and Maxillofacial Surgery, National Pirogov Memorial Medical University, Vinnytsia, Ukraine.
Scientific direction: maxillofacial surgery, medical diagnostics, engraftment of dental implants in patients with chronic liver pathology, therapeutic effect, regeneration.

https://orcid.org/0000-0001-5052-680X

M.Sc. Abilkaiyr Nazerke
e-mail: abilkaiyr.nazerke@gmail.com

Abilkaiyr Nazerke is a graduate of the Faculty of General Medicine of Kazakh National Medicine University. Currently works at the Faculty of Medicine and Healthcare of Al-Farabi Kazakh National University. She is the author or co-author of several scientific publications in the field of public health, epidemiology primary healthcare and economics in public health.

https://orcid.org/0000-0003-1603-5577

http://doi.org/10.35784/iapgos.5895 received: 14.02.2024 | revised: 21.02.2024 | accepted: 14.03.2024 | available online: 31.03.2024

ROOT SURFACE TEMPERATURE MEASUREMENT DURING ROOT CANAL OBTURATION

Les Hotra[1], Oksana Boyko[2], Igor Helzhynskyy[1], Hryhorii Barylo[1], Pylyp Skoropad[3], Alla Ivanyshyn[3], Olena Basalkevych[2]

[1]Lviv Polytechnic National University, Department of Electronic Engineering, Lviv, Ukraine, [2]Danylo Halytsky Lviv National Medical University, Department of Medical Informatics, Lviv, Ukraine, [3]Lviv Polytechnic National University, Department of Measuring Information Technology, Lviv, Ukraine

Abstract. *Prolonged exposure to elevated temperatures exceeding 47°C, which can occur during root canal obturation, can cause damage of both dental and bone tissues. In order to study the temperature distribution on the surface of the tooth root a temperature measuring device with cold-junction compensation is proposed. For in vitro measurement of the temperature distribution on the surface of the tooth, 8 thermocouples placed in direct contact with the cementum of the tooth were used. In order to eliminate the cold-junction temperature variations, the temperature equilibration device and RTD were used. The suggested linear approximation for the thermocouples' conversion function provides a nonlinearity relative error of less than 0.05% for K-type thermocouples and 0.07% for J-type thermocouples over the temperature range from 20 to 60°C.*

Keywords: temperature measurement, thermocouples, root canal obturation, linearisation

POMIAR TEMPERATURY POWIERZCHNI KORZENIA PODCZAS OBTURACJI KANAŁÓW KORZENIOWYCH

Streszczenie. *Długotrwała ekspozycja na podwyższone temperatury przekraczające 47°C, które mogą wystąpić podczas wypełniania kanałów korzeniowych, może spowodować uszkodzenie zarówno tkanek zęba, jak i kości. W celu zbadania rozkładu temperatury na powierzchni korzenia zęba zaproponowano urządzenie do pomiaru temperatury z kompensacją zimnego złącza. Do pomiaru in vitro rozkładu temperatury na powierzchni zęba wykorzystano 8 termopar umieszczonych w bezpośrednim kontakcie z cementem zęba. W celu wyeliminowania wahania temperatury zimnego złącza zastosowano urządzenie do wyrównania temperatur oraz czujnik rezystancyjny RTD. Proponowana aproksymacja liniowa funkcji przetwarzania termopary zapewnia względny błąd nieliniowości mniejszy niż 0,05% dla termopar typu K i 0,07% dla termopar typu J w zakresie temperatur od 20 do 60°C.*

Słowa kluczowe: pomiar temperatury, termopary, obturacja kanałów korzeniowych, linearyzacja

Introduction

Superior root canal obturation is a crucial prerequisite for the success of endodontic treatment. Incomplete filling can jeopardise the success of root canal treatment. Various methods and materials are employed for root canal obturation [3, 7]. The most popular obturation techniques involves the use of gutta-percha as the core material and an endodontic sealer [22]. During cold lateral compaction gutta-percha cannot fill the lateral and apical canals, and the cavities can be the origin of inflammatory processes in periodontal tissues and destruction of bone tissue [2, 5, 6]. Therefore the softening of gutta-percha by heat is used. However, usage of substantial amount of heat causes potential risks, as the excessive heat may transfer to the adjacent periodontal tissues. Prolonged exposure to elevated temperatures exceeding 47°C can cause damage to both dental and bone tissues. There is a suggestion that exceeding of the temperature by more than 10°C at the external root surface could potentially lead to the damage in cementum, periodontal ligament, and alveolar bone, possibly inducing resorption or ankylosis [17, 23]. If the damage is short-time and the temperature does not exceed 53°C, it may be reversible; however, temperatures beyond this threshold could lead to irreversible bone damage [8]. The danger of excessive temperature increase also occurs during post space preparation, when it is necessary to carry out complete removal of root filling materials to strengthen the adhesive bond with dentin and increase post-retention [19]. During the heated gutta-percha technique of the root canal filling or post space preparation, it is necessary to evaluate the temperature on the root surface.

1. Determination of root surface temperature

To examine the root surface temperature distribution, it is necessary to perform in vitro temperature measurements directly on the root surface.

In this approach, the split tooth model is employed [26], in which during obturation the extracted tooth is placed in some medium and fixed in a holder, or manually held by forceps in the hands of the researcher [15].

To measure the temperature range during in vitro root canal obturation, various primary temperature transducers such as thermoelectric, thermoresistive or semiconducting ones can be employed [4, 10–12, 14]. The temperature measurements in separate points on the root surface and inside the root canal can be effectively performed using thermocouples [20, 24]. Compactness, wide temperature range, interchangeability, relatively fast thermal response, long term of operation, reliability and affordable price are the main advantages of thermocouples.

The K type (chromel-alumel) and J type (iron-constantan) are the most prevalent thermocouples, known for their affordability, precision, dependability and wide temperature range [21].

The thermocouples used for measurement were placed in direct contact with the cementum layer on the root surface [9], as shown in Fig. 1. To minimise the impact of external temperature, the thermocouples were housed in thermally and electrically insulating tubes.

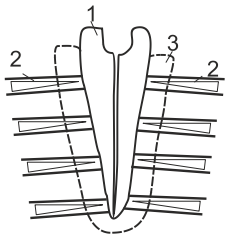


Fig. 1. The arrangement of thermocouples on the root surface: 1 – tooth, 2 – thermocouples, 3 – artificial substitute of periodontal tissues

To prevent heat dissipation loss from the tooth, it is advisable to use thermocouple electrodes with a minimal diameter. It is necessary to ensure tight fit of the measuring junction and electrodes of the thermocouple to the examined tooth surface using a heat-conducting material. The electrodes of the thermocouples are thermally insulated with a coating of thermoisolating lacquer.

The extracted tooth, along with the attached thermocouples, is then positioned in an environment with thermal conductivity similar to that of human tissues.

2. The design of temperature measuring device

For measurements of thermocouple output signals, an 8-channel microprocessor temperature measuring device has been designed. Its structural scheme (Fig. 2) consists of a temperature equalising device (TED), Resistance Temperature Detector (RTD) (R_{t0}) for measuring the temperature of the cold junctions of the thermocouples; a block of resistive temperature transducer (RTT); a 9-channel voltage commutator (VC); an analogue-to-digital converter (ADC); a microprocessor (MP); a digital readout (DR); an output interface (OI) and a supply unit (SU).

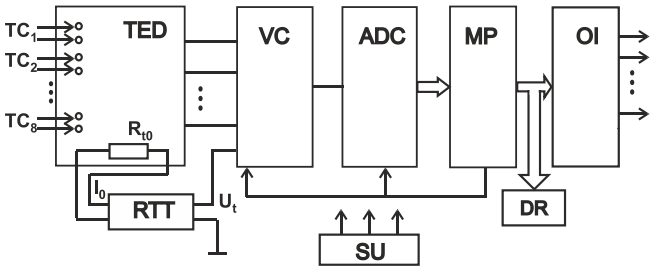


Fig. 2. Structural scheme of the temperature measuring device

The output terminals of the thermocouples are connected to the input terminals of the temperature equalising device. The values of thermo electromotive force (thermo-emf) generated in the thermocouples are a function of the temperature difference between the measuring and cold junctions:

$$e = f(t_h - t_c) = f(t_h) - f(t_c) \tag{1}$$

where t_h is the temperature of measuring (i.e. hot) junction, and t_c is the temperature of cold junction.

If the temperature of cold junction is not equal to 0ºC, the thermo-emf is equal to:

$$e = f(t_h) - e_{tc} \tag{2}$$

where e_{tc} is the value of thermo-emf at $t = t_c$.

The precision of temperature measurement is significantly influenced by the cold junction temperature, particularly in cases of lower measured temperatures [18, 25]. In order to eliminate the cold-junction temperature variations, either thermostatting of cold junction at 0ºC or electrical compensating circuits are used [13].

In the designed measuring device, the temperature equalising device was used for compensation of cold-junction temperature variations. It ensures the uniform temperature of cold junctions of the thermocouples and of the RTD. The long-term thermal stability of the circuit is achieved by placing it in a passive thermostat.

The RTD is connected to the RTT circuit, which is designed to generate a voltage proportional to the temperature of the cold junctions of the thermocouples.

The voltage drop across the resistor R_{t0} is determined by the value of the RTT reference current and is described by the following expression:

$$U_{Rt} = I_0 R_{t0}(1 + \alpha t_c) \tag{3}$$

where I_0 is the value of the reference current; R_{t0} is the resistance value of the RTD at a temperature of 0ºC, α is the temperature coefficient of resistance.

The output voltage of the RTT is described by the expression:

$$U_t = I_0 R_{t0} \alpha t_c k_1 \tag{4}$$

where k_1 is the conversion factor of the RTT.

For compensation of the influence of the temperature of cold junctions the equality $I_0 R_{t0} \alpha t_c k_1 = e_{tc}$ is required. This is achieved by choosing the coefficient k_1.

Compensation can be done in analogue or digital form (hardware or software compensations) [16]. In the first case, the RTT output is connected in series to the thermocouple outputs using a switch. In the second case it is connected to the ADC input.

The output voltage of a thermocouple exhibits a nonlinear relationship with the temperature difference between its hot and cold junctions [1]. To determine the corresponding temperature for a given measured voltage, tabulated values related to thermo-emf or approximating functions must be used. Utilising calibration charts as lookup tables demands significant memory capacity for saving the table content. Conversely, applying higher-degree polynomials for this purpose results in slower measurement rates. Therefore, the optimal approach is a linear approximation, with the accuracy of the approximation depending on the thermocouple type and the temperature range being measured.

Linearisation can be achieved by analogue or digital means. When the linear approximation function includes the coordinate origin and the termo-emf value at the endpoint of the measured temperature range, the nonlinearity error is formulated as follows:

$$\delta = \left(1 - \frac{e_k t}{e_t t_k}\right) \cdot 100\% \tag{5}$$

where e_t is the output voltage at the measured temperature t; e_k is the output voltage at the end point of the temperature range, t_k denotes the end point value of the temperature range.

Fig. 3 illustrates the correlation between the relative nonlinearity error and temperature for linear approximation functions of the K-type thermocouple at different values of t_k.

From the analysis of graphical dependencies, it is clear that each linear function provides a minimum error in separate temperature ranges. The required linearisation error can be provided when different functions, depending on the measurement range, are used to approximate.

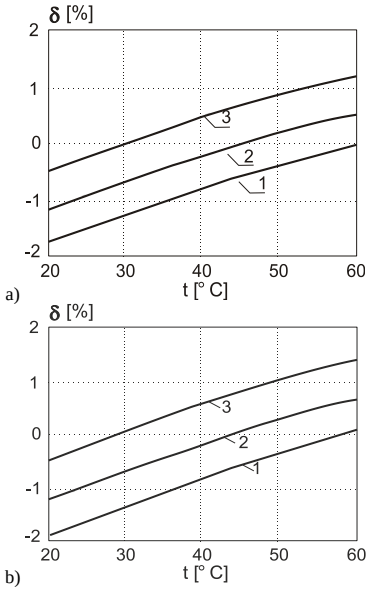


Fig. 3. The relative nonlinearity error vs. temperature at t_k = 60ºC (1), t_k = 45ºC (2), t_k = 30ºC (3) for K-type thermocouple (a) and J-type thermocouple (b)

Fig. 4 illustrates the correlation between linearisation error and temperature for a piecewise-linear approximation function with temperature spline points set at 30°, 40°, 50°, and 60°. The approximation error at the midpoint of each spline interval is zero.

Accuracy improvement and reducing the number of measuring ranges can be provided by employing an approximation function that passes through both the starting and ending points of each measuring range (Fig. 5). In this case, the relative approximation error does not exceed 0.05% for K-type thermocouple and 0.07% for J-type thermocouple (b).

Thus the comprehensive expression of approximation function for the thermo-emf of a thermocouple can be described by the function

$$e = a_i t + b_i \tag{6}$$

where a_i is the scale factor, and b_i is the constant term for the i-th range.

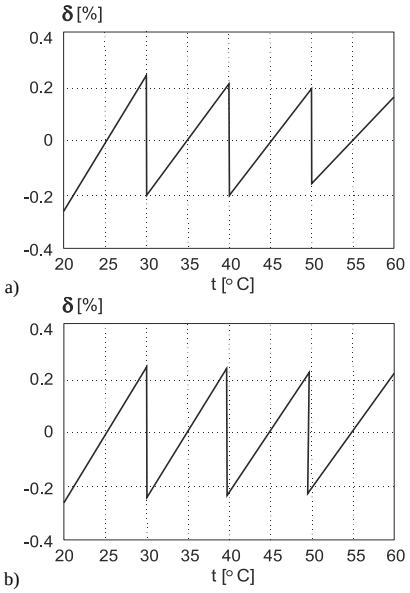


Fig. 4. The relationship between the linearisation error and temperature for piecewise-linear approximation function for K-type thermocouple (a) and J-type thermocouple (b)

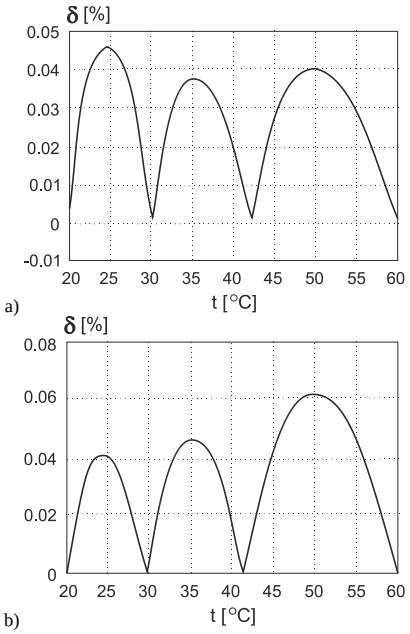


Fig. 5. The relationship between the relative approximation error and temperature for K-type thermocouple (a) and J-type thermocouple (b) in three intervals of piecewise-linear approximation

In this case, the temperature is expressed as:

$$t = \frac{e - b_i}{a_i} \tag{7}$$

The output voltage e_i of the i-th thermocouple and the output voltage of the RTT ΔU_i are passed to the input of ADC through the commutator controlled by the microprocessor. Thus, the voltage values of the thermocouples are converted into the output code of the ADC:

$$N_i = k_2 e_i \tag{8}$$

where k_2 is the conversion factor of the ADC. If $k_2 = 1$, the output code output code of the ADC is equal to the thermo-emf of the thermocouple.

When the output voltage of the RTT is converted, the output code of the ADC is equal to:

$$N_{Rt} = k_1 k_2 t_c \tag{9}$$

With the appropriate choosing the value of k_1, the output code is equal to the equivalent value of the thermo-emf at the temperature of cold junction.

The ADC output codes are passed into the MP input which performs the mathematical computation of ADC results. To decrease normal mode errors, the MP calculates the average of n voltage measurement results for each thermocouple and RTT:

$$N_{avi} = \frac{1}{n}\sum_{j=1}^{n} N_{ij} \;,\; N_{avRt} = \frac{1}{n}\sum_{j=1}^{n} N_{Rt_j} \tag{10}$$

where n is the number of the measured values at every channel.

After this, the MP calculates the equivalent values of thermo-emf of thermocouples at the temperature of cold junctions t_c:

$$N_{ec} = aN_{avRt} + b \tag{11}$$

where a and b are chosen according to the measured value of the cold-junction temperature.

Then, the MP calculates the equivalent values of thermo-emf of the thermocouples at $t_c = 0$°C:

$$N_{ei} = N_{avi} + N_{ec} \tag{12}$$

The value of the measured temperature is calculated using the formula:

$$t_i = \frac{N_{ei} + b_i}{a_i} \tag{13}$$

where the values of a and b are calculated for the corresponding range of the temperature measured value for the i-th channel.

Following the completion of the mathematical computation, the outcomes are transmitted from the microprocessor (MP) to a digital readout for all corresponding measuring channels, or for a single measuring channel. The microprocessor temperature measuring device can be connected through the output interface to a personal computer for additional processing of the measurement results during the given time period. In this case the investigation of the temperature fluctuations on the surface of the tooth root during endodontic treatment can be conducted.

3. Conclusions

A temperature-measuring device for dentistry investigations has been designed. The accuracy of a thermocouple's measurements, particularly at lower temperatures, is significantly influenced by the temperature of its cold junction and nonlinearity of the output characteristics of the thermocouple. The proposed piecewise-linear approximation of the conversion function of the thermocouple ensures the relative error of nonlinearity less than 0.05% for K-type thermocouples and 0.07% for J-type thermocouples for 3 approximation intervals over the temperature range from 20 to 60°C.

References

[1] Anandanatarajan R., Mangalanathan U., Gandhi U.: Deep Neural Network-Based Linearization and Cold Junction Compensation of Thermocouple. IEEE Transactions on Instrumentation and Measurement 72, 2022, 1–9.
[2] Balagopal S. et al.: Evaluation of remaining dentin thickness around the prepared root canals and its influence on the temperature changes on the external root surfaces during different heated gutta-percha obturation techniques. Indian Journal of Dental Research 31(6), 2020, 857–861.
[3] Bhandi S. et al.: Complete obturation–cold lateral condensation vs. thermoplastic techniques: a systematic review of micro-CT studies. Materials 14(14), 2021, 4013.
[4] Boyko O., Hotra O.: Improvement of dynamic characteristics of thermoresistive transducers with controlled heating, Przegląd elektrotechniczny 2019(5), 2019, 110–113.
[5] Diegritz C., Gerlitzki O., Fotiadou C., Folwaczny M.: Temperature changes on the root surface during application of warm vertical compaction using three different obturation units. Odontology 108, 2020, 358–365.
[6] Donnermeyer D., Schäfer E., Bürklein S.: Real-time intracanal temperature measurement during different obturation techniques. Journal of endodontics 44(12), 2018, 1832–1836.
[7] Epley S. R., Fleischman J., Hartwell G., Cicalese C.: Completeness of Root Canal Obturations: Epiphany Techniques versus Gutta-Percha Techniques. J. Endod. 32, 2006, 541–544.
[8] García-Cuerva M. et al.: Root surface temperature variation during mechanical removal of root canal filling material. An in vitro study. Acta odontologica latinoamericana – AOL 30(1), 2017.
[9] Hotra O.: Microprocessor temperature meter for dentistry investigation. Przegląd Elektrotechniczny 86 (7), 2010, 63–65.

[10] Hotra O.: Transistor-based temperature measuring device. Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska – IAPGOS 10(2), 2020, 4–7.
[11] Hotra O., Boyko O., Zyska T.: Improvement of the operation rate of medical temperature measuring devices. Proc. SPIE 92914, 2014, 92910A-92910A–6.
[12] Hotra O., Boyko O.: Analogue linearization of transfer function of resistive temperature transducers. Proc. SPIE 9662, 2015, 966247-966247–8.
[13] Hotra O., Boyko O.: Compensation bridge circuit with temperature-dependent voltage divider. Przeglad elektrotechniczny 4a, 2012, 169–171.
[14] Izhar U., Piyathilaka L., Preethichandra D. M. G.: Sensors for brain temperature measurement and monitoring–a review. Neuroscience Informatics 2(4), 2022, 100106.
[15] Lee F. S., Van-Cura J. E., Begole E. A.: A comparison of root surface temperatures using different obturation heat sources. Journal of Endodentistry 24, 1998, 617–620.
[16] Maseko M. L., Agee J. T., Davidson I.: Thermocouple signal conditioning using augmented device tables and table look-up neural networks, with validation in J-Thermocouples. IEEE 30th Southern African Universities Power Engineering Conference – SAUPEC, 2022, 1–4.
[17] Mena-Álvarez J. et al.: Comparative Analysis of Temperature Variation with Three Continuous Wave Obturation Systems in Endodontics: An In Vitro Study. Applied Sciences 12(12), 2022, 6229.
[18] Mukherjee A. et al.: An analog signal conditioning circuit for thermocouple temperature sensor employing thermistor for cold junction compensation. International conference on control, automation, robotics and embedded systems (care) 2013, 1–5.
[19] Radeva E. et al.: Evaluation of the apical seal after post space preparation: in vitro study. Journal of IMAB–Annual Proceeding Scientific Papers 25(1), 2019, 2327–2331.
[20] Radeva E. et al.: Temperature changes on the external root surface during post space preparation (in vitro study). Journal of IMAB–Annual Proceeding Scientific Papers 23(4), 2017, 1839–1844.
[21] Seung M., Choi W., Hur S., Kwon I.: Cold Junction Compensation Technique of Thermocouple Thermometer Using Radiation-Hardened-by-Design Voltage Reference for Harsh Radiation Environment. IEEE Transactions on Instrumentation and Measurement 71, 2022, 1–7.
[22] Simeone M. et al.: Temperature Profiles Along the Root with Gutta-percha Warmed through Different Heat Sources. The Open Dentistry Journal 8, 2014, 229.
[23] Singla M., Aggarwal V., Sinha N.: External root surface temperature changes during high-temperature injectable thermoplasticized root canal obturation in simulated immature teeth. Saudi Endodontic Journal 10(1), 2020, 51–55.
[24] Sivakumar A. A. et al.: Evaluation of temperature change in the root surface during post space preparation using two different drill systems – An invitro study. IOSR Journal of Dental and Medical Sciences 20 (10), 2021, 01–06.
[25] Wei G., Wang X., Sun J.: Signal processing method with cold junction compensation for thermocouple. IEEE Instrumentation and Measurement Technology Conference 2009, 1458–1462.
[26] Weller R. N. et al., A new model system for measuring intra-canal temperatures. Journal of Endodentistry 17, 1991, 491–494.

