Informatics Control Measurement
in Economy and Environment Protection

3/2023

Informatyka Automatyka Pomiary

www.e-IAPGOS.pl

w gospodarce i ochronie środowiska

ISSN 2083-0157

Kwartalnik Naukowo-Techniczny

XX Konferencja Światłowody i ich zastosowania - TAL2023

11-14 września 2023, Lublin

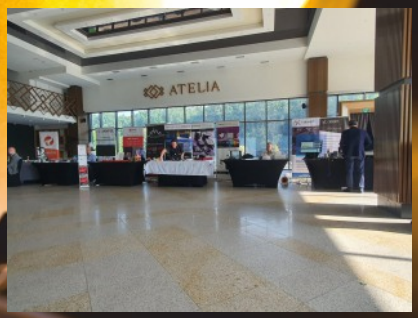
http://opticalfibers.umcs.pl/konferencja-tal-2023

XX Konferencja Światłowody i ich zastosowania - TAL 2023 11-14.09.2023 r., Lublin

Konferencja była XX z kolei spotkaniem z cyklu Krajowych Sympozjów „Światłowody i ich zastosowania". Konferencje te dotychczas odbywały się na przemian co półtora roku w Krasnobrodzie i Nałęczowie oraz w Białowieży i Lipowym Moście.

Celem konferencji jest umożliwienie bezpośredniej dyskusji merytorycznej zespołów badawczych zajmujących się:

1. rozwojem technologii i wytwarzaniem włókien światłowodowych klasycznych i mikrostrukturalnych, kabli, światłowodów planarnych, elementów optyki zintegrowanej i mikrooptyki oraz czujników światłowodowych i optycznych;
2. elementami techniki światłowodowej takimi jak sprzęgacze, złącza, wzmacniacze optyczne, urządzenia optyczne i optoelektroniczne do łączenia światłowodów ze źródłami i odbiornikami światła, multipleksery i demultipleksery itp.;
3. zastosowaniami światłowodów zwłaszcza takimi, które wymagają ścisłej współpracy ze specjalistami wytwarzającymi światłowody, tory i kable optyczne, elementy techniki światłowodowej, optolektroniki i fotoniki;
4. kształceniem w dziedzinie fotoniki w szkołach wyższych i średnich.



TAL 2023

foto: arch. TAL2023

3/2023

lipiec – wrzesień

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

PRINTING HOUSE – DRUKARNIA

PPH Remigraf Sp. z o.o.
ul. Dźwigowa 61, 01-376 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 Edukacji i Nauki z dnia 17 lipca 2023 r., Unikatowy Identyfikator Czasopisma: 200167 – z przypisaną liczbą punktów przyznawanych za publikację artykułu równą 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

http://doi.org/10.35784/iapgos.5326 *received: 08.08.2023 | revised: 09.09.2023 | accepted: 17.09.2023 | available online: 30.09.2023*

MODELING AND ANALYSIS OF SYSTOLIC AND DIASTOLIC BLOOD PRESSURE USING ECG AND PPG SIGNALS

Oleksandr Vasilevskyi[1], Emanuel Popovici[2], Volodymyr Sarana[2]
[1]University of Texas at Austin, Austin, USA, [2]University College Cork, Cork, Ireland

Abstract. *Taking into account the peculiarities of using the MAX86150 evaluation system for measuring ECG and PPG signals, mathematical models were developed for indirect determination of systolic and diastolic pressure using fingers on the hand, which were tested in the MATLAB environment. Received ECG and PPG signals. Based on the proposed mathematical models, ECG and PPG signals were processed in the MATLAB package and the results of indirect measurement of blood pressure were presented.*

Keywords: systolic pressure, diastolic pressure, ECG and PPG signals, measurement, method for determining blood pressure

MODELOWANIE I ANALIZA SKURCZOWEGO I ROZKURCZOWEGO CIŚNIENIA KRWI Z WYKORZYSTANIEM SYGNAŁÓW EKG I PPG

Streszczenie. *Biorąc pod uwagę specyfikę wykorzystania systemu oceny MAX86150 do pomiaru sygnałów EKG i PPG, opracowano modele matematyczne do pośredniego określania ciśnienia skurczowego i rozkurczowego używając palców dłoni, które zostały przetestowane w środowisku MATLAB. Otrzymano sygnały EKG i PPG. W oparciu o zaproponowane modele matematyczne, sygnały EKG i PPG zostały przetworzone w pakiecie MATLAB oraz przedstawiono wyniki pośredniego pomiaru ciśnienia krwi.*

Słowa kluczowe: ciśnienie skurczowe, ciśnienie rozkurczowe, sygnały EKG i PPG, pomiar, metoda określania ciśnienia krwi

Introduction

Determination of human blood pressure both in a static position (standing, sitting, or lying down) and during dynamic movements (when walking, running, or performing other physical exercises) is carried out by many researchers [2, 6, 8, 11, 17, 20, 35]. There are many different methods for determining systolic and diastolic pressure. All of them are implemented using different devices and sensors, as well as different mathematical expressions (formulas) for calculating blood pressure [2, 6, 11, 17]. At the same time, the accuracy of determining blood pressure (BP) values is different.