M.Sc. Les Hotra
e-mail: les.m.hotra@lpnu.ua

Les Hotra graduated from Department of Applied Mathematics, Lviv Polytechnic National University. He is currently a postgraduate student at Lviv Polytechnic National University.
His areas of scientific interest cover mathematical modelling and electronics including biomedical devices.

https://orcid.org/0009-0005-1351-1883

Prof. Oksana Boyko
e-mail: oxana_bojko@ukr.net

Oksana Boyko is currently a Head of the Medical Informatics Department of Danylo Halytsky Lviv National Medical University.
Her areas of scientific interest cover mathematical modelling, biomedical devices and medical information systems including elements of artificial intelligence. She is the author of over 200 scientific and methodological works.

https://orcid.org/0000-0002-8810-8969

Prof. Igor Helzhynskyy
e-mail: iigorg@ukr.net

Igor Helzhynskyy is a doctor of solid-state electronics, professor in the Department of Electronic Engineering of Lviv Polytechnic National University. He has been participated in numerous Ukrainian and international projects related to materials science, engineering, in particular organic and gibrid light-emitting devices for organic electronics. His research area focuses on WOLED, PhOLED, QLED and electronics.

https://orcid.org/0000-0002-1931-6991

Prof. Hryhorii Barylo
e-mail: gbarylo@polynet.lviv.ua

Hryhorii Barylo has worked for many companies and institutions. He was the Head of the Microprylad production enterprise in Lviv. Since 2008 he has been working at the Department of Electronic Engineering of Lviv Polytechnic National University. His research activity is focused on the problem of the use of impedance spectrometry in sensor technology, materials science, biological and medical research. His scientific approaches are based on the results of mathematical modelling, elements of artificial intelligence, systems and achievements of Internet technologies.

https://orcid.org/0000-0001-5749-9242

Prof. Pylyp Skoropad
e-mail: pylyp.i.skoropad@lpnu.ua

Pylyp Skoropad is currently a professor in the Department of Measuring Information Technologies of Lviv Polytechnic National University. His areas of scientific interest cover contact and contactless thermometry (from low to high temperatures) and materials science.
He is an author of more than 200 scientific publications, including 2 monographs and 10 patents of Ukraine.

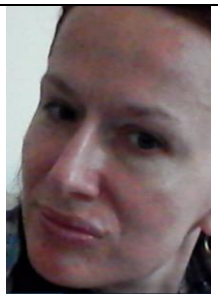
https://orcid.org/0000-0003-3559-6580

Ph.D. Alla Ivanyshyn
e-mail: alla.v.hunkalo@lpnu.ua

Alla Ivanyshyn is currently an Associate Professor in the Department of Measuring Information Technologies of Lviv Polytechnic National University (Ukraine). Her areas of scientific research include expert systems, fuzzy logic, quality monitoring, and management. She is an author of more than 60 scientific publications, including over 30 articles. She is the coordinator of the international educational project Erasmus + Jean Mone 101085516 – QMSEEI (2022-2025). She has 15 years of experience in scientific and pedagogical work.

https://orcid.org/0000-0002-3302-7889

Ph.D. Olena Basalkevych
e-mail: olena.basalkevych@gmail.com



Olena Basalkevych is currently an Assistant in the Medical Informatics Department of Danylo Halytsky Lviv National Medical University.
Ph.D. in Philology, Modelling of the Thesaurus of Qualitative Adjectives in Older Scots (Mathlinguistics)
Her areas of scientific research include expert systems, artificial intelligence and mathematical modelling.
She is the author of 30 publications, 11 in Scopus, 5 reviews.

https://orcid.org/0000-0001-5886-6374

http://doi.org/10.35784/iapgos.5555 *received: 09.11.2023 | revised: 15.02.2024 | accepted: 09.03.2024 | available online: 31.03.2024*

EVALUATING THE FEASIBILITY OF THERMOGRAPHIC IMAGES FOR PREDICTING BREAST TUMOR STAGE USING DCNN

Zakaryae Khomsi, Mohamed El Fezazi, Achraf Elouerghi, Larbi Bellarbi
Mohammed V University in Rabat, Ecole Nationale Supérieure d'Arts et Métiers (ENSAM), Ecole Nationale Supérieure d'Informatique et d'Analyse des Systèmes (ENSIAS), Electronic Systems Sensors and Nanobiotechnologies (E2SN), Rabat, Morocco

Abstract. *Early-stage and advanced breast cancer represent distinct disease processes. Thus, identifying the stage of tumor is a crucial procedure for optimizing treatment efficiency. Breast thermography has demonstrated significant advancements in non-invasive tumor detection. However, the accurate determination of tumor stage based on temperature distribution represents a challenging task, primarily due to the scarcity of thermal images labeled with the stage of tumor. This work proposes a transfer learning approach based on Deep Convolutional Neural Network (DCNN) with thermal images for predicting breast tumor stage. Various tumor stage scenarios including early and advanced tumors are embedded in a 3D breast model using the Finite Element Method (FEM) available on COMSOL Multiphysics software. This allows the generation of the thermal image dataset for training the DCNN model. A detailed investigation of the hyperparameters tuning process has been conducted to select the optimal predictive model. Thus, various evaluation metrics, including accuracy, sensitivity, and specificity, are computed using the confusion matrix. The results demonstrate the DCNN model's ability to accurately predict breast tumor stage from thermographic images, with an accuracy of 98.2%, a sensitivity of 98.8%, and a specificity of 97.7%. This study indicates the promising potential of thermographic images in enhancing deep learning algorithms for the non-invasive prediction of breast tumor stage.*

Keywords: image analysis, classification, tumor prediction, transfer learning, thermography

OCENA WYDAJNOŚCI OBRAZÓW TERMOGRAFICZNYCH DO PRZEWIDYWANIA STOPNIA GUZA PIERSI PRZY UŻYCIU DCNN

Streszczenie. *Wczesny i zaawansowany rak piersi stanowią odrębne procesy chorobowe. Dlatego też identyfikacja stadium nowotworu jest kluczową procedurą dla optymalizacji skuteczności leczenia. Termografia piersi wykazała znaczny postęp w nieinwazyjnym wykrywaniu nowotworów. Jednak dokładne określenie stopnia zaawansowania nowotworu na podstawie rozkładu temperatury stanowi trudne zadanie, głównie ze względu na niedobór obrazów termicznych oznaczonych stopniem zaawansowania nowotworu. W niniejszej pracy zaproponowano podejście uczenia transferowego oparte na głębokiej konwolucyjnej sieci neuronowej (DCNN) z obrazami termicznymi do przewidywania stadium guza piersi. Różne scenariusze stadium nowotworu, w tym guzy wczesne i zaawansowane, są osadzone w trójwymiarowym modelu piersi przy użyciu metody elementów skończonych (MES) dostępnej w oprogramowaniu COMSOL Multiphysics. Pozwala to na wygenerowanie zestawu danych obrazów termicznych do trenowania modelu DCNN. Przeprowadzono szczegółowe badanie procesu dostrajania hiperparametrów w celu wybrania optymalnego modelu predykcyjnego. W związku z tym różne wskaźniki oceny, w tym dokładność, czułość i swoistość, są obliczane przy użyciu macierzy pomyłek. Wyniki pokazują zdolność modelu DCNN do dokładnego przewidywania stadium guza piersi na podstawie obrazów termograficznych, z dokładnością 98,2%, czułością 98,8% i swoistością 97,7%. Badanie to wskazuje na obiecujący potencjał obrazów termograficznych w ulepszaniu algorytmów głębokiego uczenia się w celu nieinwazyjnego przewidywania stadium guza piersi.*

Słowa kluczowe: analiza obrazu, klasyfikacja, przewidywanie nowotworów, uczenie transferowe, termografia

Introduction

The Breast cancer is a chronic and common disease that often affects women worldwide. In 2020, the World Health Organization (WHO) reported 685,000 deaths and 2.3 million new cases related to breast cancer [3]. Several factors, including age, genetics, hormonal imbalances, and environmental exposures, can contribute to breast cancer [10]. Early-stage cancer primarily exhibits localized growth without dissemination, while advanced-stage cancer is characterized by the spread of the disease [7]. The disease's progression from early to advanced stages can lead to more aggressive treatments and poorer outcomes. For this reason, the accurate determination of breast cancer stage, typically based on tumor size, is essential for choosing the most appropriate treatment, and improving the patient's outcomes [1].

Traditional methods, such as biopsy, ultrasound, and mammography, have long been employed for this purpose. The objective of a biopsy is to identify the type and stage of the tumor and evaluate its potential invasiveness. This procedure is conducted by a medical professional, such as a surgeon, who carefully selects the site for tissue collection, often guided by imaging techniques like ultrasound or CT scans [2]. Once the sample is obtained, it is then subjected to laboratory analysis, including histological examination and molecular testing. Potential risks associated with biopsies include moderate pain, post-procedural bleeding, and risk of infection at the biopsy site.

Mammography is a widely used breast cancer screening tool, utilizing X-rays to create detailed breast tissue images through differential X-ray attenuation. The procedure involves breast compression between plates to reduce scatter and ensure consistent thickness, with resulting images analyzed by radiologists for breast tumor stage detection [14]. However, mammography has limitations such as reduced sensitivity in dense breast tissue, ionizing radiation exposure, and patient discomfort due to breast compression [18]. These challenges underscore the need for non-invasive screening alternatives for improved breast cancer staging.

Breast thermography is a promising method for early breast cancer detection, characterized by its non-invasive nature, absence of radiation exposure, and cost-effectiveness [9]. It relies on the concept that cancerous cells exhibit higher metabolic activity, generating more heat than normal cells [11]. Thermography offers a non-invasive and radiation-free approach to detect the existence of potential tumors, contrasting with the radiation exposure of mammography and the operator dependency of ultrasound, enhancing patient comfort and safety. However, the determination of the stage of a breast tumor based on thermography is a critical challenge due to the scarcity of thermographic images with labeled tumor stages. Addressing this issue is crucial for enhancing the artificial intelligence algorithms to predict tumor stage by thermal imaging. Advanced simulation methods offer a promising solution to overcome this limitation [12].

In this study we developed a comprehensive approach to predict the stage of breast tumors based on thermal image dataset. Our key contributions could be summarized as follows:

- We developed a well-labeled thermal images dataset including several scenarios of early and advanced-stage breast tumors using FEM. That allows an understanding of thermal patterns obtained from various scenarios of breast cancer.
- We proposed a pre-trained model (MobileNet) for predicting tumor stage based on thermal images. This provides a good compromise between better prediction performance and computational efficiency.

- We selected the optimal predictive model through a detailed investigation of the hyperparameters tuning process. This contributes to improved generalization for the proposed DCNN model to perform well on unseen data.

The rest of this paper is structured as follows. Section 1 details our methodology from the data generation process to the formulation of the DCNN model. In section 2, we discuss the obtained results during the hyperparameters tuning process to select the optimal model. In section 3, we present the concluding remarks of the entire study.

1. Methodology

The methodology includes several key phases: Data generation process using FEM, data preparation involving labeling and splitting, and the development of the DCNN model.

1.1. FEM-based breast tissue modeling

First, we designed a hemispherical model (Figure 1) for simulating the breast geometry including tissues such as skin, gland, fat, and muscle in the COMSOL software. Then, we applied the Pennes equation (1) [5] for simulating and studying thermal distribution in breast tissues for different tumor scenarios. Table 1 summarizes the thermal properties considered in the modeling [16].

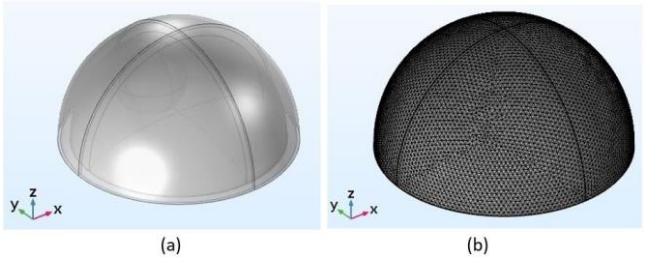


Fig. 1. (a) Breast geometry in 3D modeling. (b) Generated mesh

$$\nabla(k_i \cdot \nabla T_i) + c_b \cdot \rho_b \cdot \omega_{b,i}(T_b - T_i) + q_{m,i} = 0 \quad (1)$$

where i represents the breast layers. The thickness of skin, fat, gland, and muscle is set to 1.6 mm, 5 mm, 43.4 mm, and 15 mm respectively. T_i, and $q_{m,i}$ is respectively the temperature and the metabolic heat generation rate of the tissues. T_b is the arterial blood temperature (37°C), k_i is the thermal conductivity for each tissue, $\omega_{b,i}$ is the blood perfusion rate for each tissue. A room temperature of 25°C is set into the model, which reflects the common range for medical room temperatures.

Table 1. Thermal characteristics of breast tissue

	q_m (W/m^3)	k (W/m.K)	ω_b (ml.s^{-1}.ml^{-1})
Skin	368.1	0.45	0.00018
Fat	400	0.21	0.00022
Gland	700	0.48	0.00054
Muscle	700	0.48	0.00270
Tumor	70,000	0.62	0.01600

1.2. Data collection

Researchers categorized cancer into early and advanced-stage tumors (table 2) [6]. The first tumor category (T1-stage) consists of tumor sizes less than 20 mm. This early-stage category require less aggressive treatment in many cases. The second category (T2-stage) includes tumors between 20 mm and 50 mm. This last may require more aggressive treatments, including high-dose of chemotherapy and radiotherapy. The other categories contain tumors larger than 50 mm, which represent tumor metastases that are very difficult to treat.

Table 2. Range of tumor stage T1 and T2

Tumor stage	T1	T2
Range of tumor size (mm)	2 - 20	21 - 40
Image data	590	569
Dissemination risk	Low	High

In this study, we will focus on predicting T1 and T2 tumor stage categories, which can improve the prognosis and the chances of recovery. Using FEM modeling previously described in section 1.1, we include various permutations of tumor sizes and locations resulting in a dataset of 1159 different cases of thermographic images. Each image has a resolution of 224×224 pixels as shown in figure 2.

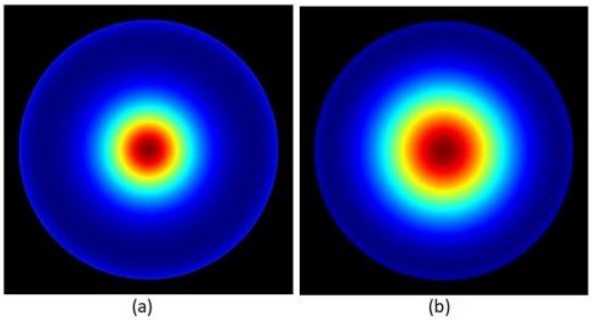


Fig. 2. Example of thermal images generated: (a) T1-stage tumor. (b) T2-stage tumor

The early-stage class contained 590 images, while the advanced-stage class contained the remaining 569 images. The dataset was labled and randomly split into training and validation sets with a ratio of 85:15, with 984 images used for training and 175 images for test. The training set was used to train our DCNN model, while the test set was used to evaluate the model's performance.

1.3. Architecture of the pre-trained DCNN model

Our proposal involved the use of CNN-based MobileNet architecture available on Teachable Machine 2.0, which is a web-based tool that uses Google Cloud services to simplify and accelerate the creation of machine learning models for image classification. MobileNet uses depthwise separable convolutions (DSC) [17] that reduce the number of parameters and computations required while maintaining accuracy [13]. Figure 3 shows the MobileNet principle.

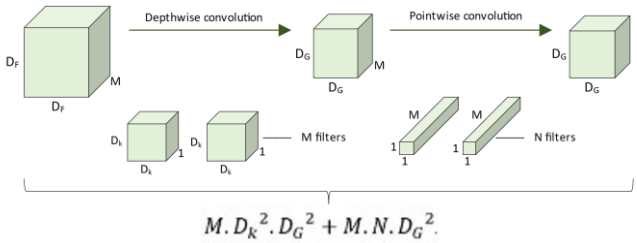


Fig. 3. Depthwise separable convolution principle

In depth wise convolution, one filter is applied to each input channel independently with fewer parameters. Then, in pointwise convolution, 1×1 filters are applied to combine the output feature maps for dimensionality reduction. The mathematical expression of depthwise and pointwise convolution is formulated in (2) and (3). Finally, we get (4) [19].

$$D_{x,y,m} = \sum_m \sum_{i,j} K_{i,j,m} \cdot I_{x+i-1,y+j-1,m} \quad (2)$$

$$P_{x,y,m} = \sum_m K'_{m,n} \cdot D_{x,y,m} \quad (3)$$

$$P_{x,y,m} = \sum_m \sum_{i,j} K'_{m,n} \cdot K_{i,j,m} \cdot I_{x+i-1,y+j-1,m} \quad (4)$$

where, $D_{x,y,m}$ is output of depthwise convolution, $P_{x,y,m}$ is output of pointwise convolution, I is input image, K is depth wise convolution kernel, K' is pointwise convolution kernel.

Figure 4 represents the global architecture adopted on Teachable Machine 2.0 for tumor stage detection. The model's architecture is composed of multiple layers of DSC with rectified linear activation function (ReLU) (5) [15], followed by global average pooling, a fully connected layer, and a softmax activation function (6) [20] for the classification of images. We developed the initial dense layer and the last softmax layer with two classes (T1-stage and T2-stage thermal images). The model learns to extract features in a more efficient way by separately applying filters to each channel of the input, rather than convolving the entire input with a single filter. This reduces the number of computations needed in the convolutional layer, making the model faster and more memory-efficient, while still preserving important features in the data [8].

$$ReLU(z) = \begin{cases} 0 \; if \; z < 0 \\ z \; if \; z \geq 0 \end{cases} \quad (5)$$

$$softmax(z)_{[i]} = \frac{e^{z[i]}}{\sum_j e^{z[j]}} \quad (6)$$

where, z is the input.

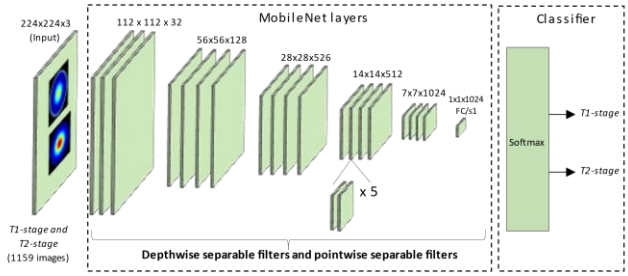


Fig. 4. Architecture adopted for tumor stage prediction

1.4. Hyperparameters tuning process

Hyperparameter tuning is a crucial step in the development of an accurate deep learning models [4]. Figure 5 indicates the process followed for selecting the best parameter combination. As shown in the flowchart (Figure 5), we considered independ-ent elements for evaluation, which included several combina-tions of epochs (Ep) and learning rates (LR). These elements were chosen to evaluate the test accuracy and loss of the model. By systematically varying these hyperparameters at 16 batch size, we achieved the optimal model able to detect accurately the tumor stage based on thermal images.

2. Results and discussion

We trained our deep learning model using different combinations of learning rates and epochs to investigate their impact on the performance of the model. Especially, we considered three different learning rates (0.00001, 0.0001, and 0.001), and nine different epoch numbers from 20 to 100 with an increment step of 10 for each combination. Figure 6 shows all the obtained combinations of training and test loss curves. As shown in Figure 6, increasing the learning rate up to 0.001 caused large fluctuations and noise on the loss curves. However, when the learning rate decreases to 0.0001 and 0.00001, the training and testing loss curves become smooth with faster convergence to the global minimum. We notice that the loss curves confronted a significant impact by adjusting the learning rate. Thus, tuning the epoch numbers leads to achieving the optimal performance of the model. Figure 7 includes the representation of the confusion matrix obtained from the testing set of each LR and epoch combination. We reported the overall test accuracy in table 3.

From table 3, we notice that the test accuracy is considerably influenced by both the LR and the number of epochs. Thus, the higher accuracy is obtained by running more epochs. Especially, on 80 epochs for the LR (= 0.0001), the model achieved the best performance with an accuracy of 98.28%.

Table 3. Comparative analysis for DCNN model selection

Epoch	Learning rate		
	0.00001	0.0001	0.001
20	86.28%	94.85%	89.14%
30	92%	94.85%	94.28%
40	92%	97.14%	93.71%
50	91.42%	97.71%	96.57%
60	89.71%	92.57%	93.71%
70	94.85%	95.42%	91.42%
80	93.71%	**98.28%**	97.14%
90	92%	97.14%	96%
100	93.71%	95.42%	96%

We summarized the performance metrics for this model in table 4. The table includes six metrics: accuracy, positive predictive value (PPV), negative predictive value (NPV), sensitivity (recall), specificity, and F1-score. Our results show that the model with the LR (= 0.0001) and 80 epochs achieved an accuracy of 98.2%, a sensitivity of 98.8%, a specificity of 97.7%, and an F1-score of 98.7% on the test set, indicating strong performance for tumor staging.

Table 4. Performance results for the optimal model (TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative, n: numerical application)

Parameter	Formula	n	Value
Accuracy	(TP+TN)/Total	(87+85)/175	0.982
Sensitivity (Recall)	TP/(TP+FN)	87/(87+1)	0.988
Specificity	TN/(TN+FP)	85/(85+2)	0.977
PPV	TP/(TP+FP)	87/(87+2)	0.977
NPV	TN/(TN+FN)	85/(85+1)	0.988
F1-score	2*(PPV*Recall)/ (PPV+Recall)	2*(0.977*0.988)/ (0.997+0.998)	0.987

Overall, our results suggest that careful tuning of the hyperparameters contributes for achieving optimal performance with a good fit. Thus, a learning rate of 0.0001 with 80 epochs was proved the optimal combination for our task.

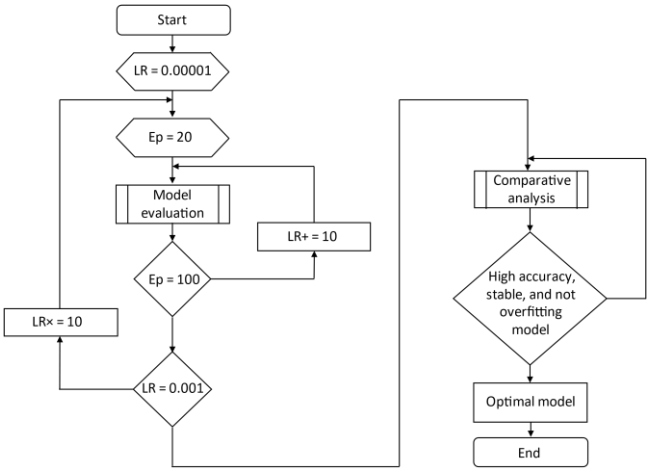


Fig. 5. Flowchart of the hyperparameters tuning process (LR: learning rate, Ep:number of epoch)

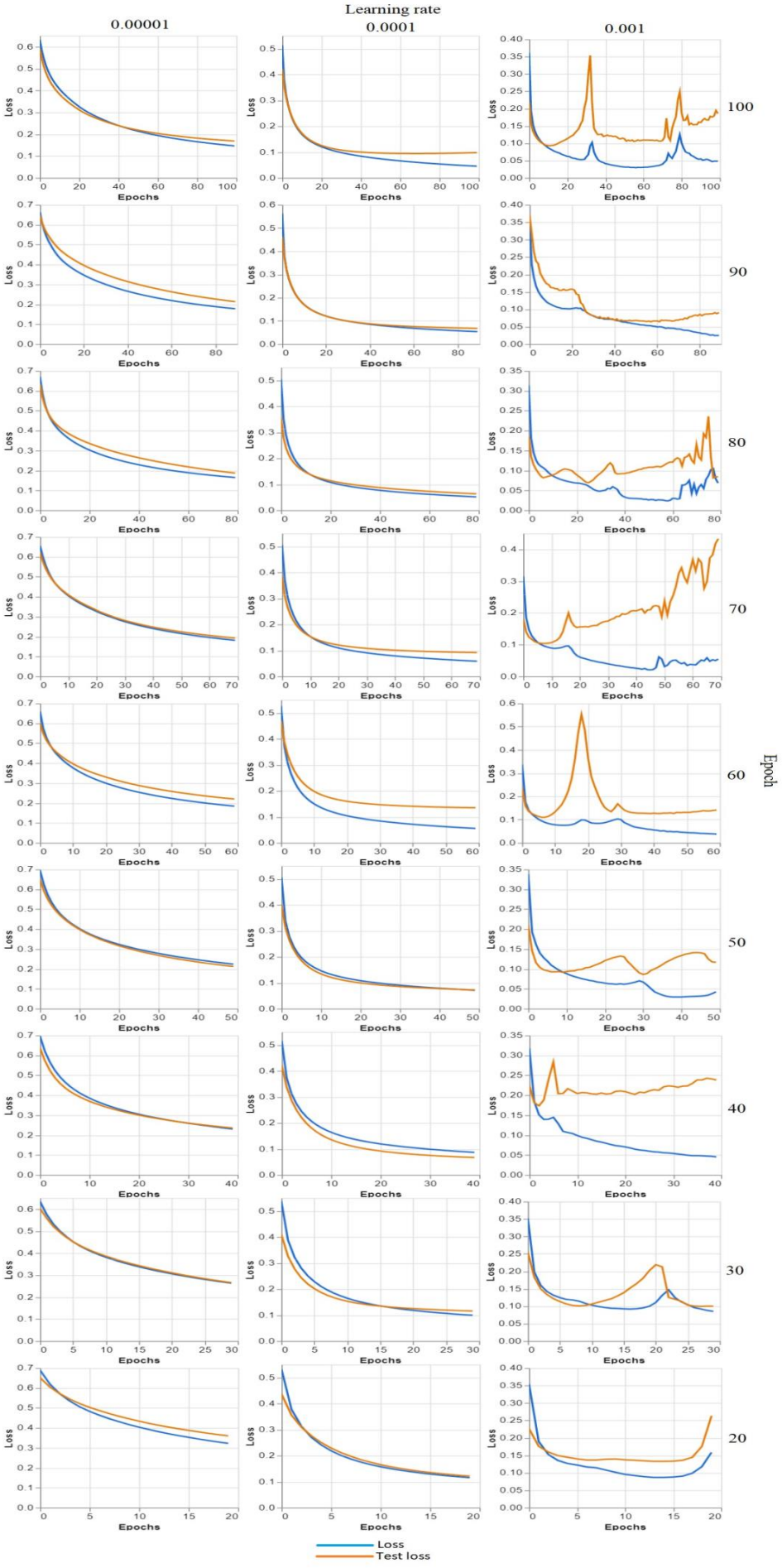


Fig. 6. Loss curves for the proposed model depending on different learning rates (horizontal) and number of epochs (vertical)

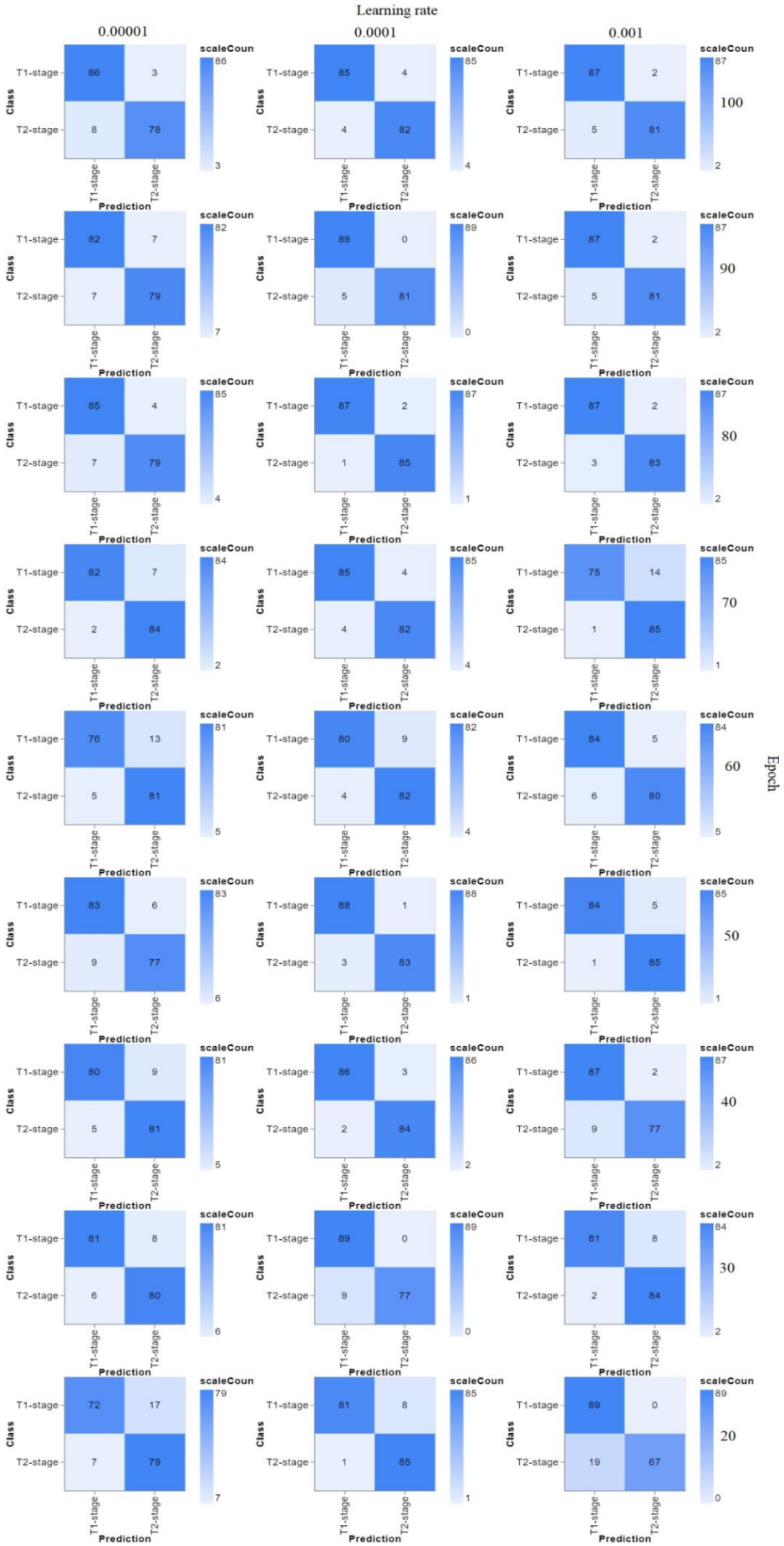


Fig. 7. The evaluation of the confusion matrix for the proposed model on different epochs (vertical) and learning rates (horizontal). The x-axis represents the model's predictions, while the y-axis represents the actual class of the dataset

3. Conclusion

In this paper, we introduced an appropriate DCNN-based methodology to enhance the prediction of breast tumor stage using thermographic images. Using the COMSOL software, we first developed a 3D breast model incorporating skin, fat, mammary gland, and muscle. Then, we included diverse scenarios of early and advanced tumor stages (T1-stage and T2-stage) to generate the thermal image dataset. Each image was labeled according to the corresponding tumor stage category for training the DCNN model. Multiple combinations of learning rates and number of epochs are investigated for adopting the optimal predictive model. According to the results, our approach shows the potential of the proposed DCNN model to predict tumor stage based on thermographic images with a reasonable accuracy of 98.2%. The optimum predictive model is obtained by running more epochs (80) in the LR of 0.0001. This study provides insights into the promising feasibility of thermographic images in combination with transfer learning techniques for tumor stage prediction. In future works, we aim to include the proposed framework in a smart thermographic system for investigating decisions on realistic subjects.

Acknowledgment

We thank Al-Khawarizmi program to encourage research in Artificial Intelligence and its applications.

References

[1] Ahlawat P. et al.: Tumour Volumes: Predictors of Early Treatment Response in Locally Advanced Head and Neck Cancers Treated with Definitive Chemoradiation. Reports of Practical Oncology and Radiotherapy 21(5), 2016, 419–426 [https://doi.org/10.1016/j.rpor.2016.04.002].
[2] Alghamdi S. et al.: The Impact of Reporting Tumor Size in Breast Core Needle Biopsies on Tumor Stage: A Retrospective Review of Five Years of Experience at a Single Institution. Annals of Diagnostic Pathology, vol. 38, 2019, 26–28 [https://doi.org/10.1016/j.anndiagpath.2018.10.002].
[3] De Miglio M. R., Mello-Thoms C.: Editorial: Reviews in Breast Cancer. Frontiers in Oncology 13, 2023, 1161583 [https://doi.org/10.3389/fonc.2023.1161583].
[4] Farhangi F.: Investigating the Role of Data Preprocessing, Hyperparameters Tuning, and Type of Machine Learning Algorithm in the Improvement of Drowsy EEG Signal Modeling. Intelligent Systems with Applications 15, 2022, 200100 [https://doi.org/10.1016/j.iswa.2022.200100].
[5] Gavazzi S. et al.: Advanced Patient-Specific Hyperthermia Treatment Planning. International Journal of Hyperthermia 37(1), 2020, 992–1007 [https://doi.org/10.1080/02656736.2020.1806361].
[6] Giuliano A. E. et al.: Breast Cancer-Major Changes in the American Joint Committee on Cancer Eighth Edition Cancer Staging Manual. CA: A Cancer Journal for Clinicians 67(4), 2017, 290–303 [https://doi.org/10.3322/caac.21393].
[7] Horvath L. E. et al.: The Relationship between Tumor Size and Stage in Early versus Advanced Ovarian Cancer. Medical Hypotheses 80(5), 2013, 684–687 [https://doi.org/10.1016/j.mehy.2013.01.027].
[8] Huang W. et al.: Wearable Health Monitoring System Based on Layered 3D-Mobilenet. Procedia Computer Science 202, 2022, 373–378 [https://doi.org/10.1016/j.procs.2022.04.051].
[9] Jacob G. et al.: Breast Cancer Detection: A Comparative Review on Passive and Active Thermography. Infrared Physics and Technology 134, 2023, 104932 [https://doi.org/10.1016/j.infrared.2023.104932].
[10] Jones S. C. et al.: Australian Women's Perceptions of Breast Cancer Risk Factors and the Risk of Developing Breast Cancer. Women's Health Issues 21(5), 2011, 353–360 [https://doi.org/10.1016/j.whi.2011.02.004].
[11] Kandlikar S. G. et al.: Infrared Imaging Technology for Breast Cancer Detection – Current Status, Protocols and New Directions. International Journal of Heat and Mass Transfer 108, 2017, 2303–2320 [https://doi.org/10.1016/j.ijheatmasstransfer.2017.01.086].
[12] Khomsi Z. et al.: Towards Development of Synthetic Data in Surface Thermography to Enable Deep Learning Models for Early Breast Tumor Prediction. Masrour T. et al. (eds): Artificial Intelligence and Industrial Applications. Springer Cham, Switzerland, 2023, 356–365 [https://doi.org/10.1007/978-3-031-43520-1_30].
[13] Lu S. Y. et al.: A Classification Method for Brain MRI via MobileNet and Feedforward Network with Random Weights. Pattern Recognition Letters 140, 2020, 252–260 [https://doi.org/10.1016/j.patrec.2020.10.017].
[14] Magario M. B. et al.: Mammography Coverage and Tumor Stage in the Opportunistic Screening Context. Clinical Breast Cancer 19(6), 2019, 456–459 [https://doi.org/10.1016/j.clbc.2019.04.014].
[15] Muruganandam S. et al.: A Deep Learning Based Feed Forward Artificial Neural Network to Predict the K-Barriers for Intrusion Detection Using a Wireless Sensor Network. Measurement: Sensors 25, 2023, 100613 [https://doi.org/10.1016/j.measen.2022.100613].
[16] Ragab M. et al.: Heat Transfer in Biological Spherical Tissues during Hyperthermia of Magnetoma. Biology 10(12), 2021, 1–16 [https://doi.org/10.3390/biology10121259].
[17] Rahman M. H. et al.: Real-Time Face Mask Position Recognition System Based on MobileNet Model. Smart Health 28, 2023, 100382 [https://doi.org/10.1016/j.smhl.2023.100382].
[18] Sardanelli F., Helbich T. H.: Mammography: EUSOBI Recommendations for Women's Information. Insights into Imaging 3(1), 2012, 7–10 [https://doi.org/10.1007/s13244-011-0127-y].
[19] Wang H. et al.: A Model for Detecting Safety Hazards in Key Electrical Sites Based on Hybrid Attention Mechanisms and Lightweight Mobilenet. Energy Reports 7, 2021, 716–724 [https://doi.org/10.1016/j.egyr.2021.09.200].
[20] Zhu D. et al.: Efficient Precision-Adjustable Architecture for Softmax Function in Deep Learning. IEEE Transactions on Circuits and Systems II: Express Briefs 67(12), 2020, 3382–3386 [https://doi.org/10.1109/TCSII.2020.3002564].

M.Sc. Zakaryae Khomsi
e-mail: zakaryae_khomsi@um5.ac.ma

Received the Master degree in electrical engineering from ENSAM, Mohammed V University, Rabat, Morocco, in 2018. He is currently Ph.D. student at ENSAM-ENSIAS and member of E2SN team, Mohammed V University, Rabat, Morocco.
His research interests are in biomedical engineering, image analysis and processing for early detection of breast tumors.

https://orcid.org/0000-0003-2321-9622

M.Sc. Mohamed El Fezazi
e-mail: elfezazi.med@gmail.com

Received the master degree in electrical engineering from ENSAM, Mohammed V University, Rabat, Morocco, in 2018. He is currently Ph.D. student at ENSAM-ENSIAS and member of E2SN team, Mohammed V University, Rabat, Morocco.
His research interests are in biomedical engineering, signal processing and telemedicine.

https://orcid.org/0000-0001-6072-325X

M.Sc. Achraf Elouerghi
e-mail: achraf_elouerghi@um5.ac.ma

Received his master degree in electrical engineering from ENSAM, Mohammed V University, Rabat, Morocco, in 2017. Currently Ph.D. student with E2SN research team at ENSAM, Mohammed V University in Rabat, Morocco.
His research activities are biomedical engineering, embedded systems, manufacturing engineering, radar, telecommunication systems, microelectronics, and semiconductor engineering.

https://orcid.org/0000-0001-5880-0172

Prof. Larbi Bellarbi
e-mail: l.bellarbi@um5r.ac.ma

Professor at the electrical engineering department of ENSAM, Mohammed V University in Rabat, Morocco. His research interests include electrical and biomedical engineering, telecommunications, electronic, and embedded systems. He is an academic member of the Electronic Systems, Sensors, and Nanobiotechnology (E2SN) research group.

https://orcid.org/0009-0004-5074-3117

http://doi.org/10.35784/iapgos.5804 *received: 16.01.2024 | revised: 12.02.2024 | accepted: 29.02.2024 | available online: 31.03.2024*

A COMPREHENSIVE STUDY: INTRACRANIAL ANEURYSM DETECTION VIA VGG16-DENSENET HYBRID DEEP LEARNING ON DSA IMAGES

Sobhana Mummaneni, Sasi Tilak Ravi, Jashwanth Bodedla, Sree Ram Vemulapalli, Gnana Sri Kowsik Varma Jagathapurao

Velagapudi Ramakrishna Siddhartha Engineering College, Department of Computer Science and Engineering, Vijayawada, India

Abstract. *An intracranial aneurysm is a swelling in a weak area of a brain artery. The main cause of aneurysm is high blood pressure, smoking, and head injury. A ruptured aneurysm is a serious medical emergency that can lead to coma and then death. A digital subtraction angiogram (DSA) is used to detect a brain aneurysm. A neurosurgeon carefully examines the scan to find the exact location of the aneurysm. A hybrid model has been proposed to detect these aneurysms accurately and quickly. Visual Geometry Group 16 (VGG16) and DenseNet are two deep-learning architectures used for image classification. Ensembling both models opens the possibility of using diversity in a robust and stable feature extraction. The model results assist in identifying the location of aneurysms, which are much less prone to false positives or false negatives. This integration of a deep learning-based architecture into medical practice holds great promise for the timely and accurate detection of aneurysms. The study encompasses 1654 DSA images from distinct patients, partitioned into 70% for training (1157 images) and 30% for testing (496 images). The ensembled model manifests an impressive accuracy of 95.38%, outperforming the respective accuracies of VGG16 (94.38%) and DenseNet (93.57%). Additionally, the ensembled model achieves a recall value of 0.8657, indicating its ability to correctly identify approximately 86.57% of true aneurysm cases out of all actual positive cases present in the dataset. Furthermore, when considering DenseNet individually, it attains a recall value of 0.8209, while VGG16 attains a recall value of 0.8642. These values demonstrate the sensitivity of each model to detecting aneurysms, with the ensemble model showcasing superior performance compared to its individual components.*

Keywords: DenseNet, DSA, hybrid model, intracranial aneurysm, VGG16

KOMPLEKSOWE BADANIE: WYKRYWANIE TĘTNIAKA WEWNĄTRZCZASZKOWEGO ZA POMOCĄ HYBRYDOWEGO GŁĘBOKIEGO UCZENIA SIĘ VGG16-DENSENET NA OBRAZACH DSA

Streszczenie. *Tętniak wewnątrzczaszkowy to obrzęk w słabym obszarze tętnicy mózgowej. Główną przyczyną tętniaka jest wysokie ciśnienie krwi, palenie tytoniu i uraz głowy. Pęknięcie tętniaka jest poważnym stanem nagłym, który może prowadzić do śpiączki, a następnie śmierci. W celu wykrycia tętniaka mózgu stosuje się cyfrową angiografię subtrakcyjną (DSA). Neurochirurg dokładnie bada skan, aby znaleźć dokładną lokalizację tętniaka. Zaproponowano model hybrydowy do dokładnego i szybkiego wykrywania tych tętniaków. Visual Geometry Group 16 (VGG16) i DenseNet to dwie architektury głębokiego uczenia wykorzystywane do klasyfikacji obrazów. Połączenie obu modeli otwiera możliwość wykorzystania różnorodności w solidnej i stabilnej ekstrakcji cech. Wyniki modelu pomagają w identyfikacji lokalizacji tętniaków, które są znacznie mniej podatne na fałszywie dodatnie lub fałszywie ujemne. Ta integracja architektury opartej na głębokim uczeniu się z praktyką medyczną jest bardzo obiecująca dla szybkiego i dokładnego wykrywania tętniaków. Badanie obejmuje 1654 obrazów DSA od różnych pacjentów, podzielonych na 70% do treningu (1157 obrazów) i 30% do testowania (496 obrazów). Złożony model wykazuje imponującą dokładność 95,38%, przewyższając odpowiednie dokładności VGG16 (94,38%) i DenseNet (93,57%). Dodatkowo, złożony model osiąga wartość pełności 0,8657, co wskazuje na jego zdolność do prawidłowej identyfikacji około 86,57% prawdziwych przypadków tętniaka spośród wszystkich rzeczywistych pozytywnych przypadków obecnych w zbiorze danych. Ponadto, biorąc pod uwagę DenseNet indywidualnie, osiąga on wartość pełności 0,8209, podczas gdy VGG16 osiąga wartość pełności 0,8642. Wartości te pokazują czułość każdego modelu w wykrywaniu tętniaków, przy czym model zespołowy wykazuje lepszą wydajność w porównaniu z jego poszczególnymi komponentami.*

Słowa kluczowe: DenseNet, DSA, model hybrydowy, tętniak wewnątrzczaszkowy, VGG16

Introduction

An enormous weak spot in the brain artery causes an intracranial aneurysm, which is a dangerous condition. Weakness of the artery walls in some areas can lead to aneurysm formation, posing a great risk to those affected. When an intracranial aneurysm ruptures, it causes a medical emergency that can lead to serious damage, including the possibility of coma and death. For prompt intervention and improved patient outcomes, intracranial aneurysms must be accurately and quickly detected. Among the major risk factors for the development of an intracranial aneurysm are head trauma, smoking, and high blood pressure. These factors can lead to wall disruption, highlighting the need to identify and monitor aneurysms in individuals with these conditions. The identification of intracranial aneurysms was made meticulously and precisely possible through the utilization of Cutting-edge imaging techniques like magnetic resonance imaging (MRI), computed tomography (CT) scans, and exceptionally accurate digital subtraction angiography (DSA). However, these methods may not always provide the accuracy or speed required to diagnose aneurysms, especially in emergencies. In recent years, the intersection of medical imaging and artificial intelligence (AI) has opened the door to better and more effective diagnoses. Deep learning-based designs attract attention due to their ability to analyse medical images accurately and quickly. Among these models, the two main options are Visual Geometry Group 16 (VGG16) and DenseNet [5], which are widely used in image classification [1]. VGG16 and DenseNet have been successful in many types of image classification. However, accurate and timely detection of intracranial aneurysms is a unique challenge that can be solved by combining the advantages of these two methods. This study leverages the power of deep learning by providing composite models that leverage diversity inference provided by VGG16 and DenseNet [8].