Therefore, the development and study of a mathematical model for accurate determination of human systolic and diastolic pressure on the fingers of the hand only by ECG and PPG signals for use in an electronic stethoscope [6, 35] with a diameter of 50 mm to expand its functionality is an urgent scientific task.

As a result of the analysis and research of existing methods and algorithms for determining blood pressure, it was found that many of them are difficult to apply in practice and have low accuracy and sensitivity in determining systolic and diastolic pressure for people of different ages with different values of blood pressure (BP). At the same time, the mathematical apparatus for determining blood pressure used by different researchers is interpreted and applied in different ways. In addition, different researchers use different equipment and the resulting waveforms as a result of the experiments are also different. Waveforms sometimes differ from those theoretical (fundamental) information that is described in the literature [2, 8, 17].

Therefore, in order to expand the functionality of the electronic stethoscope [35], we developed and studied in the MATLAB environment an algorithm for indirectly determining systolic and diastolic pressure from ECG and PPG signals received from the fingers, which is based on two different approaches to calculating blood pressure.

1. System for measuring ECG and PPG signals

The MAX86150 Evaluation System equipment was used for the experiments (**Błąd! Nie można odnaleźć źródła odwołania.**). The MAX86150 Evaluation System provides a proven platform to evaluate the MAX86150 integrated photoplethysmogram (PPG) and 1-lead electrocardiogram (ECG) sensor module [34]. The Evaluation System consists of two boards that connect through header pins: a MAX32630FTHR microcontroller board and a MAX86150 evaluation kit. The MAX32630FTHR houses a microcontroller with preloaded firmware, Bluetooth communication, and power management. The sensor board contains the MAX86150 module and two stainless steel dry electrodes for ECG measurement. The Evaluation kit is powered by the included lithium ion battery, which is charged with a micro-USB cable [34]. Based on this evaluation system, we obtained the results of measurements of ECG and PPG signals using only fingers. The measurements were carried out on several patients of different ages (42 years old – patient 1 and 66 years old – patient 2). The obtained measurement results were saved in CSV files, and according to the algorithm proposed below, using the MATLAB platform, the values of systolic (SBP) and diastolic (DBP) pressure of these patients were determined.

2. Mathematical models for determining blood pressure from ECG and PPG signals

As a result of long-term measurement of ECG and PPG signals using fingers on the hand, a database of two patients was obtained, signal fragments from which are shown in Fig. 2.

For clarity and a detailed description of the methodology for determining systolic and diastolic pressure, we have identified one period of the ECG and PPG signals, which is shown in Fig. 3.

Fig. 2 and Fig. 3 show that the PPG signals we received using the MAX86150 Evaluation System have a slightly different shape (inverse) and differ from the PPG signal shapes that were obtained in literature sources [2, 17, 34]. This may be due to the use of different equipment and software when measuring ECG and PPG signals.

To determine the pulse transit time interval, which characterizes the systolic pressure ptt, we calculated the difference between the t_2 interval and the t_1 interval

$$ptt = t_2 - t_1 \quad (1)$$

At the same time, the t_2 interval was defined as the difference between the ECG signal peak and the PPG signal minimum, and the t_1 interval was defined as the difference between the ECG signal peak and the PPG signal maximum (Fig. 3).

Then the number of samples corresponding to the interval ptt between the extrema of the ECG and PPG signals was converted at time ts (ms) taking into account the sampling frequency (which was 200 kHz) according to the formula

$$ts = (ptt/200)*1000 \quad (2)$$

Fig. 1. The MAX86150 Evaluation System equipment was used for the experiments

Taking into account the above formulas (1) and (2), as well as the recommendations of previous studies by other authors [5, 15, 20], we determined the systolic pressure (SBP) using two different formulas

$$SBP_1 = a_1 \cdot ts + b_1 \quad (3)$$

$$SBP_2 = a_2/t_2 + b_2 \quad (4)$$

The diastolic pressure (DSP) of the two patients was also determined using two different formulas

$$DBP_1 = a_1 \cdot t_1 + b_1 \quad (5)$$

$$DBP_2 = a_2/t_1 + b_2 \quad (6)$$

There are a large number of mathematical models for indirectly determining blood pressure from ECG and PPG signals, which are presented in [32]. Various styles of equations are applied, such as linear, quadratic, exponential and others, which are based on different deduction processes. For example, the model (4) reflects the reverse correlation between PTT and SBP shown by large amounts of studies, based on the fact that a high SBP will reduce the time consumed by the pressure pulse to propagate from the proximal to the distal sites, and vice versa [13].

As a result of the analysis and experimental studies of various styles of equations, as well as based on the recommendations of other researchers [13, 15, 20, 32], we have chosen only two models for indirect determination of blood pressure, which are represented by equations (3) – (6). Other styles of equations were either weakly sensitive to changes in blood pressure or did not significantly affect the accuracy of the indirect determination of systolic and diastolic pressure.