The main objective is to implement a model that can quickly and accurately detect intracranial aneurysms in DSA images while minimizing negative-related risks. The integration of deep learning-based architectures into the realm of medical practice offers significant potential to reshape the landscape of aneurysm diagnosis, emphasizing the importance of early detection and precise localization. This study offers a comprehensive exploration of the proposed hybrid model, delving into the methodology, evaluation, and results. Furthermore, it highlights the potential implications of this discovery in clinical practice, underscoring the significance of early detection. The ensuing sections of this study will provide detailed insights into the methods, evaluations, and outcomes, shedding light on how this hybrid model can significantly enhance the precise identification of intracranial aneurysms, particularly in DSA images, thereby advancing the field of medical imaging and patient care.

1. Literature review

Cheol Kim et al. [12] utilized the Alexnet_v2 architecture, achieving superior accuracy over human assessors in a computer-aided system. Despite this success, limitations such as a small dataset and variability from user-selected ROIs were acknowledged, signaling the need for broader studies to enhance model robustness and generalization.

Gurunathan et al. [10] proposed CNN-based models, boasting high accuracy rates of 98.6% for normal and 98% for abnormal cases. However, the challenges associated with demanding data requirements and complexities in model interpretation underscore the ongoing efforts required for optimization and practical application.

Hossain et al. [11] presented a deep CNN architecture for brain tumor detection, emphasizing its accuracy advantages. Yet, they acknowledged the complexity and resource requirements as significant drawbacks. The study's comprehensive assessment, utilizing F1 scores and balanced statistics, highlighted the need for a nuanced evaluation approach.

Stember et al. [20] employed U-net and CNN for Magnetic Resonance Angiography. However, the study revealed challenges, particularly in achieving a lower F1 score, pointing to the importance of dataset quality in model performance.

Xinke Liu et al. [14] introduced a modified UNet-based 3D CNN with excellent segmentation accuracy, particularly in handling 3D DSA pictures. Despite its strengths, challenges such as lower accuracy for smaller aneurysms were noted. The evaluation, encompassing F1 scores and error analysis, identified specific areas for improvement.

Haihan Duan et al. [7] implemented cascade CNNs for intracranial aneurysm identification, highlighting their precision but acknowledging computational complexity and extended training cycles as drawbacks. The study emphasized the critical role of the dataset in ensuring model robustness and generalization.

Chen et al. [6] introduced a system for early detection using contrast-unenhanced magnetic resonance angiography, showcasing high sensitivity and low false positive rates. This suggests practical applicability for automated screening in routine examinations, addressing the need for efficient identification of suspected aneurysm areas.

Takahiro Nakao et al. [17] proposed a deep neural network model for diagnosing cerebral aneurysms, emphasizing the high accuracy achieved. However, challenges such as the model's limitations in detecting tiny aneurysms underscore the importance of broader datasets for increased generalizability and therapeutic effectiveness.

Yuan et al. [22] introduced DCAU-Net for segmenting intracranial aneurysms, showcasing superior accuracy and efficiency. The model's utilization of dense blocks and Convolution Block Attention Module (CBAM) contributed to enhanced segmentation performance. Comparative experiments highlighted its advantages over other models, emphasizing its potential in clinical applications.

Jun Hyong Ahn et al. [2] introduced a Multiview CNN-ResNet50 model for classifying rupture risk in tiny intracranial aneurysms. The algorithm demonstrated high sensitivity, specificity, accuracy, and F1 score, suggesting its utility in assessing rupture likelihood.

Daiju Ueda, MD et al. [21] developed a deep-learning framework for detecting brain aneurysms using time-of-flight MRI, achieving improved detection rates. The focus on avoiding misses, though resulting in lower specificity, prompted suggestions for further improvements, including incorporating additional imaging data for comprehensive evaluations.

Agus Eko Minarno et al. [15] applied SVM and DenseNet to MRI images for brain tumor categorization and achieved an impressive 99.65% accuracy. This study underscores the effectiveness of deep learning strategies in accurately identifying brain tumor categories.

Bincy Chellapandi et al. [5] discussed plant disease detection using a self-built CNN model and achieved superior performance with over 93% accuracy. This study also explored the application of transfer learning to enhance the accuracy of plant disorder detection.

Zeng et al. [23] proposed a method for enhancing detection accuracy using successive image fusion (SIF), achieving an impressive 98.89% accuracy with 3D-RA projection images. The study demonstrated the robustness and effectiveness of the proposed framework.

Belaid et al. [4] presented a model that combines VGG16 CNN and GLCM features for brain tumor classification and achieved a 96% accuracy. This study highlighted the potential of this approach for accurate tumor categorization.

Sourodip et al. [9] utilized an enhanced U-Net architecture with VGG-16 for brain tumor categorization via MRI, outperforming basic U-Net and other CNNs with pixel accuracies of 0.997. The study highlighted the superior performance of the enhanced U-Net architecture.

Wufeng Liu et al. [13] achieved a 96% accuracy rate for rice leaf disease identification using a hybrid neural network model with enhanced attention. The study demonstrated the model's efficiency in identifying and classifying diverse types of rice leaf diseases.

Yufan Zhou et al. [24] proposed a method for diagnosing brain tumors using composite 3D MR images, achieving accurate tumor diagnosis without detailed annotations. The study emphasized the potential of dense convolutional neural networks and recurrent neural networks for accurate diagnosis and categorization.

Martinson Ofori et al. [18] presented an approach for reducing DL model complexity in precision agriculture systems by combining ensemble learning, model compression, and transfer learning for sustainability and cost-effectiveness. This study addressed the limitations of DL models in precision agriculture.

Muhammad Mujahid et al. [16] utilized machine learning and signal processing for early-stage Alzheimer's disease detection, achieving promising accuracies of 97.35% and 99.64%, respectively, in terms of the area under the curve (AUC). The study explored different datasets and techniques, demonstrating the potential for improving the accuracy of diagnosing Alzheimer's disease.

Rahil Shahzad et al. [19] employed a deep learning system to identify active aneurysms in aSAH patients via CT angiography; the system was effective and had high sensitivity for detecting aneurysms of unusual sizes. The study demonstrated the model's accuracy in detecting and classifying large arteries.

Al Okashis et al. [3] developed an autonomous model for hemorrhage detection via MRI, achieving 89.2% accuracy and 100% sensitivity. The study evaluated the impacted brain area by applying segmentation-based feature extraction for efficient detection.

2. Proposed methodology

The intended system architecture strives to guarantee accurate and fast detection of intracranial aneurysms. The system is designed by ensembling two widely used pre-trained models, VGG16 (Visual Geometry Group 16) and DenseNet, which are commonly employed for image classification. By concatenating the outputs of both models, the system leverages feature extraction from both networks. Fig.1 illustrates the process flow diagram of the methodology. Integrating these advanced CNN models for analyzing medical images results in robust, error-free, and timely detection of aneurysms in the intracranial section.

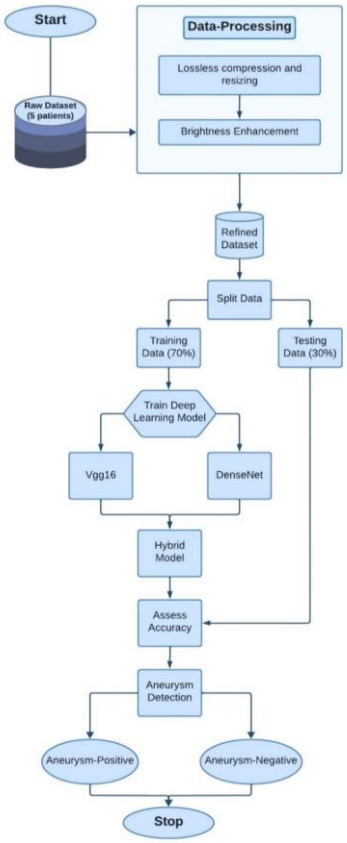


Fig. 1. Process flow model

2.1. Raw dataset

The dataset utilized in this study was gathered from various hospitals. It contains images of the Vascular system of the intracranial section of the Brain. These images are captured using the Digital Subtraction Angiographic Technique which incorporates advanced X-ray Technology, Contrast dye, and monitoring equipment that ensures safe imaging of Brain Vascularity. The dataset contains 1654 DSA images composed from the unique cases of 5 patients. The dataset consists of two classes: the positive class, comprising 446 images containing aneurysms, and the negative class, comprising 1208 images that do not contain aneurysms. Fig. 2 represents the sample images of raw dataset.

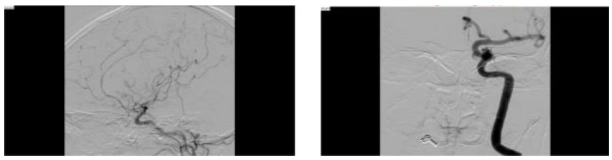


Fig. 2. Raw dataset

2.2. Data preprocessing

The collected DSA images have varying dimensions, contrasts, and levels of sharpness. To standardize the dataset, it has undergone two preprocessing steps.

2.2.1. Lossless compression and resizing

The initial step involves standardizing all the images to a uniform size. Within this dataset, the images include details such as scan layer numbers, which must be eliminated through cropping. After the cropping process, all the images are resized to dimensions of 512×512 using the Lanczos filter. The Lanczos filter is employed for resizing the images to the specified dimensions while also preserving image quality and minimizing aliasing effects. This is achieved through a weighted convolution filter that calculates new pixel values by considering a limited number of neighboring pixel values, resulting in smoother images.

2.2.2. Brightness enhancement

The DSA images exhibit low contrast and variations in brightness due to the different diameters of blood vessels. Additionally, smaller structures are hardly visible in these images. These issues can be addressed by applying a CLAHE (Contrast Limited Adaptive Histogram Equalization) filter. An excessive brightness ratio of 1.2 has been employed to overcome this problem, enhancing visibility without over-amplifying the blood vessels. This technique improves contrast and visibility by dividing the image into tiny blocks and equalizing the histogram of each block. Fig. 3 represents a sample of refined dataset.

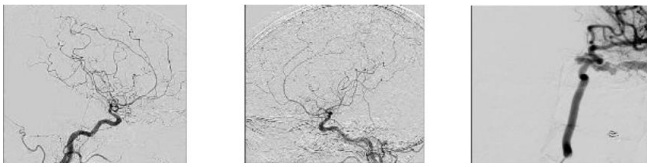


Fig. 3. Processed images

2.3. Split data

The preprocessed and refined images have been split into training data and testing data. In total, there are 1,654 images. Out of these, 70% of the data, which amounts to 1,158 images, is designated as the training dataset, while the remaining 30%, comprising 496 images, is assigned to the test dataset. This division ensures a robust evaluation of the model's performance and allows for effective training on a diverse range of examples. By carefully partitioning the dataset in this manner, we aim to optimize the model's ability to generalize to unseen data, thereby enhancing its overall efficacy in real-world applications.

2.4. Train deep learning models

To assess the accuracy of various deep learning models, multiple models were trained with the same dataset to compare their performance and select the one with the highest accuracy. The preprocessed data was trained by choosing diverse model architectures. The trained models were evaluated on the test data by comparing key metrics. After comparing the results, the model with higher performance is selected. This rigorous evaluation process ensures that the chosen model not only performs well on the training data but also generalizes effectively to unseen test data, enhancing its reliability for real-world applications. By systematically testing different architectures and comparing their performance, we aim to identify the most suitable model for the given task, thereby maximizing the accuracy and efficiency of our deep learning system.

2.4.1. DenseNet

The data was trained utilizing a pre-existing Deep Learning model DenseNet-121 [8]. To initiate model training, the pre-trained DenseNet-121 model was loaded with pre-trained weights from the ImageNet dataset. Data augmentation was implemented through the `ImageDataGenerator`, introduces variations to the training images. Techniques such as rotation, shifting, shearing, zooming, and flipping, coupled with rescaling, contribute to a more diverse dataset. The setup involved excluding the upper classification layers and configuring it to accept input images with specific dimensions. The layers of the DenseNet-121 model were locked, ensuring that they would not be modified during the training process. Custom classification layers were added above the pre-trained model. To reduce the spatial dimensions of the feature maps, GlobalAveragePooling2D was applied. Following that, dense layers incorporated with Rectified Linear Unit (ReLU) Trigger purposes, and a final dense layer with a sigmoid activation function was added to produce binary classification results. The model was compiled using the Adam optimizer employing a learning rate of 0.0001. For binary classification, the binary cross entropy loss function

was utilized, and precision was chosen as the evaluation metric. After completing the training process, the model demonstrated a success rate of 93.57%.

Algorithm 1: Image Classification using DenseNet121.

Input: train_ dir: Training data directory; test_dir:Testing data directory;img_width, img_height:Image dimensions; batch_size: Batch size;
Output: Trained DenseNet121-based model;
Step 1: Import Libraries;Load TensorFlow and the required modules;
Step 2: Data and Parameters:Specify the data directories, image parameters, and batch size;
Step 3: Data Augmentation;Prepare the data generators with data augmentation;
Step 4: Load the DenseNet121 Model;The DenseNet121 model is loaded with pretrained weights;
Step 5: Customize the top layers;Add custom classification layers;
Step 6: Compile the model;Compile the model with the specified optimizer and loss function;
Step 7: Callbacks;Define the model checkpoint and early stopping callbacks;
Step 8: Train Model;Train the model with specified settings;
Step 9: Assess Model;The model's performance was assessed on the test dataset;

2.4.2. VGG16

The data were trained utilizing a pre-existing deep learning model VGG16. The training process commenced by loading the pretrained VGG16 model with weights initially trained on the ImageNet dataset. Data augmentation was implemented through the `ImageDataGenerator`, introduces variations to the training images. Techniques such as rotation, shifting, shearing, zooming, and flipping, coupled with rescaling, contribute to a more diverse dataset. The model's configuration excluded the top classification layers and allowed input images with specified dimensions. The layers of the VGG16 model were frozen to prevent updates during training, and the model deliberately excluded the uppermost classification layers. Additional custom classification layers were placed on the pretrained model. A flattening layer was applied to the base model's output. Subsequently, two dense layers with ReLU activation functions were introduced, and a final dense layer with a sigmoid activation function was included to produce binary classification results. The model was constructed with the Adam optimizer using a designated learning rate of 0.0001. Binary cross entropy served as the loss function for binary classification, and accuracy was chosen as the evaluation metric. Following the training process, the model achieved an accuracy of 94.38%.

Algorithm 2: Image Classification Using VGG16

Input: train_ dir: Training data directory; test_dir:Testing data directory;img_width, img_height:Image dimensions; batch_size: Batch size;
Output: Trained VGG16-based model;
Step 1: Import Libraries;Load necessary TensorFlow libraries;
Step 2: Data and Parameters:Specify the data directories and image parameters;
Step 3: Data Augmentation;Prepare the data generators with augmentation;
Step 4: Load the VGG16 Model;The VGG16 model is loaded with pretrained weights;
Step 5: Customize the top layers; add custom classification layers;
Step 6: Compile the model;Compile the model with the specified optimizer and loss function;
Step 7: Callbacks;Define the model checkpoint and early stopping callbacks;
Step 8: Train Model;Train the model with specified settings;
Step 9: Assess Model;The model's performance was assessed on the test dataset;

2.4.3. Hybrid model

After training the two models, DenseNet121 and VGG16, respectively, achieved accuracies of 93.17% and 92.32%. To create a hybrid model, both pretrained models, DenseNet121 and VGG16, are loaded with weights from the ImageNet dataset. Both models are connected to exclude their top classification layers and to accept input from the previously defined input layer. For transfer learning, all the layers of both models are frozen so that they cannot be changed during the training process. Global average pooling layers are added to both the DenseNet121 and VGG16 models [22], which diminishes the spatial dimensions of the feature maps. Now, the outputs of the global average pooling of both models are concatenated into a single 1×1 tensor. Two custom dense layers are added to it: the first is ReLU, with 256 units. Each unit computes a weighted sum of input features, allowing nonlinearity in the model and the ability to learn complex relationships. The final layer is the sigmoid layer with a single unit, which is typically used for binary classification tasks. The hybrid model is prepared by using pretrained models as input layers and custom dense layers as output layers. The model is assembled in combination with. The Adam optimizer, incorporating a designated learning rate of 0.0001. The accuracy of the hybrid model was evaluated. It achieved an accuracy of 95.38%.

Algorithm3: CombinedModel: DenseNet121 and VGG16

Input: train_ dir: Train datadir; test_dir: Test data dir;img_width,img_height:Img dimensions; batch_ size: Batch size;
Output: Combined model for binary classification;
Step 1: Import Libraries; Load TensorFlow and modules;
Step 2: Define the input and layers; define the input shape and create layers;
Step 3: Load the pretrained models;
The DenseNet121 and VGG16 models are loaded with pretrained weights;
Step 4: Freeze layers: Freeze layers in both models;
Step 5: Combine the outputs;
Combined_output = Concatenate ([GlobalAvgPool (DenseNet121 output), GlobalAvgPool (VGG16 output)])
Step 6: Custom dense layers;
x = Dense (combined output, 256*;* activation =*′ relu′*)output = Dense (*x*, 1*;* activation =*′ sigmoid′*);
Step 7: Create the combined model; Create the combined model with input and output;
Step 8: Compile the model; Compile the model with the specified optimizer and loss function;
Step 9: Data generators; Define the data generators for the training and test data;
Step 10: ModelCheckpoint; Define a ModelCheckpoint to save the best model;
Step 11: Train and evaluate; Train the model with data generators and evaluate the best model**;**

3. Results and comparitive analysis

The dedicated test dataset is subsequently used to assess the trained models. This stage facilitates a detailed evaluation of each model's performance. Various crucial performance indicators, such as the ROC-AUC, precision, accuracy, F1 score, and recall curve, were computed to provide a comprehensive assessment of each model's capabilities. Analyzing these metrics helps identify the model that excels in our specific task, simplifying the decision-making process for further deployment or analysis. Table 1 shows the comparison of the various models on the above-mentioned parameters. By meticulously examining these metrics, we gain insights into the strengths and weaknesses of each model, enabling us to make informed decisions regarding their suitability for practical implementation in real-world scenarios.

Table 1. Performance metrics

Model	Accuracy	Precision	Recall	F1 SCORE	ROC-AUC
Hybrid	0.9538	0.9587	0.8657	0.9098	0.9712
DenseNet	0.9357	0.9322	0.8209	0.8730	0.9675
VGG16	0.9438	0.9206	0.8642	0.8923	0.9650

Figures 4a, b, and c show the confusion matrices, which represent the performances of VGG16, DenseNet, and the hybrid model, respectively. These matrices offer insights into the models' classification accuracy by illustrating the distribution of true

positive, true negative, false positive, and false negative predictions across different classes. Additionally, in figures 5a, b, and c, the ROC-AUC curves are shown for VGG16, DenseNet, and the hybrid model, respectively. These curves provide a graphical representation of the models' performance in terms of true positive rate against false positive rate across different threshold values. Analyzing both confusion matrices and ROC-AUC curves aids in understanding the strengths and weaknesses of each model's classification capabilities, facilitating informed decision-making for model selection and further optimization.

Table 2 illustrates a comparison of the accuracies of architectures proposed in various studies previously discussed. Upon examination, it is evident that the ensembled model of VGG16 and DenseNet exhibits higher accuracy compared to the others. The hybrid model benefits from the complementary strengths of both VGG16 and DenseNet, leveraging VGG16's robust feature extraction capabilities and DenseNet's dense connectivity to achieve superior performance.

By combining the strengths of these two architectures, the hybrid model achieves enhanced accuracy, demonstrating the effectiveness of ensemble methods in deep learning.

By combining the strengths of these two architectures, the hybrid model achieves enhanced accuracy, demonstrating the effectiveness of ensemble methods in deep learning. Ensemble methods leverage the diverse perspectives and capabilities of multiple models to create a more robust and reliable system. In this case, integrating features from both DenseNet and VGG16 allows the hybrid model to capture a wider range of image characteristics and patterns, leading to improved performance in detecting intracranial aneurysms. This approach not only increases accuracy but also enhances the model's resilience to noise and variability in the data, making it more suitable for real-world applications where robustness is critical.

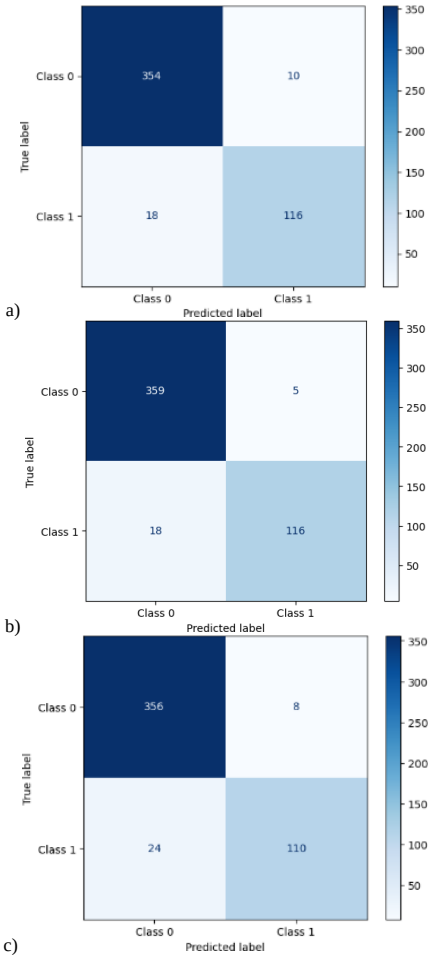


Fig. 4. Confusion matrices of: a) DenseNet, b)VGG16, c) hybrid model

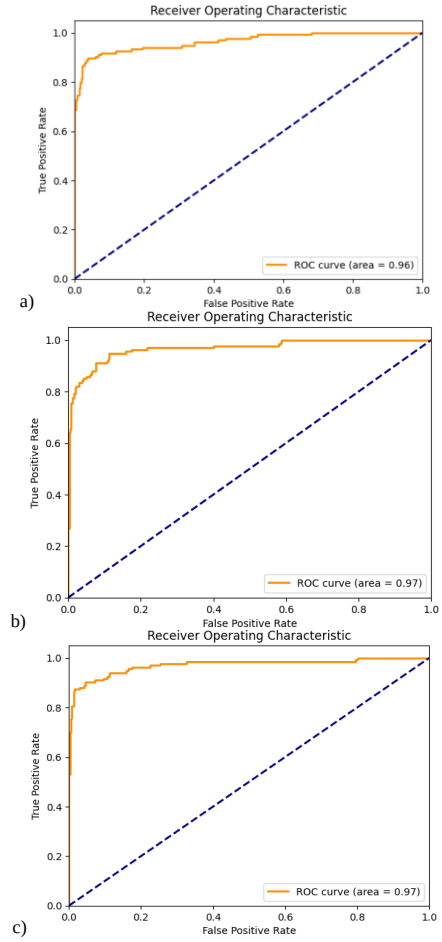


Fig. 5. ROC-AUC Curves of: a) DenseNet, b) VGG16, c) hybrid model

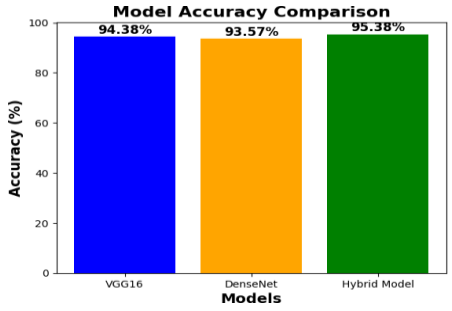


Fig. 6. Bar Graph representing accuracy comparison

Table 2. Comparative analysis

Reference number	Architecture	Accuracy
[18]	Ensembled model (VGG, Resnet, DenseNet)	91.2%
[2]	A multi-view CNN-ResNet50	81.72%
[12]	CNN	76.84%
[14]	CNN and MIP	94.2%
[7]	Cascade CNN	93.5
	Ensembled Model (VGG16, DenseNet)	95.38%

4. Conclusion

DenseNet and VGG16 stand out as prominent deep learning models widely employed for image classification, with this study focusing on the detection of intracranial aneurysms in DSA images. Remarkably, the DenseNet and VGG16 models showcased outstanding performance, achieving accuracies of 93.57% and 94.38%, respectively. DenseNet's proficiency in feature extraction, facilitated by its deep network architecture, and VGG16's high efficiency due to its straightforward design contributed to their respective successes. Specifically, for recall values, VGG16 demonstrated a recall of 0.8642, while DenseNet exhibited a slightly lower but still commendable recall of 0.8209. Combining the strengths of both models in a hybrid approach

yielded a significantly higher accuracy of 95.38%. Notably, the hybrid model, with a recall value of 0.8657, not only surpassed individual model performances but also showcased enhanced robustness and superior generalizability to previously unseen data. In conclusion, the hybrid model successfully fulfil all the objectives of detecting aneurysms in DSA images with remarkable efficiency, demonstrating the synergistic power of leveraging diverse deep learning architectures. The ensembled model, by combining the strengths of DenseNet and VGG16, achieves a level of performance that exceeds what either model can achieve individually. Ensembling is a useful technique for improving model performance by leveraging the complementary strengths of multiple models, in this case, enhancing feature extraction and detection sensitivity for more effective identification of intracranial aneurysms in DSA images.

References

[1] Ahmed F. et al.: Identification and Prediction of Brain Tumor Using VGG-16 Empowered with Explainable Artificial Intelligence. International Journal of Computational and Innovative Sciences 2(2), 2023, 24–33.

[2] Ahn J. H. et al.: Multi-view convolutional neural networks in rupture risk assessment of small, unruptured intracranial aneurysms. Journal of Personalized Medicine 11(4), 2021, 239.

[3] Al Okashi O. M. et al.: An ensemble learning approach for automatic brain hemorrhage detection from MRIs. 12th International Conference on Developments in eSystems Engineering – DeSE, IEEE, 2019.

[4] Belaid O. N., Loudini M.: Classification of brain tumor by combination of pre-trained VGG16 CNN. Journal of Information Technology Management 12(2), 2020, 13–25.

[5] Chellapandi B., Vijayalakshmi M., Chopra S.: Comparison of pre-trained models using transfer learning for detecting plant disease. International Conference on Computing, Communication, and Intelligent Systems – ICCCIS, IEEE, 2021.

[6] Chen G. et al.: Automated computer-assisted detection system for cerebral aneurysms in time-of-flight magnetic resonance angiography using fully convolutional network. BioMedical Engineering OnLine 19(1), 2020, 1–10.

[7] Duan H. et al.: Automatic detection on intracranial aneurysm from digital subtraction angiography with cascade convolutional neural networks. Biomedical Engineering Online 18, 2019, 1–18.

[8] Ghaleb Al-Mekhlafi Z. et al.: Hybrid Techniques for Diagnosing Endoscopy Images for Early Detection of Gastrointestinal Disease Based on Fusion Features. International Journal of Intelligent Systems 2023, 8616939.

[9] Ghosh S., Chaki A., Santosh K. C.: Improved U-Net architecture with VGG-16 for brain tumor segmentation. Physical and Engineering Sciences in Medicine 44(3), 2021, 703–712.

[10] Gurunathan A., Krishnan B.: Detection and diagnosis of brain tumors using deep learning convolutional neural networks. International Journal of Imaging Systems and Technology 31(3), 2021, 1174–1184.

[11] Hossain T. et al.: Brain tumor detection using convolutional neural network. 1st international conference on advances in science, engineering and robotics technology – ICASERT, IEEE, 2019, 1–6.

[12] Kim H. C. et al.: Machine learning application for rupture risk assessment in small-sized intracranial aneurysm. Journal of Clinical Medicine 8(5), 2019, 683.

[13] Liu W., Yu L., Luo J.: A hybrid attention-enhanced DenseNet neural network model based on improved U-Net for rice leaf disease identification. Frontiers in Plant Science 13, 2022, 922809.

[14] Liu X. et al.: Deep neural network-based detection and segmentation of intracranial aneurysms on 3D rotational DSA. Interventional Neuroradiology 27(5), 2021, 648–657.

[15] Minarno A. E. et al.: Classification of Brain Tumors on MRI Images Using DenseNet and Support Vector Machine. JOIV: International Journal on Informatics Visualization 6(2), 2022, 404–410.

[16] Mujahid M. et al.: An Efficient Ensemble Approach for Alzheimer's Disease Detection Using an Adaptive Synthetic Technique and Deep Learning. Diagnostics 13(15), 2023, 2489.

[17] Nakao T. et al.: Deep neural network-based computer-assisted detection of cerebral aneurysms in MR angiography. Journal of Magnetic Resonance Imaging 47(4), 2018, 948–953.

[18] Ofori M.: Transfer-Learned Pruned Deep Convolutional Neural Networks for Efficient Plant Classification in Resource-Constrained Environments. Masters Theses & Doctoral Dissertations 371, 2021.

[19] Shahzad R. et al.: Fully automated detection and segmentation of intracranial aneurysms in subarachnoid hemorrhage on CTA using deep learning. Scientific Reports 10(1), 2020, 21799.

[20] Stember J. N. et al.: Convolutional neural networks for the detection and measurement of cerebral aneurysms on magnetic resonance angiography. Journal of Digital Imaging 32, 2019, 808–815.

[21] Ueda D. et al.: Deep learning for MR angiography: automated detection of cerebral aneurysms. Radiology 290(1), 2019, 187–194.

[22] Yuan W. et al.: DCAU-Net: Dense convolutional attention U-Net for segmentation of intracranial aneurysm images. Visual Computing for Industry, Biomedicine, and Art 5(1), 2022, 1–18.

[23] Zeng Y. et al.: Automatic diagnosis based on spatial information fusion feature for intracranial aneurysm. IEEE Transactions on Medical Imaging 39(5), 2019, 1448–1458.

[24] Zhou Y. et al.: Holistic brain tumor screening and classification based on DenseNet and recurrent neural network. Crimi A., Bakas S., Kuijf H., Keyvan F., Reyes M., van Walsum T. (eds): Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2018. Lecture Notes in Computer Science 11383. Springer, Cham [https://doi.org/10.1007/978-3-030-11723-8_21].

Prof. Sobhana Mummaneni
e-mail: sobhana@vrsiddhartha.ac.in

Dr. Sobhana Mummaneni is currently working as an associate professor in the Department of Computer Science and Engineering, V. R. Siddhartha Engineering College, Vijayawada, India. She received Ph.D. degree in Computer Science and Engineering in 2018 from Krishna University. She has 16 years of teaching experience. She published 35 papers in National and International journals and published 7 patents.
Her research interests lie in areas such as artificial intelligence, machine learning, data analytics, cyber security, and software engineering.

http://orcid.org/0000-0001-5938-5740

Eng. Sasi Tilak Ravi
e-mail: rsasitilak0987@gmail.com

Sasi Tilak Ravi is a third-year B.Tech. student specializing in Computer Science and Engineering at V. R Siddhartha Engineering College. He is currently Marketing Lead for Google Developer Student Club and has organized various technical events.
He is passonate about deep learning and artificial intelligence. He is interested in developing various machine learning projects

https://orcid.org/0009-0005-3342-2984

Eng. Jashwanth Bodedla
e-mail: jashwanthchowdary333@gmail.com

Jashwanth Bodedla is a third-year B.Tech. student specializing in Computer Science and Engineering at V. R. Siddhartha Engineering College.
He is interested in cloud computing and machine learning.

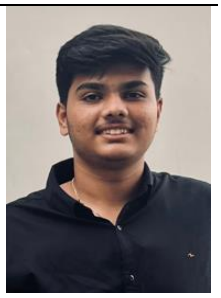
https://orcid.org/0009-0008-6654-1076

Eng. Sree Ram Vemulapalli
e-mail: vemulapallisreeram19@gmail.com

Sree Ram Vemulapalli is a third-year B.Tech. student specializing in Computer Science and Engineering at V. R. Siddhartha Engineering College.
He is passionate about Full Stack Web Development. He is interested in creating various websites.

https://orcid.org/0009-0000-1916-4433

Eng. Gnana Sri Kowshik Varma Jagathapurao
e-mail: kowshikjkvarma@gmail.com



Gnana Sri Kowshik Varma Jagathapurao is a third-year B.Tech. student specializing in Computer Science and Engineering at V. R. Siddhartha Engineering College.
He is passionate in cyber security. He loves to explore various AI Tools.

https://orcid.org/0009-0009-9684-6994

http://doi.org/10.35784/iapgos.5824 received: 22.01.2024 | revised: 16.02.2024 | accepted: 26.02.2024 | available online: 31.03.2024

DEFORMATIONS OF SOIL MASSES UNDER THE ACTION OF HUMAN-INDUCED FACTORS

Mykola Kuzlo[1], Viktor Moshynskyi[2], Nataliia Zhukovska[3], Viktor Zhukovskyy[3]
[1]National University of Water and Environmental Engineering, Department of Highways, Bases and Foundations, Rivne, Ukraine,
[2]National University of Water and Environmental Engineering, Department of Land Management, Cadastre, Land Monitoring and Geoinformatics, Rivne, Ukraine,
[3]National University of Water and Environmental Engineering, Department of Computer Sciences and Applied Mathematics, Rivne, Ukraine

Abstract. *Significant changes in the stress-strain state cause a change in the soil profile of the massif, which is affected by various physical and chemical factors. In particular, groundwater filtration, mass transfer, heat transfer, dissolution and leaching of soil masses. This can lead to various types of accidents. Therefore, the study of the stress-strain state of the soil massif is an important topic. Nonlinear dependences in the form of polynomials of the modulus of deformation and Lame coefficients on the concentration of salt solutions and their temperature have been received in this research based on experimental research and their statistical processing. This allowed improving the mathematical model of the stress-strain state of the soil taking into account the nonlinear deformation processes occurring in the soil masses under the presence and filtration of saline solutions in non-isothermal conditions.*

Keywords: mathematical models, statistical analysis, stress-strain state, deformation

DEFORMACJE MASY GLEBY POD DZIAŁANIEM CZYNNIKÓW CZŁOWIEKA

Streszczenie. *Znaczące zmiany stanu naprężeniowo-odkształceniowego powodują zmianę profilu glebowego masywu, na który wpływają różne czynniki fizyczne i chemiczne. W szczególności filtracja wód gruntowych, przenikanie masy, przenoszenie ciepła, rozpuszczanie i wymywanie mas glebowych. Może to prowadzić do różnego rodzaju wypadków. Dlatego ważnym tematem jest badanie stanu naprężeniowo-odkształceniowego masywu glebowego. W niniejszych badaniach otrzymano nieliniowe zależności w postaci wielomianów modułu odkształcenia i współczynników Lame'a od stężenia roztworów soli i ich temperatury w oparciu o badania eksperymentalne i ich obróbkę statystyczną. Pozwoliło to na udoskonalenie modelu matematycznego stanu naprężeniowo-odkształceniowego gruntu uwzględniającego nieliniowe procesy odkształceń zachodzących w masach gruntu pod wpływem obecności i filtracji roztworów soli w warunkach nieizotermicznych.*

Słowa kluczowe: modele matematyczne, analiza statystyczna, stan naprężenia-odkształcenia, odkształcenie

Introduction

Human activity leads to changes in hydrogeological conditions [5, 12]. And the change in hydrogeological conditions on Earth is accelerating. The reason for such changes is the excessive strengthening of a previously unknown geological agent. This new geological agent is human activity. In the research paper [19], human technical activity is considered a driving geological force that not only changes the face of the earth's surface but also makes significant changes in the upper part of the earth's crust, which is comparable to scale and consequences with geological processes.

The problem is greatly complicated if the soil is contaminated with various solutions of industrial facilities' slurry ponds, which accumulate industrial waste. These objects often reach quite large sizes, and they constantly or periodically get a large amount of so-called "spent" liquids used in technological processes, which are characterized by high mineralization and temperature.

Currently, the data have been used which are developed for soil mechanics to assess the soil mass deformation which is saturated with saline solutions [8, 9]. And this is the subject to the action of low-concentration natural groundwater [4].

Also some researchers are working on geotechnical issues of landslides in Ukraine [11] and decreasing service life of buildings under the regular explosion loads [10]. At the same time others proved the theory that mining activities always have and exclusively caused considerable changes in the environment [7]. Moreover, the rapid growth of population, urban planning, agricultural and industrial sectors also may have a significant effect [2].

Scientists pay attention also to asphalt mixes. In paper [3] they developed mathematical models for not isothermal conditions of marginal soils deformation (foamed and emulsified sulfur asphalt soils mixes) using specific software. Paper [1] provides a new numerical model in discrete element analysis for simulating flow time and number tests of asphalt mixes.

However, our previous experiments have shown, the compressibility of soils subject to the action of saline solutions depends on many factors. It depends from the degree of load and the concentration of saline solutions and their temperature as well [13, 14, 17].

Human-induced impact on soil masses leads to various factors: hydrodynamic forces of the filtration flow, changes in the filtration and deformation characteristics of the soil and so on. The change of these factors can be important, which leads to significant deformations of the earth's surface [6, 21, 23]. These deformations complicate regular operation and, in some cases, lead to accidents of industrial and energetic objects and can cause significant economic damage and even victims. In difficult military times, the issue of the stress-deformation state of soil massifs also acquires new relevance. This especially applies to areas of hostilities affected by massive artillery strikes.

The work aims to study the influence of the concentration of salt solutions and their temperature on the deformation processes of the soil and to create on this basis a computational-theoretical apparatus for forecasting the deformations of soil masses.

1. Physical experiments

To establish the influence of solutions' concentration and their temperature on the deformation characteristics of the soil, experimental studies were conducted in the geotechnical laboratory of the National University of Water and Environmental Engineering (Rivne, Ukraine).

The experiments were performed on a compression-filtration device according to state standard methods [18]. Issued in January 1991 in Kyiv, this standard outlines the specific methods to accurately assess these soil properties, which are crucial for construction and engineering purposes. The standard ensures that soil testing is conducted consistently and reliably, providing essential data for evaluating soil behavior under various conditions, which is vital for the design and safety assessment of construction projects.

Soil samples of disturbed structures with specified values of density and humidity were used for analysis. The soils for the study were sandy clays with the number of plasticity I_p = 7.0%, the porosity coefficient e = 0.55. Soil pastes for the experiment had been saturating with NaCl saline solutions' concentration 0; 22; 44; 65; 90; 110; 130; 145; 165 g/l for 2 days. The temperature value was taken in the range from 22°C to 88°C.

Results of experimental data have been shown in tables 1–3.

Table 1. The value of the deformation modulus E(c, T)

c, g/l	Deformation modulus $E(c, T)$, kPa			
	T = 22°C	T = 31°C	T = 60°C	T = 88°C
0	6555	6003	5692	5117
22	5520	5290	5175	4743
44	5002	4887	4738	4427
65	5060	4773	4427	4140
90	5164	4876	4485	4025
110	5348	4922	4571	4111
130	5474	5060	4715	4312
145	5681	5232	4916	4571
165	6038	5462	5175	4830

Table 2. The value of Lame parameter λ(c, T)

c, g/l	Lame parameter $\lambda(c, T)$, kPa			
	T = 22°C	T = 31°C	T = 60°C	T = 88°C
0	5663	5187	4918	4422
22	4770	4571	4471	4098
44	4322	4222	4094	3826
65	4372	4124	3826	3577
90	4462	4213	3875	3478
110	4620	4252	3950	3552
130	4729	4372	4074	3726
145	4909	4521	4248	3950
165	5217	4719	4471	4173

Table 3. The value of Lame parameter μ(c, T)

c, g/l	Lame parameter $\mu(c, T)$, kPa			
	T = 22°C	T = 31°C	T = 60°C	T = 88°C
0	2427	2222	2107	2107
22	2043	1959	1915	1756
44	1852	1810	1754	1638
65	1874	1767	1638	1532
90	1912	1805	1660	1490
110	1980	1822	1692	1522
130	2026	1874	1745	1596
145	2104	1937	1820	1692
165	2235	2022	1915	1788

2. Results and discussion

The following dependences of the Young's deformation modulus and Lame parameters on the concentration of saline solutions and temperature were obtained as a result of experimental data statistical mathematical processing:

$$E(c,T) = (a_{18} \cdot c^2 + a_{17} \cdot c + a_{16}) \cdot T^2 + \\ +(a_{15} \cdot c^2 + a_{14} \cdot c + a_{13}) \cdot T + a_{12} \cdot c^2 + a_{11} \cdot c + a_{10} \quad (1)$$

where

$a_{18} = 1.179 \cdot 10^{-5}$, $a_{17} = -7.755 \cdot 10^{-4}$, $a_{16} = 0.024$, $a_{15} = -1.713 \cdot 10^{-3}$, $a_{14} = 0.138$, $a_{13} = -17.539$, $a_{12} = 0.198$, $a_{11} = -30.741$, $a_{10} = 6.546 \cdot 10^{3}$;

$$\lambda(c,T) = (a_{28} \cdot c^2 + a_{27} \cdot c + a_{26}) \cdot T^2 + \\ +(a_{25} \cdot c^2 + a_{24} \cdot c + a_{23}) \cdot T + a_{22} \cdot c^2 + a_{21} \cdot c + a_{20} \quad (2)$$

where

$a_{28} = 1.177 \cdot 10^{-5}$, $a_{27} = -9.23 \cdot 10^{-4}$, $a_{26} = 0.018$, $a_{25} = -1.558 \cdot 10^{-3}$, $a_{24} = 0.128$, $a_{23} = -14.343$, $a_{22} = 0.166$, $a_{21} = -25.543$, $a_{20} = 5.611 \cdot 10^{3}$;

$$\mu(c,T) = (a_{38} \cdot c^2 + a_{37} \cdot c + a_{36}) \cdot T^2 + \\ +(a_{35} \cdot c^2 + a_{34} \cdot c + a_{33}) \cdot T + a_{32} \cdot c^2 + a_{31} \cdot c + a_{30} \quad (3)$$

where

$a_{38} = 1.242 \cdot 10^{-5}$, $a_{37} = -1.983 \cdot 10^{-3}$, $a_{36} = 0.084$, $a_{35} = -1.3 \cdot 10^{-3}$, $a_{34} = 0.191$, $a_{33} = -12.748$, $a_{32} = 0.085$, $a_{31} = -13.816$, $a_{30} = 2.533 \cdot 10^{3}$.

The results of experimental data mathematical processing with the choice of the optimal scale have been shown in figures 1–3.

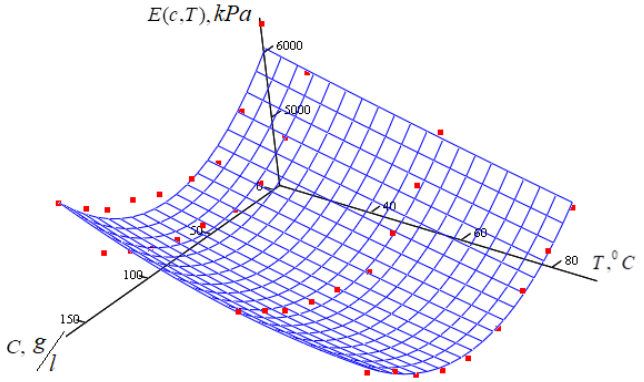


Fig. 1. The graphic of Young's deformation modulus E(c, T), kPa on pollution concentration and temperature

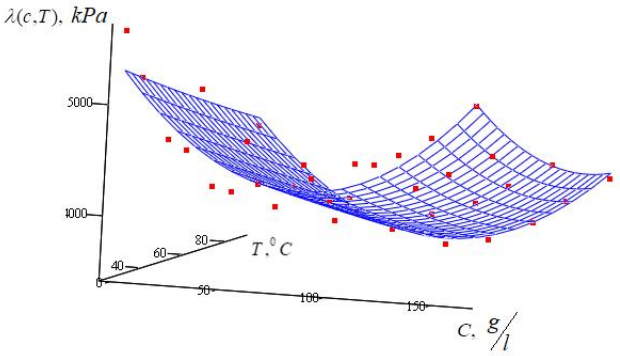


Fig. 2. The graphic of Lame's parameter dependency λ(c, T), kPa on pollution concentration and temperature

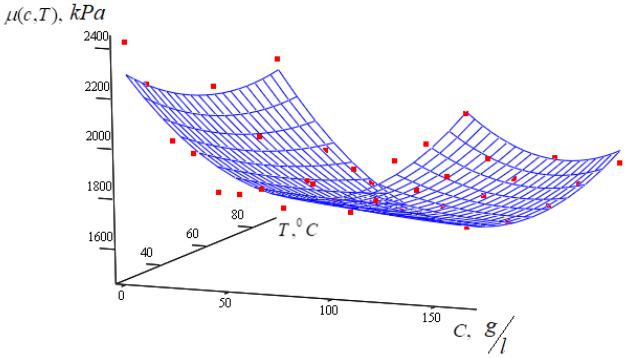


Fig. 3. The graphic of Lame's parameter dependency μ(c, T), kPa on pollution concentration and temperature

The established dependences of the Lame coefficients $\lambda(c, T)$ and $\mu(c, T)$ on the filtration solution concentration and its temperature are in the area of scientific and practical interest. It can be further used in the creation of underground hydromechanics mathematical models and estimation of the stress-deformed state of soil environments. Accordingly, a new scientific field of these dependencies is emerging in new and updated mathematical models.

2.1. Mathematical model

Let us consider a one-dimensional problem for determining the vertical displacements of the soil layer based on the problem complexity. The soil mass thickness l has been given, in which the processes of heat transfer and mass transfer of contaminated substances take place (Fig. 4).

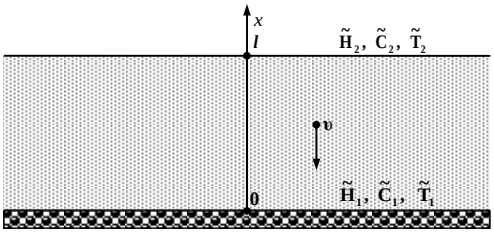


Fig. 4. Soil mass scheme in one-dimensional case

On Fig. 4. there is the scheme of soil massif in the one-dimensional case under the existence of pressures $\tilde{H}_2 > \tilde{H}_1$, the concentration of contaminated substances $\tilde{C}_2 > \tilde{C}_1$ and temperature $\tilde{T}_2 > \tilde{T}_1$ in the upper and lower ones of its limits.

The mathematical model of the one-dimensional problem of the stress-deformed state in the layer and soil mass thickness l with the spread of contaminated substances and nonisothermal conditions in conventional symbols can be described in the following boundary value problem [16, 20, 22]:

$$(\lambda+2\mu)\frac{d^2U}{dx^2}+\frac{d(\lambda+2\mu)}{dx}\frac{dU}{dx}- -\left(\frac{d(\lambda+2\mu)}{dx}T+(\lambda+2\mu)\frac{\partial T}{\partial x}\right)\alpha_T = X \quad (4)$$

$$\varepsilon=\frac{dU}{dx},\ \sigma=E\left(\frac{dU}{dx}-\alpha_T\bar{T}\right),\ x\in(0,l), \quad (5)$$

$$L_1U(0)=0,\quad L_2U(l)=0, \quad (6)$$

and equations describing convective diffusion taking into account heat and mass transfer and convective heat and mass transfer with appropriate boundary conditions.

Here x – vertical coordinate, $x\in(0;l)$; $U(x)$ – transfer in soil mass relative to the axis Ox; X – volumetric force, which is determined with the formula $X=\gamma_{sb}+\frac{dp}{dx}$; λ, μ – Lame parameters; $c(x,t)$ – concentration of salt solution in soil mass; α_T – coefficient of linear heat expansion; $T(x,t)$ – temperature; γ_{sb} – specific soil gravity in weighed state; E – Young's module; $\varepsilon(x)$, $\sigma(x)$ – deformations and strains in the layer of soil mass relatively.

A finite-difference approximation of equations (4)–(6) by the finite difference method has been performed. The method of multiple calculations has been used to find the values of vertical displacements of the soil mass from equation (4).

The numerical experiments have been performed to establish the effect on the soil mass stress-deformed state of changes in the contaminated substance's temperature, in the presence of this factor, and in its absence.

For the case when the temperature of the pollutants corresponds to the environment, the dependences of the Lame parameters and the Yung's modulus on the concentration of the pollutants have been accepted based on the experimental data which are given in [15]:

$$\lambda(c)=a_3^2\cdot c^3+a_2^2\cdot c^2+a_1^2\cdot c+a_0^2 \quad (7)$$

where a_3^2 = -1798.96, a_2^2 = 4314.732, a_1^2 = -2615.37, a_0^2 = 2545.743;

$$\mu(c)=a_3^3\cdot c^3+a_2^3\cdot c^2+a_1^3\cdot c+a_0^3 \quad (8)$$

where a_3^3 = -1205.28, a_2^3 = 2880.321, a_1^3 = -1741.92, a_0^3 = 1696.324;

$$E(c)=a_3^4\cdot c^3+a_2^4\cdot c^2+a_1^4\cdot c+a_0^4 \quad (9)$$

where a_3^4 = -0.000393, a_2^4 = 0.1878866, a_1^4 = -22.70202, a_0^4 = 4410.552.

Numerous experiments have been carried out for the following input data:

l = 10 m, γ_{sb} = 1.1·10^4 $^{Pa}/_{m}$, α_T = 1·10^{-6} $^{1}/_{deg}$

As soil environments can be affected by temperature, the effect of temperature and the concentration of contaminants on the values of vertical displacements must be taken into account.

The results of numerous experiments have been highlighted in Fig. 5.

This chart was received in our developed software complex. This software complex has been created as a web platform for online mathematical and computer modeling. It consists of the backend and frontend parts. The backend was written in C# programming language and used ASP.NET. Core technology, while the frontend uses Material framework and communicate with backend via specific API functions.

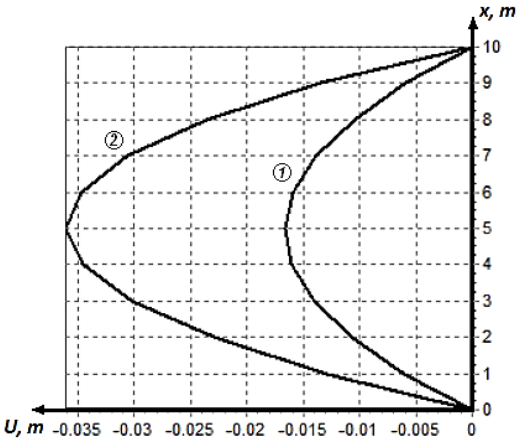


Fig. 5. The graphics of transfer distribution in the conditions of heat mass transfer under the filtration of salt solutions in nonisothermal conditions for t = 1440 days: 1 – under $\lambda=\lambda(c)$, $\mu=\mu(c)$; 2 – under $\lambda=\lambda(c,T)$, $\lambda=\lambda(c)$, $\mu=\mu(c,T)$

Experimental studies have shown that considering the concentration of pollutants reduces the values of Lame parameters compared to not considering the concentration. Taking into account the dependencies of the Lame coefficients on the concentration of pollutants and temperature, as seen in Fig. 5, displacement values increase compared to displacement values considering only the dependence of the Lame coefficients on concentration.

3. Conclusion

The obtained dependences of the deformation modulus and Lame coefficients in the form of polynomials on the concentration of saline solutions and their temperature allowed to improve mathematical models of the stress-deformed state of the soil taking into account nonlinear deformation processes occurring in soil masses under the presence and filtration of saline solutions.

Incorporating the effect of temperature on soil environments, it's crucial to consider how both temperature and contaminant concentration influence vertical displacement values. This aspect is particularly important when modeling soil behavior under varied environmental conditions. Additionally, in scenarios where pollutant temperatures align with the ambient environment, understanding the relationship between pollutant concentration and soil properties like Lame parameters and Young's modulus becomes vital. These relationships are key to accurately predicting soil response and ensuring the integrity of structures built on or within these soils.

4. Acknowledgment

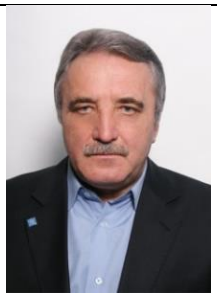
The authors would like to acknowledge for the funding scientific projects of young scientists. The publication contains the research results conducted within the project name "Information and analytical system of organic agriculture and ensuring environmental sustainability of soils" (#0120U000235).

The publication has become possible because of the courage of the Armed Forces of Ukraine, which provide security and allow us to work under a clear sky, on safe ground and away from the tyrannical government.

References

[1] Abdo A. M. A.: Utilizing a Simple Numerical Model in Discrete Element Analysis to Simulate Flow Time and Number Tests of Asphalt Mixes. Jou. Eng. Res. 11(2), 2017, 39 [https://doi.org/10.24200/tjer.vol11iss2pp39-49].
[2] Ait M'Barek S. et al.: Multi-Site Calibration and Validation of SWAT Model for Hydrologic Modeling and Soil Erosion Estimation: A Case Study in El Grou Watershed, Morocco. Ecol. Eng. Environ. Technol. 22(6), 2021, 45–52 [https://doi.org/10.12912/27197050/141593].
[3] Al-Abdul Wahhab H. I., Abdullah G. M. S.: Prediction of Permanent Deformation of Foamed and Emulsified Sulfur Asphalt Soils Mixes. Jou. Eng. Res. 15(2), 2018, 1 [https://doi.org/10.24200/tjer.vol13iss1pp1-21].
[4] Boyko I. P. et al.: Influence of hydrogeological conditions on the deformation of the fundamentals of existing foundations. Fundamentals and foundations 24, 1991, 23–30.
[5] Gerasimov I. et al.: Hydrodynamic features of the floating water outlet from a pumping station. Journal of Water and Land Development 48 (1) 2020, 162–171 [https://doi.org/10.24425/jwld.2021.136159].
[6] Hetnarski R. B.: Encyclopedia of thermal stresses. Springer Reference, Dordrecht 2014.
[7] Hutniczak A. et al.: The concept of reclamation in the life sciences to new environmental challenges. Inż. Ekolog. 20(4), 2019, 14–20 [https://doi.org/10.12912/23920629/113633].
[8] Jafari M. et al.: Dynamic approaches for system identification applied to deformation study of the dams. Acta Geod. Geophys. 50(2), 2015, 187–206 [https://doi.org/10.1007/s40328-014-0091-3].
[9] Jafari M.: Deformation modelling of structures enriched by inter-element continuity condition based on multi-sensor data fusion. Applied Mathematical Modelling 40(21-22), 2016, 9316–9326 [https://doi.org/10.1016/j.apm.2016.06.010].
[10] Kaliukh I. et al.: Decreasing Service Life of Buildings Under Regular Explosion Loads. Cybernetics and Systems Analysis 54(6), 2018, 948–956 [https://doi.org/10.1007/s10559-018-0098-9].
[11] Kaliukh I. et al.: Geotechnical Issues of Landslides in Ukraine: Simulation, Monitoring and Protection. Proceedings of China-Europe Conference on Geotechnical Engineering. Springer International Publishing, Cham 2018, 1466–1469.
[12] Klimov S. et al.: Limiting horizontal water filtration using drainage-screened modules. Journal of Water and Land Development 43(1), 2019, 90–95 [https://doi.org/10.2478/jwld-2019-0066].
[13] Kuzlo M. et al.: Deformations of Soil Massifs Under the Existence of Saline Solutions with Different Concentration and Temperature. Proceedings of the 2nd International Conference on Building Innovations. Springer International Publishing, Cham 2020, 123–131.
[14] Kuzlo M. T., Filatova I. A.: Investigation of the influence of the concentration of saline solutions on the deformation characteristics of soils. Bulletin National University of Water and Environmental Engineering, 31, 2006, 175–181.
[15] Kuzlo M. T.: Experimental and theoretical studies of soil masses' deformations under the action of human induced factors. NUWEE, Rivne 2019.
[16] Kuzlo M. T. et al.: Mathematical modelling of soil massif's deformations under its drainage. International Journal of Applied Mathematics 31(6), 2018 [https://doi.org/10.12732/ijam.v31i6.5].
[17] Kuzlo M. T. et al.: Experimental studies of saline solutions filtrations depending on their concentration and 114an dunder114. Automobile roads and road construction 107, 2020, 57–62.
[18] Methods of laboratory determination of durability and deformation characteristics. 2nd ed. (GOST 12248 – 96). State Committee of Ukraine for Urban Development and Architecture, Kyiv 01/04/1991.
[19] Sergeeva E. M. (ed): Soil Science. Moscow State University, Moscow 1983.
[20] Vlasyuk A. et al.: Mathematical Modelling of Spatial Deformation Process of Soil Massif with Free Surface. Advances in inteligent systems and computing IV, 1080, 2020, 107–120 [https://doi.org/10.1007/978-3-030-33695-0_9].
[21] Vlasyuk A. P. et al.: Computer modeling of heat and mass transfer effect on the three-dimensional stressed-strained state of soil massif. 18th International Multidisciplinary Scientific Geoconference SGEM, 18(1.2), 2018, 153–160 [https://doi.org/10.5593/sgem2018/1.2/S02.020].
[22] Vlasyuk A. P. et al.: Mathematic and computer modeling of cohesion effect forces on spatial deformation processes of soil massif. Math. Model. Comput. 7(1), 2020, 196–205 [https://doi.org/10.23939/mmc2020.01.196].
[23] Yamada Y., Ishihara K.: Anisotropic deformation characteristics of 114an dunder three dimensional stress conditions. Soils and Foundations 19(2), 1979, 79–94.

Prof. Mykola Kuzlo
e-mail: m.t.kuzlo@nuwm.edu.ua



He is an expert in hydro-melioration, holding a Dr. Sc. in Technical Sciences (2015) and a Professorship (2016). He graduated from the Ukrainian Institute of Water Management Engineers in Rivne (1984) and has been working there since 1989. He serves as the Head of the Department of Automobile Roads, Foundations, and Bases. His research focuses on predicting soil mass deformations, evaluating the stability of soil slopes, and assessing natural slopes' stability under changing hydrogeological conditions and anthropogenic factors.

https://orcid.org/0000-0002-1016-0396

Prof. Viktor Moshynskyi
e-mail: v.s.moshynskyi@nuwm.edu.ua

He is the Rector of the National University of Water and Environmental Engineering. With a career starting in 1987 as a hydrogeologist, he earned a Doctor in Agricultural Sciences in 2002. His academic career includes roles as an Associate Professor in agrochemistry, soil science, and agriculture, and later as Head of the Department of Land Management and Cadastre. Research interests: environmental protection, agricultural melioration, land management, and mathematical modeling in agriculture and environmental activities.

https://orcid.org/0000-0002-1661-6809

Ph.D. Eng. Nataliia Zhukovska
e-mail: n.a.zhukovska@nuwm.edu.ua

She is an associate professor at the Department of Computer Science and Applied Mathematics. She earned her Ph.D. in Technical Sciences in 2016 with a dissertation on "Mathematical Modeling of Heat and Mass Transfer in Filtration of Salt Solutions and their Effect on Deformation Processes in Soil Masses".
N. Zhukovska specializes in mathematical modeling and computational methods, with over 70 scientific and methodological works. She is a laureate of the 2018 Young Scientists Award from the Rivne Regional State Administration.

https://orcid.org/0000-0001-7839-0684

Ph.D. Eng. Viktor Zhukovskyy
e-mail: v.v.zhukovskyy@nuwm.edu.ua

He is an associate professor at the Department of Computer Science and Applied Mathematics. He earned his Ph.D. in Mathematical Modeling and Computational Methods in 2018 with his dissertation on "Mathematical and Computer Modeling of Mass Transfer of Salt Solutions in Catalytic and Dispersed Media of Microporous Structure". He is a laureate of the 2018 Young Scientists Award from the Rivne Regional State Administration.

https://orcid.org/0000-0002-7088-6930

http://doi.org/10.35784/iapgos.5858 *received: 02.02.2024 | revised: 20.02.2024 | accepted: 18.03.2024 | available online: 31.03.2024*

RUNNING A WORKFLOW WITHOUT WORKFLOWS: A BASIC ALGORITHM FOR DYNAMICALLY CONSTRUCTING AND TRAVERSING AN IMPLIED DIRECTED ACYCLIC GRAPH IN A NON-DETERMINISTIC ENVIRONMENT

Fedir Smilianets, Oleksii Finogenov
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

Abstract. *This paper introduces a novel algorithm for dynamically constructing and traversing Directed Acyclic Graphs (DAGs) in workflow systems, particularly targeting distributed computation and data processing domains. Traditional workflow management systems rely on explicitly defined, rigid DAGs, which can be cumbersome to maintain, especially in response to frequent changes or updates in the system. Our proposed algorithm circumvents the need for explicit DAG construction, instead opting for a dynamic approach that iteratively builds and executes the workflow based on available data and operations, through a combination of entities like Data Kinds, Operators, and Data Units, the algorithm implicitly forms a DAG, thereby simplifying the process of workflow management. We demonstrate the algorithm's functionality and assess its performance through a series of tests in a simulated environment. The paper discusses the implications of this approach, especially focusing on cycle avoidance and computational complexity, and suggests future enhancements and potential applications.*

Keywords: distributed computing, directed acyclic graph, pipeline processing

OBLICZANIE PRZEPŁYWÓW PRACY BEZ PRZEPŁYWÓW PRACY: PODSTAWOWY ALGORYTM DYNAMICZNEGO KONSTRUOWANIA I PRZESZUKIWANIA NIEJAWNEGO SKIEROWANEGO GRAFU ACYKLICZNEGO W ŚRODOWISKU NIEDETERMINISTYCZNYM

Streszczenie. *W artykule przedstawiono nowy algorytm dynamicznego konstruowania i przejść skierowanych grafów acyklicznych (DAG) w systemach zarządzania przepływem pracy, w szczególności tych ukierunkowanych na domeny obliczeń rozproszonych i przetwarzania danych. Tradycyjne systemy zarządzania przepływem pracy opierają się na jawnie zdefiniowanych, sztywnych grafach DAG, które mogą być uciążliwe w utrzymaniu, zwłaszcza w odpowiedzi na częste zmiany lub aktualizacje systemu. Proponowany algorytm pozwala uniknąć konieczności jawnego konstruowania SAG, zamiast tego wybierając dynamiczne podejście, które iteracyjnie buduje i wykonuje przepływy pracy w oparciu o dostępne dane i operacje. Korzystając z kombinacji jednostek, takich jak typ danych, operator i element danych, algorytm niejawnie buduje DAG, upraszczając w ten sposób proces zarządzania przepływami pracy. Demonstrujemy funkcjonalność algorytmu i oceniamy jego wydajność za pomocą serii testów w symulowanym środowisku. W artykule omówiono implikacje tego podejścia, ze szczególnym uwzględnieniem unikania pętli i złożoności obliczeniowej, a także zasugerowano dalsze ulepszenia i potencjalne zastosowania.*

Słowa kluczowe: obliczenia rozproszone, skierowane grafy acykliczne, potokowe przetwarzanie danych

Introduction

Workflows, pipelines and systems enabling them are a mainstay of many different areas involving distributed computation and data processing. Primarily, the systems in question are based on Directed Acyclic Graphs, which in this domain represent activities and order-of-operations relationships between them. Thus, the graph of the workflow or a pipeline needs to be directed to maintain the order of operations, and acyclic to not cause an endless loop of computation.

In common workflow management systems the DAGs are explicit, rigid, developed and maintained by the operator of the system, which results in painstaking and meticulous development, maintenance and update process. An extreme scenario of that might be when a component commonly used in the pipelines receives a breaking change, rising a need to address the breaking change in every pipeline.

Thus, operator-defined DAG workflow runners, while gracefully solving tasks they are intended for, impose a non-insignificant burden, which should be considered and addressed when designing, maintaining and improving applications that rely on those systems.

This paper explores the possibility to execute workflows without actually defining and maintaining directed acyclic graphs by dynamically constructing and traversing a Directed Acyclic Graph from a set of defined Operators with defined input and output Data Kinds.

While exploring and constructing DAGs in an automated way is frequently done in various domains, like cumulative risk assessment [1] as well as epidemiology [4], the attention of these researches is mostly driven towards causal inference and evidence synthesis, which is an environment principally different from workflow construction and execution, since causal graphs are mainly a statistics tool.

1. Algorithm description

1.1. Conceptual overview

The main goal of the proposed algorithm is to enable a way of dynamically constructing and traversing a Directed Acyclic Graph using user defined data types and operations available on those, with the possibility for the operations to be non-deterministic in regards to their outputs, so that the operator could return all, some or none of its possible output types, and output them in any quantity desired. The traversal should be seeded by a set of input data marked with their corresponding types, and the algorithm should then be able to iterate over the data, and the provided operators until it exhausts all calculations possible with the input data, data returned by all executed operators and all available operators. The proposed algorithm should also be able to detect and avoid cycles so that the users of the envisioned system using this algorithm would be safeguarded against endless cyclic calculations and equipment and maintenance time required to prevent, detect and stop those.

To achieve the goals and requirements established, presented algorithm is designed around working with several primary entities, which together are used to give a complete description of a given execution with a set of given input, intermediary and output data:

- **Data Kind** is a semantic label that is used to annotate the required inputs of a given Operator, and any data that exists within the scope of a given execution. It should also be stated that the algorithm described here is not concerned with data validation, so Data Kinds are merely labels. One should conceptualize those as Types in Charles Sanders Peirce's Type-Token distinction.
- **Operator** is an atomic operation over a set of strictly defined inputs of various Data Kinds, and may or may not output a set

of outputs of arbitrary Kinds. Neither the size nor the Kind labels of the outputs of a given operator are deterministic from the perspective of the algorithm described here.

- **Data Unit** is a piece of data. If Data Kinds are Peirce's Types, then Data Units are Tokens. A given Data Unit consists of a Data Kind label for the item, a reference to a file containing said data, and an ancestry list, which describes the sequence of Operators which resulted in the existence of this Data Unit.
- **Step** is an execution of a given Operator with Data Units matched to the inputs of an Operator.

Evidently, this set of entities does not describe a Directed Acyclic Graph explicitly, instead opting for entities that are easy to describe and maintain for the user of the envisioned system exploiting proposed algorithm, thus the DAG exists from the combination of these entities implicitly, rather than explicitly, although it is possible to reconstruct it from these entities for visualization purposes.

1.2. Algorithm details

For convenience, the algorithm is split into several functions. In the following pseudocode, `O` is the set of all Operators in the system, `K` is the set of all Data Kinds, `S` is the set of all computed steps in a given execution, `U` is the set of all Data Units (input and output to the Operators) in a given execution.

Algorithm 1 describes the general order of operations of the proposed algorithm's environment. On each turn, we solve the next steps using the current Data Pool `U`, and execute those to get new Data Units to add to `U` for the next turn of computations.

It should also be mentioned that the main loop does not necessarily need to wait for all the computed steps to finish until solving the next turn of new steps. To cut out the wait time on longer operators, we suggest implementing this in a way that would compute the next steps every time at least one of the operators finishes executing. Since both the wait time optimization, and the pseudocode for the `ExecuteSteps` function are somewhat implementation specific and don't constitute the main proposal of this paper – we will omit them here.

Algorithm 1: Main loop

```
Data: O, K, U satisfying ∀u ∈ U, u.dataKind ∈ K containing input data provided by user
Result: U containing results of exhaustive computation with o ∈ O.
start:

S ← ∅;
repeat
    N ← SolveTurn(O,K,S,U);
    if N ≠ ∅ then
        U ← U ∪ ExecuteSteps(N,U);
        S ← S ∪ N;
    end
until N ≠ ∅
end

Function SolveTurn(O, K, S, U):
/* Get those operators, whose input kinds are satisfied with at least one unit in execution's data pool */
    O_I ← GetAvailableOperators(O, U);
    opsWithInputOptions ← map o ∈ O with
        SelectOperatorInputs(o, U);
    if opsWithInputOptions = ∅ then
        return ∅;
    end
/*  Per operator, make a cartesian product of all available units per input to create suitable sets of inputs that would constitute a Step to execute */
opsWithInputCombinations ← map
opsWithSatisfiedInputs with f(o) =
                (operator:o.operator, inputCombinations:
            {i_1…i_n ∈ o.unitsPerInput|i_1 × i_2 × … × i_n});
    newSteps ← ∅;
    foreach o ∈ opsWithInputCombinations do
        foreach i ∈ o.inputCombinations do
            if
                i does not contain duplicate units
then
                append (operator:o, input:i) to
                    availableNewSteps;
            end
        end
    end
    N ← {n ∈ newSteps | n ∉ S};
    return N
Function GetAvailableOperators(O,U):
    kindsPresent ← map u ∈ U to u.dataKind;
    return
            {o ∈ O| o.inputDataKinds ⊆ kindsPresent};
Function SelectOperatorInputs(o,U):
    matchingDataUnits ← filter U for u ∈ U,
        u.dataKind ∈ o.inputDataKinds;
    unitsPerInput ← empty Map;
    foreach input ∈ o.inputs do
        unitsPerInput[input.name] ←
                        empty List;
        foreach unit ∈ matchingDataUnits do
            if unit.dataKind = input.dataKind
                and ¬o.name ∈ unit.ancestors then
                append unit.id to
                    unitsPerInput[input.name];
            end
        end
    end
    return (operator, unitsPerInput);
```

Function `SolveTurn` in Algorithm 1 contains the main order of operations of proposed solution to the problem of discovery and traversal of an implied Directed Acyclic Graph. The turn is solved by first figuring out the operators that have the necessary inputs to be executed – so that each of the inputs has at least one Data Unit in the pool that matches the input's data kind. After that what we have is a set of data units per input satisfying operator's requirements. To produce a set of viable next steps to execute we need to transform the sets of units per operator into set of pairings, where the operator is paired with one data unit per input. Since we need to exhaust all possible ways to compute a given operator with given inputs, we accomplish that transformation with a cartesian product of N sets, where N is the amount of inputs of the operator. Once we have the steps we can attempt, the only thing left is to filter out those operator-units pairs which have not been yet executed.

By doing so, each time the Data Pool is appended with new Units we still retry all possible combinations. If we went the way of only checking the new Unit Kinds for possible operators, we would miss executing the operators that had part of their inputs satisfied with input data of the execution, part with (N-1)-th turn, and part with N-th turn.

2. Experimental setup

To test the robustness of this algorithm in a semi-realistic scenario, a test bench was devised. This test bench would randomly generate 50 data kinds and 25 operators. Operators are randomly generated to have 1-4 inputs. These inputs can be either from the starting or middle third of the Data Kind array, to facilitate formation of longer executions. The execution is then seeded with three Data Units. After each pass, a random amount of Data Units ranging from 0 to twice the amount of produced valid next steps would be added to the data pool as descendant from random steps to simulate random operator outputs.

The algorithm was implemented in TypeScript, and executed as part of an HTTP server on a GKE non-Autopilot cluster with one node. Time was measured using `performance.now()` precision timing API. This implementation showed adequate results – 0.5–1 ms for executions with <5 turns and <10 steps, 4–8 ms for executions with <5 turns and <50 executed steps

and 15–30 ms for executions with <15 turns and <50 executed steps. Time taken on this scale can be considered negligible when compared to time which would be spent by the system to spawn, prepare and execute Kubernetes jobs or other ways of executing operators.

3. Discussion

3.1. Cycle avoidance

Cycle avoidance in this approach is guaranteed by only suggesting units to operator inputs that do not have this operator in their ancestry. While in this basic approach it is suggested to only keep operators in the ancestry, it might be beneficial to track the ancestry unit-wise as well, which would enable the users of the system exploiting this algorithm to run an execution on a set of different datas of same type and still be able to track down which results were achieved from which input data, which while not good practice, is something that could be expected from real world use.

Another problem of cycle avoidance in this algorithm is the case of a hypothetical operator X taking units of kinds A and B as inputs, and providing A as the output. Currently, as illustrated by figure 1, if the operator X took A_1 and B_1, and produced A_2, and, by addition from the user or from some other operator, data unit B_2 would then appear, operator X would only be re-run with A_1 and B_2 as it's inputs, while in a theoretical scenario it might be beneficial to run X with A_2 and B_2 as well, despite A_2 being produced by operator X itself, as illustrated by figure 2.

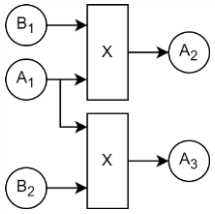


Fig. 1. Current solution to the hypothetical operator X scenario

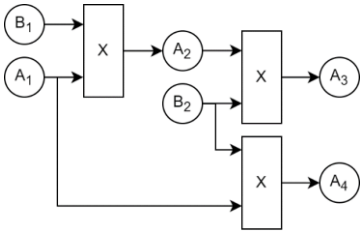


Fig. 2. Possible solution to the hypothetical operator X scenario

While it's assumed a described scenario is highly unlikely, it could be argued that a use case described in Fig. 2 is a more correct way to solve such a theoretical scenario.

3.2. Complexity

The complexity of presented algorithm is hard to point out precisely due to the nature of processing several entities at the same time and spawning new entities during the process, but we can still analyse some of the aspects of it.

The theoretical narrow point of presented algorithm is the computation of the cartesian product of sets of data units per operator input, which has a complexity of $O(2^n)$ with n being the number of operator inputs per specific operator. While $O(2^n)$ a horrible complexity, we consider that it's highly unlikely this will be a problem in the real world usage scenarios, since an operator with hundreds of inputs is quite hard to imagine and justify.

From the perspective of operator count (n) and data unit (m) count the complexity of this algorithm can be described as $O(nm)$.

3.3. Running at scale

While the pseudocode provided, and the preliminary implementation of this algorithm are designed and presented from a perspective of a central service which would orchestrate the computation, it is obvious that there are uses and scenarios where one service would not be enough to handle the amount of data flowing through it.

From our perspective, this algorithm can also be implemented in a decentralised and distributed manner by utilizing Pub/Sub [3] or other distributed communication model based tools to trigger both operator execution and the solution for the next possible steps. Given that operator executors and the solvers will be implemented in a stateless manner themselves, instead outsourcing state and data handling to other distributed systems, there should be no problems employing this approach in a decentralised manner.

3.4. Future work and potential applications

The presented algorithm is a basic attempt at the described problem. While it is a solution, there most probably are at least slightly better ways to compute an implied DAG in such a manner, from adding more features and possible nodes to the graph by implementing a more thought-through version of cycle avoidance, to optimizations in ways the operators, data units and possible next steps are filtered down to produce valid next steps not executed before.

The algorithm presented is intended to be used in software requiring automated workflow management systems as a replacement or a complement for CWL, Nextflow and other similar systems commonly used in data engineering pipelines and bioinformatical/biostatistical computations [5, 6], as well as for running data through neural networks [2]. While it is obvious this approach would not be a replacement for a significant majority of use-cases of workflow systems in general, the utilization of this algorithm in an envisioned system is predicted to enable easier development of prototypes and small-scale systems for scientific and bioinformatical computations relying on data pipelines with lots of common components by cutting down the time and effort spent on development and maintenance of pipelines. We see the presented approach as being especially valuable in circumstances with constantly changing requirements that grow from the new and new opportunities discovered during the process of developing with the suggested approach.

Designed around non-deterministic outputs, the presented algorithm also allows flexible and effortless branching and expansion in a way that is not possible with CWL. Figure 3 describes a use case for a workflow system employing presented algorithm to run three-dimensional CT scan analysis.

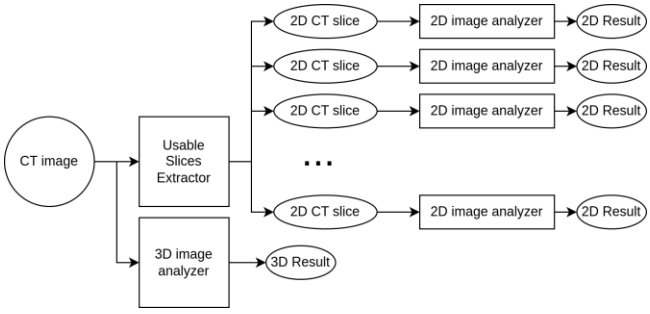


Fig. 3. Usage scenario for CT scan analysis

In this scenario, usable slice extractor would, for example, filter out the slices which contain the area of interest – for example, lungs in chest CT imaging, and pass them on to other analyzers. As it's impossible to predict how many slices would be actually containing the area of interest due to inherent variety in human bodies, the extractor would output an unpredictable

amount of Data Units, for each of which the system will automatically find and execute the according analysis operators. The system would also allow easy extensibility with additional three- and two-dimensional analyzers.

4. Conclusion

An algorithm for dynamic construction and traversal of a workflow DAG was presented, with an aim to reduce user effort in developing and maintaining workflows. The described algorithm iterates over data and operators in the system, selecting possible to run data-operator combinations on each step of the algorithm until there is no more data-operator pairings possible and not previously executed. The algorithm also avoids cyclic computations, although in a way which would not allow some theoretical uses. Experimental results indicate that the algorithm is robust and avoids stalling and endless computation and is able to generate and execute complex workflows, supporting operators with uncertain outputs. While the computation complexity of parts of this algorithm are not ideal, practical implications of it seem insignificant in real world situations. Future work will focus on refining cycle avoidance mechanisms, optimizing performance, and development of a Kubernetes-based system utilizing the presented algorithm.

Acknowledgements

Special thanks to Anton Shpigunov, Radek Janik, and Anton Zhdan-Pushkin for their valuable perspectives and insights during the development of the algorithm and the writing of this paper. Their contributions were instrumental in shaping both the technical and narrative aspects of this work.

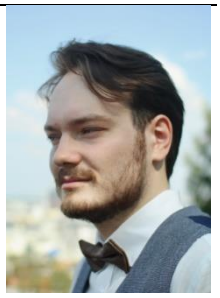
References

[1] Brewer L. E. et al.: Causal inference in cumulative risk assessment: The roles of directed acyclic graphs. Environment International 102, 2017, 30–41 [https://doi.org/https://doi.org/10.1016/j.envint.2016.12.005].
[2] Colonnelli I. et al.: Bringing AI pipelines onto cloud-HPC: setting a baseline for accuracy of COVID-19 diagnosis. ENEA CRESCO in the Fight Against COVID-19, 2021, 66–73 [https://doi.org/10.5281/ZENODO.5151511].
[3] Eugster P. Th. et al.: The many faces of publish/subscribe. ACM Comput. Surv. 35(2), 2003, 114–131.
[4] Ferguson K. D. et al.: Evidence synthesis for constructing directed acyclic graphs (ESC-DAGs): a novel and systematic method for building directed acyclic graphs. International Journal of Epidemiology 49(1), 2019, 322–329 [https://doi.org/10.1093/ije/dyz150].
[5] Georgeson P. et al.: Bionitio: demonstrating and facilitating best practices for bioinformatics command-line software. GigaScience 8(9), 2019, giz109 [https://doi.org/10.1093/gigascience/giz109].
[6] Jackson M. et al.: Using prototyping to choose a bioinformatics workflow management system. PLOS Computational Biology 17(2), 2021.

M.Sc. Fedir Smilianets
e-mail: fedor.smile@gmail.com

Fedir Smilianets received his M.Sc. from National Tecnhical University of Ukraine "Igor Sikorsky Polytechnic Institute". Since 2021, he is a Ph.D. student at the department of Computer Science and Software Engineering, also working as a teaching assistant there since 2023.
His research interests include machine learning and systems for data processing in the cloud.

https://orcid.org/0000-0002-0061-7479



Ph.D. Oleksii Finogenov
e-mail: fenyatrashbox@gmail.com

Ph.D., docent at Department of Computer Science and Software Engineering of National Technical University of Ukraine "Igor Sikorsky Polytechnic Institute".
Research interests include numerical analysis, optimization methods, parallel computing, decision making systems.

https://orcid.org/0000-0002-1708-5632

http://doi.org/10.35784/iapgos.5423 *received: 22.10.2023 | revised: 11.02.2024 | accepted: 03.03.2024 | available online: 31.03.2024*

INTELLIGENT DATA ANALYSIS ON AN ANALYTICAL PLATFORM

Dauren Darkenbayev[1], Arshyn Altybay[1], Zhaidargul Darkenbayeva[2], Nurbapa Mekebayev[3]
[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan, [2]Kazakh Ablai Khan University of International Relations and World Languages, Almaty, Kazakhstan,
[3]Kazakh National Women's Teacher Training University, Almaty, Kazakhstan

Abstract. *The article discusses methods for processing unstructured data using an analytical platform. The authors analyze existing methods and technologies used to implement data processing and propose new approaches to solving this problem. The possibilities of using analytical platforms to solve the problem of processing source data are considered. The purpose of the article is to explore the possibilities of data import, partial preprocessing, missing data recovery, anomaly removal, spectral processing and noise removal. The authors explored how analytics platforms can function without a data warehouse, obtaining information from any other sources, but the most optimal way is to use them together, and how big data and unstructured data can be processed using an analytics platform. The authors solved a specific problem related to processing problems and proposed ways to solve them using an analytical platform. Particular attention is paid to a complete set of mechanisms that allows you to obtain information from any data source, carry out the entire processing cycle and display the results. Overall, the paper represents an important contribution to the development of raw data processing technologies. The authors plan to continue research in the field of processing big unstructured data.*

Keywords: raw data, processing, analytical platform, technology, analysis

INTELIGENTNA ANALIZA DANYCH NA PLATFORMIE ANALITYCZNEJ

Streszczenie. *W artykule omówiono metody przetwarzania surowych danych z wykorzystaniem platformy analitycznej. Autorzy analizują istniejące metody i technologie stosowane do realizacji przetwarzania danych i proponują nowe podejścia do rozwiązania tego problemu. Rozważane są możliwości wykorzystania platform analitycznych do rozwiązania problemu przetwarzania surowych danych. Celem artykułu jest zbadanie możliwości importu danych, częściowego przetwarzania wstępnego, przywracania brakujących danych, usuwania anomalii, przetwarzania spektralnego i usuwania szumu. Autorzy sprawdzili, jak platformy analityczne mogą funkcjonować bez hurtowni danych, otrzymując informacje z innych źródeł, jednak najbardziej optymalnym sposobem jest ich wspólne wykorzystanie oraz jak duże zbiory danych można przetwarzać za pomocą platformy analitycznej. Autorzy omawiają możliwe problemy związane z problemami przetwarzania i sugerują sposoby ich rozwiązania. Szczególną uwagę zwrócono na kompletny zestaw mechanizmów, który pozwala na pozyskanie informacji z dowolnego źródła danych, przeprowadzenie całego cyklu przetwarzania i wyświetlenie wyników. Ogólnie rzecz biorąc, artykuł stanowi ważny wkład w rozwój technologii przetwarzania surowych danych. Artykuł kończy się przyszłościowym planem dalszych badań w tym obszarze.*

Słowa kluczowe: surowe dane, przetwarzanie, platforma analityczna, technologia, analiza

Introduction

The modern world and information systems are data-driven, generated in huge quantities every day [1]. Therefore, to obtain truly valuable results in modern realities in industrial, business or scientific problems, it is very important to be able to effectively process the big data available to us, using the tools available for this, and some scientists use parallel data processing on many computing nodes [2].

This article contains the theoretical material necessary to understand the basic principles of how the analytical platform works, as well as solving practical problems that allow you to start working with raw data and processing it. The review material covers issues of Big Data processing.

Big Data is one of the key tools of digitalization. Their use in government administration and business began at the turn of 2010. But the relevance and possibilities of using this technology are only increasing over time [5].

The classic tool for working with large volumes of information – structured databases – cannot process such volumes of information. This was the main reason for the emergence of Big Data technology. This term refers to working with a large volume of loosely structured information stored in different formats and frequently updated. "Big Data" may include text documents, video and audio recordings, program code, etc. The main problem here is adequate analysis tools that make it possible to compare these data with each other and ensure their useful use [6].

Unlike DBMS with a set of Data Mining algorithms, analytical platforms are initially focused on data analysis and are designed to create ready-made analytical solutions.

An analytical platform is an information and analytical system, as well as a specialized software solution that contains all the tools for carrying out the process of extracting patterns from "raw" data; the process of extracting some patterns from the entire data array is carried out: a means of consolidating information in a single source (storage data), extraction, transformation, transformation of data, Data Mining algorithms, visualization, shifting simple and complex methods and models [3].

The main purpose of a scientific article is data import, partial preprocessing, restoration of missing data, removal of anomalies, spectral processing, noise removal. For this purpose, the Deductor analytical platform was used.

Deductor is a program that implements the functions of importing, processing, visualizing and exporting data. Deductor can function without a data warehouse, receiving information from any other sources, but the most optimal way is to use them together.

In recent years, there has been a lot of thought about the role of technology in the analysis of search data. When we create technologies of all types, we must always think carefully about their consequences [8]. Typically, new technologies are created to solve specific problems or meet specific needs, so we try to determine how well they will succeed in doing so. But this is not enough. We also need to consider possible downsides-the ways in which these technologies can cause harm. This is especially true for information technology, particularly technologies for creating meaning from data, but this is rarely done by the companies that produce it [4].

1. Literature review

The relevance of the selected study is studied in the works of such scientists as Franks B. [9], Lubanovic B. [11], Rastorguev V. [12], Rimmer J. [13], Saar-Tsechansky M., Provost F. [14] etc.

Today, there are many developments in data processing, such as SAS [13], Statistica Data Miner [10] etc.

However, there are problems in processing large raw data [18]. Big Data processing technologies are developing every day and many soft-ware systems are being developed for processing big unstructured data [17].

2. Formulation of the problem

Big Data is everywhere and everywhere [7]. Every day we are faced with big data and problems of processing it. There are many analytical platforms, but their mathematical model is a trade secret, so the authors decided to use a ready-made analytical platform. Using the analytical platform, the authors wanted to show the process of processing large unstructured data [16].

Deductor Studio is a program that implements the functions of importing, processing, visualizing and exporting data. Deductor Studio can function without a data warehouse, receiving information from any other sources, but the most optimal way is to use them together [15]. Deductor Studio includes a full set of mechanisms that allows you to obtain information from an arbitrary data source, carry out the entire processing cycle (cleaning, transforming data, building models), display the results in the most convenient way (OLAP, charts, trees) and export the results externally. This is entirely consistent with the concept of Knowledge Discovery from Databases (KDD) [19].

Given a data set containing columns such as "Argument", "Sinus", "Anomalies", "More noise", "Medium noise", "Small noise". The separator between columns is a tab character. The Argument column is assigned values from 0 to 2.96 in increments of 0.02. For any twenty argument values, skip entering data in sine values. The values in the Anomalies column are equal to the values in the Sine column, but have no missing data, but 10 values deviate sharply from the true value of the sine of the argument. The values for the columns "More noise", "Medium noise", "Small noise" have values close to the value of the sine of the argument, but have some deviation (dispersion and are selected from the range -1.5 to 1.5) (Fig. 1). Import data, create a file, process data, restore missing sinus values, perform partial processing, remove anomalies and noise.

Argument	Sinus	Anomalies	Big noises	Medium noises	Small noises
0.14	0.139543115	0.139543115	-0.006298541	0.126864534	0.174892379
0.16	0.159318207	0.159318207	0.33294777	0.085238625	0.145575649
0.18	0.179029573	0.179029573	0.279505602	0.085723008	0.134508346
0.2	0.198669331	0.198669331	0.258280061	0.277157184	0.176827573
0.22	0.218229623	0.5	0.139744277	0.28969863	0.231458805
0.24	0.237702626	0.237702626	0.317646146	0.177087062	0.282627785
0.26	0.257080552	0.257080552	0.266913669	0.19389017	0.288192043
0.28	0.276355649	0.276355649	0.131932825	0.325878695	0.243445006
0.3		0.2955200207	0.403496144	0.30362014	0.30123014
0.32		0.314566561	0.430780976	0.306273759	0.312952364

Fig. 1. Data set

Often the source data is not complete enough or has various noises and is not suitable for analysis, and the quality of the data affects the quality of the results. So the issue of preparing data for subsequent analysis is very important. Typically, "raw" data contains various noises, behind which it is difficult to see the overall picture, as well as anomalies – the influence of random or rare events. Obviously, the influence of these factors on the general model must be minimized, because a model that takes them into account will be inadequate.

3. Partial preprocessing

Partial preprocessing is used to recover missing data, edit anomalous values, and spectral data processing (for example, data smoothing). This step is often carried out first. In data processing, in the case of missing data, a repeating number is often written in their place, so blanks are filled, and it is better to monitor the filling of blanks. In many cases, those gaps affect the results of data processing and it can be poor quality data. Poor quality data directly leads to poor results.

4. Recovering missed data

It often happens that some data in a column is missing for some reason (the data is unknown, or they forgot to enter it, etc.). Normally, this would cause all rows that contain missing data to be removed from processing. But Deductor Studio mechanisms allow you to solve this problem. One of the partial processing steps is responsible for restoring missing values. If the data is ordered (for example, by time), then it is recommended to use approximation to recover missing values. The algorithm itself will select a value that should replace the missing value based on nearby data. If the data is not ordered, then you should use the maximum likelihood mode, when the algorithm substitutes the most probable values for the missing data, based on the entire sample.

5. Removing anomalies

Anomalies are deviations from the normal behaviour of something. This could be, for example, a sharp deviation of a value from its expected value. Automatic editing of anomalous values is carried out using robust filtering methods, which are based on the use of robust statistical estimates, such as the median. In this case, it is possible to set an empirically selected criterion for what is considered an anomaly. For example, setting the degree of suppression of anomalous data to "weak" means the most tolerant attitude towards the value of permissible emissions. Essentially, anomalies should not have any effect on the result at all. If they are present during the construction of the model, then they have a very large influence on it. They must first be eliminated. They also spoil the statistical picture of the data distribution. Data with anomalies, as well as a histogram of their distribution are presented in Fig. 2.

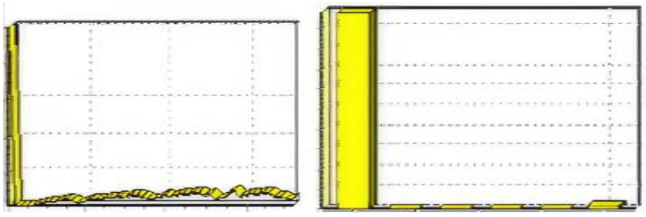


Fig. 2. Anomaly data and histogram distribution

Obviously, anomalies do not allow us to determine both the nature of the data themselves and the statistical picture. The data after eliminating the anomalies is presented in Fig. 3.

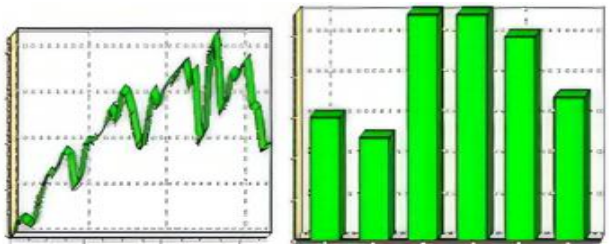


Fig. 3. Results of anomaly removal

6. Spectral processing

Data smoothing is used to remove noise from the original data set. The Deductor Studio platform offers several types of spectral processing: data smoothing by specifying the bandwidth, noise subtraction by specifying the degree of noise subtraction, and wavelet transform by specifying the depth of decomposition and wavelet order.

Spectral analysis or spectral analysis is analysis in terms of the spectrum of frequencies or associated quantities, such as energies, eigenvalues, etc. In specific fields, it may refer to:

Spectroscopy in chemistry and physics, a method for analyzing the properties of substances by their electromagnetic interactions.

Spectral estimation – in statistics and signal processing, an algorithm that estimates the strength of various frequency components (power spectrum) of a signal in the time domain. This can also be called frequency domain analysis.

A spectrum analyzer is a hardware device that measures the magnitude of the input signal as a function of frequency over the full frequency range of the instrument.

Spectral theory in mathematics is a theory that extends eigenvalues and eigenvectors to linear operators in Hilbert space and, more generally, to elements of Banach algebra.

In nuclear and particle physics, gamma-ray spectroscopy, and high energy astronomy, the analysis of pulse amplitude analyzer output for characteristic features such as spectral lines, edges, and various physical processes that create continuous shapes.

7. Noise removal

Noise in the data not only hides the overall trend, but also manifests itself when building a forecast model. Because of them, the model may turn out to have poor generalizing qualities. Spectral processing allows you to do this by specifying these fields as the processing type "Noise Subtraction" The settings have some flexibility. Thus, there is a large, medium and small degree of noise subtraction. The analyst can choose a degree that suits him. In some cases, wavelet transform gives good results for removing noise.

8. Processing results

After launching the import wizard, we will specify the import type "Text file with delimiters" and proceed to setting up the import. Let's indicate the name of the file from which we need to get data. In the viewing window of the selected file, you can see the contents of this file; we will also process the data from the "TestForPPP.txt" file.

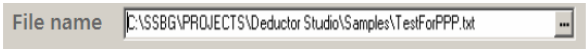


Fig. 4. Uploading a file for processing

It contains a table with the following fields: "Argument" – argument, "Sinus" – sinus values of the argument (some values are empty), "Anomalies" – sinus with outliers, "Big noises" – sinus values with large noises, "Medium noises" – sine values with medium noise, "Small noises" – sinus values with small noise. All data can be seen on the chart after importing from a text file.

After importing the file, you can see that in the "Sinus" column contains empty values. In the diagram above you can see that some sine values are missing. For further processing it is necessary to restore them. To do this, run the Partial Processing Wizard.

Since the data in the source set is ordered, at the next step of the processing wizard, select the "Sinus" field and specify the "Approximation" processing type for it. Since in this case nothing else is required, we leave the remaining processing parameters disabled. Having gone to the page for launching the processing process, we execute it by clicking on start, and then select the type of visualization of the processed data. After completing the processing process, as can be seen from Fig. 10, the gaps in the data in the diagram have disappeared, which is what needed to be done.

Next, remove anomalies from the "Anomalies" field imported table. In the partial preprocessing wizard at the third step select the "Anomalies" field and indicate to it the type of processing "Removal of anomalous phenomena", the degree of suppression "Large".

After completing the processing process, the diagram shows that the outliers have disappeared, leaving only small disturbances that can be easily smoothed out using spectral processing.

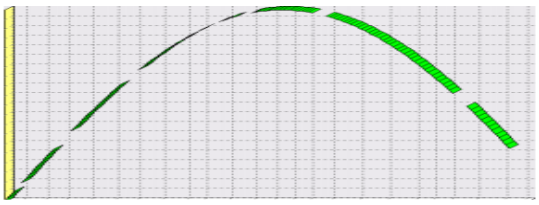


Fig. 5. Column with missing data

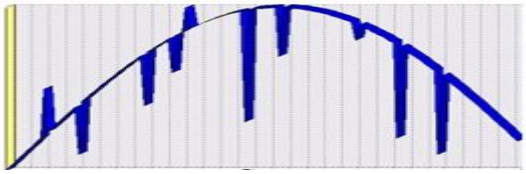


Fig. 6. Column with anomalies

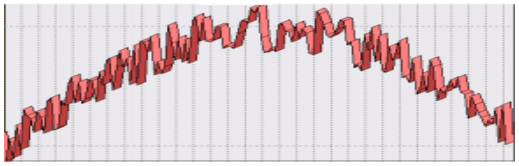


Fig. 7. Column with big ears

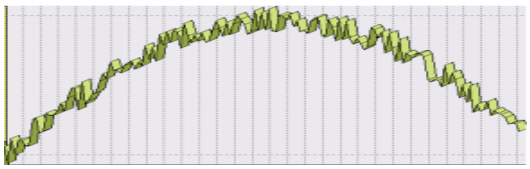


Fig. 8. Column with medium spikes

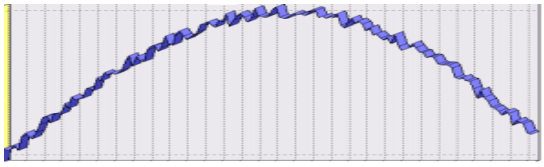


Fig. 9. Column with low noise

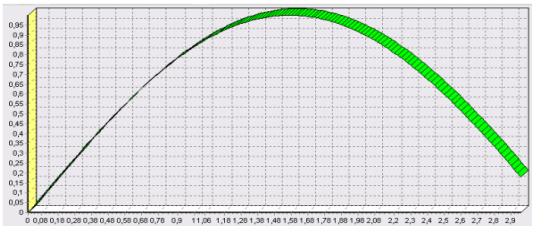


Fig. 10. Diagram after processing process

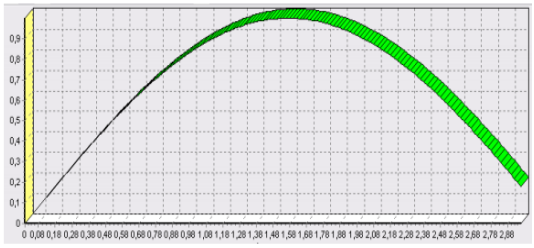


Fig. 11. Results after removing the anomaly

As can be seen in Fig. 11, the anomalies were eliminated, however minor disturbances remained. Let's smooth them out using partial processing. To do this, after removing the anomalies, we will run the partial processing wizard again. In it, in the fourth step, we select the "Anomalies" field and indicate to it the processing type "Wavelet transform" with default parameters (decomposition depth 3, wavelet order 6).

After processing, you can see in the diagram that there is no emissions and compare the result with the reference sine value (column "Sinus"). In Fig. 12 green (light) graph – sine values, blue (dark) – smoothed sine values after eliminating anomalies.

As was given in the partial processing dataset, as shown earlier, there are 3 columns of noise: "Big noise", "Medium noise", and "Small noise" – respectively a sine with large, medium and small noise. It is clear that for further work with the data, these noises must be eliminated. Thus, in the fourth step of the partial processing wizard, select the fields "Big noise", "Medium noise" and "Small noise" in turn, set the type of processing to "Subtraction of noise" and indicate the degree of suppression – "Large", "Medium" and " Small" respectively.

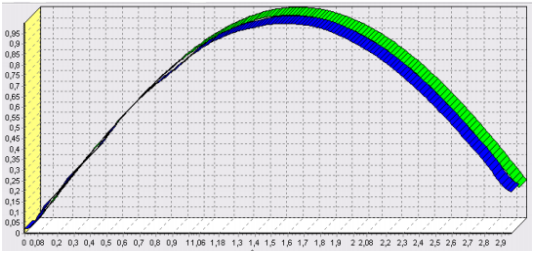


Fig. 12. Results after processing

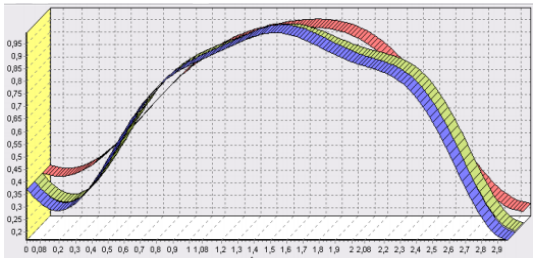


Fig. 13. Results of partial processing

Now let's remove noise using wavelet transform. In the partial processing wizard, select the "Large" Fields noise", "Medium noise" and "Small noise", indicate the type "Wavelet transform" processing, leaving the default processing parameters (decomposition depth – 3, wavelet order – 6). In the diagram you can see that the data has been smoothed (Fig. 14). The quality of noise smoothing can be improved in this way by selecting satisfactory processing parameters.

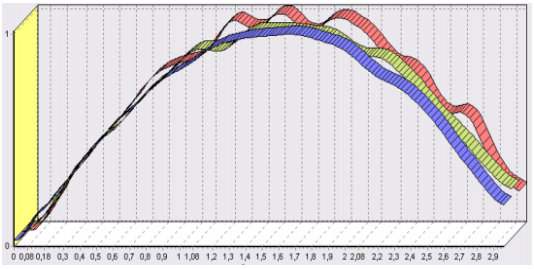


Fig. 14. Results of noise removal

9. Conclusion

The rate of data growth has become significant in the last decade. We have to work with data in different formats. Servers monitor incoming emails and transactions. It can be said that the Internet has become a major source of data storing large amounts of data. Processing data using computers is one of the main tasks of many information systems.

The technologies implemented in Deductor allow you to go through all the stages of building an analytical system on the basis of a single architecture: from data consolidation to building models and visualizing the results obtained.

Before the advent of analytics platforms, data analysis was carried out mainly in statistical packages. Their use required high user qualifications. Most algorithms implemented in statistical packages did not allow for efficient processing of large volumes of information. To automate routine operations, it was necessary to use built-in programming languages.

The article solves the problems of processing raw data. The authors concluded that such analytics platforms are needed to process unstructured data, eliminate processing errors, and obtain accurate results. The results of the study are used by the authors of this article for further modeling of big data processing and for the development of analytical platforms.

References

[1] Abdiakhmetova Z. M.: Wavelet data processing in the problems of allocation in recovery well logging. Journal of Theoretical and Applied Information Technology 95(5), 2017, 1041–1047.
[2] Altybay A. et al: Numerical Simulation and Parallel Computing of the Acoustic Wave Equation. AIP Conference Proceedings 3085(1), 2024, 020006.
[3] Balakayeva G. et al: Development of an application for the thermal processing of oil slime in the industrial oil and gas sector. Informatics, Control, Measurement in Economy and Environmental Protection 13(2), 2023, 20–26.
[4] Balakayeva G. et al: Digitalization of enterprise with ensuring stability and reliability. Informatics, Control, Measurement in Economy and Environmental Protection 13(1), 2023, 54–57 [http://doi.org/10.35784/iapgos.3295].
[5] Balakayeva G., Darkenbayev D.: The solution to the problem of processing Big Data using the example of assessing the solvency of borrowers. Journal of Theoretical and Applied Information Technology 98(13), 2020, 2659–2670.
[6] Balakayeva G. T. et al: Using NoSQL for processing unstructured Big Data. News of the NAS of the Republic of Kazakhstan 6(438), 2019, 12–21.
[7] Big Data Big Opportunity [http://www.oracle.com] (28.01.2012).
[8] Darkenbayev D. K.: Numerical solution of the regression model for analysis and processing of Big Data. Vestnik KazNRTU 6(130), 2018, 132–139.
[9] Franks B.: The Taming of Big Data: How to Extract Knowledge from Arrays of Information Using Deep Analytics. Mann, Ivanov and Ferber, 2014, 180.
[10] Highlights: Unique Features of Statistica Data Miner [http://www.statsoft.com] (01.02.2014).
[11] Lubanovic B.: Introducing Python: Modern Computing in Simple Packages 2nd Edition. O'Reilly Media, 2019.
[12] Rastorguev V.: DataMining technology for data analysis in credit scoring methods. Banking Technologies (11), 2003, 14–18.
[13] Rimmer J.: Contemporary changes in credit scoring. Credit Control 26 (4), 2005, 56–60.
[14] Saar-Tsechansky M., Provost F.: Active sampling for class probability estimation and ranking. Machine Learning 54(2), 2004, 153–178.
[15] Semenov Yu. A.: Large amounts of data (big data) [http://book.itep.ru] (21.04.2013).
[16] Usachev S.: Credit scoring: desktop or enterprise solutions. Banks and technologies (4), 2008, 50–54.
[17] [http: //www.basegroup.ru].
[18] [http://www.nosql-database.org].
[19] [https://basegroup.ru/deductor/components/studio].

Ph.D. Dauren Darkenbayev
e-mail: dauren.kadyrovich@gmail.com

Ph.D., associate professor of the Department of Computer Science, Faculty of Information Technologies, Al-Farabi Kazakh National University, Almaty, Kazakhstan.
Research interests: Big Data processing, mathematical and computer modeling, development of computer systems for the educational process, machine learning, deep learning in inverse problems.

https://orcid.org/0000-0002-6491-8043

Ph.D. Arshyn Altybay
e-mail: arshyn.altybay@gmail.com

Ph.D., associate professor of the Department of Computer Science, Faculty of Information Technologies, Al-Farabi Kazakh National University and researcher at the Institute of Mathematics and Mathematical Modeling, , Almaty, Kazakhstan.
Research interests: numerical simulation of PDE, numerical analysis, high-performance computing, numerical methods, parallel programming, machine learning, deep learning in inverse problems.

https://orcid.org/0000-0003-4939-8876

Ph.D. Zhaidargul Darkenbayeva
e-mail: zhaidargul.d@mail.ru

Candidate of Philological Sciences, associate professor at Kazakh Ablai Khan University of International Relations and World Languages.
Research interests: information technology, processing of voice and text data, semantic features of phraseological units and computational linguistics. Speech recognition using Big Data technology, machine learning, deep learning.

https://orcid.org/0000-0003-3756-0581

Ph.D. Nurbapa Mekebayev
e-mail: nurbapa@mail.ru

Ph.D. in Computer Science from Al-Farabi Kazakh National University, Almaty, Kazakhstan, in 2020. Currently, he is a senior researcher at the Institute of Information and Computing Technologies in Almaty, Kazakhstan, and an associate professor at the Department of Computer Science of the Kazakh National Women's Teacher Training University in Almaty, Kazakhstan.
Research interests: machine learning, deep learning in inverse problems and computational linguistics.
https://orcid.org/0000-0002-9117-4369