In the first method for calculating blood pressure (SBP_1 and DBP_1), the coefficients a1 and b1 of the linear equation (3) were taken equal to $a_1 = -0.5$ and $b_1 = 164$ to determine the systolic pressure of SBP_1, as well as $a_1 = -0.05$ and $b_1 = 125$ to determine the diastolic pressure DBP_1 (5), based on their numerical values previously presented (calculated) in the literature [3, 14].

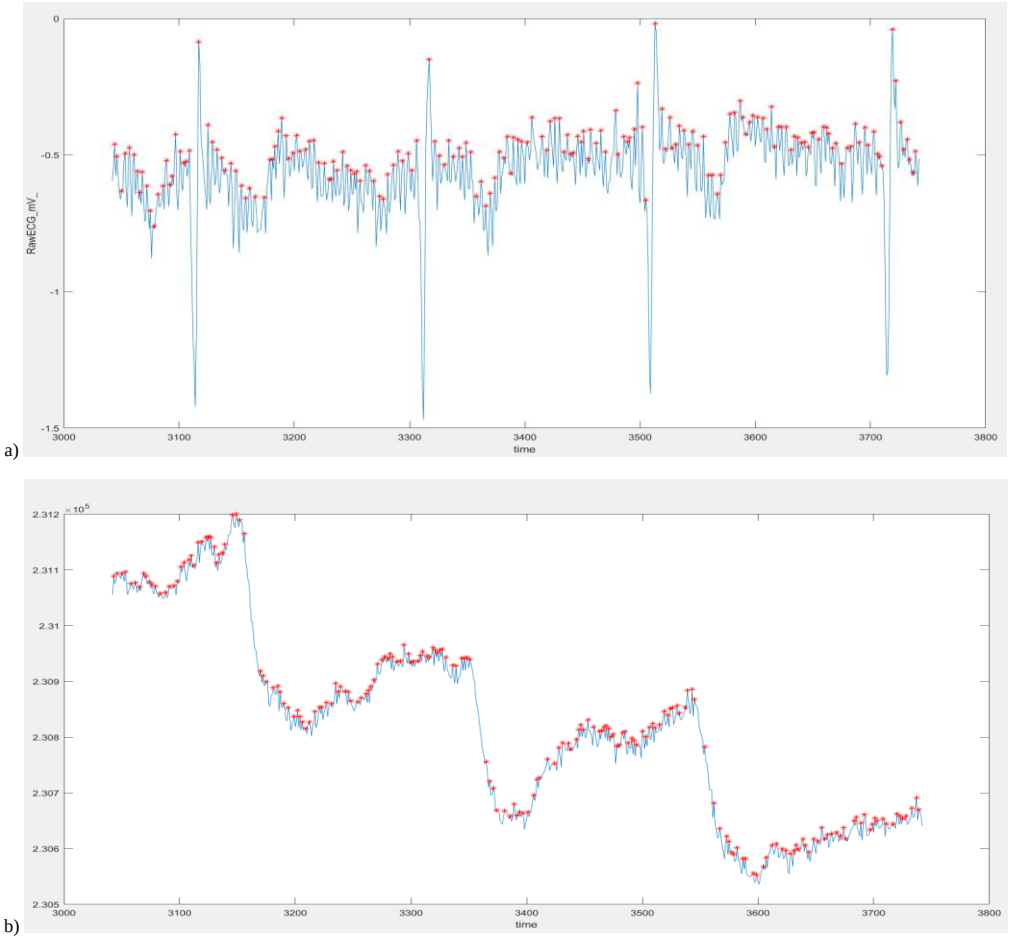


Fig. 2. Illustration a typical waveform of the PPG and ECG: a) the ECG signal; b) the PPG signal

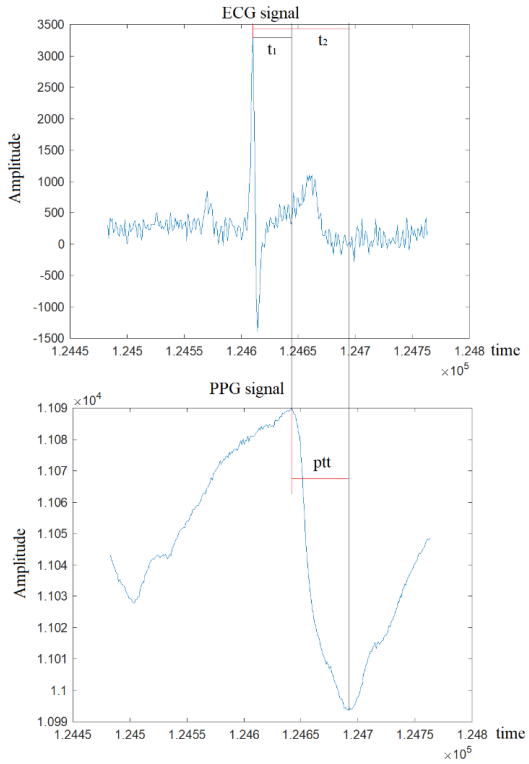


Fig. 3. Calculation of PTT from the time taken for the heart beat pulse to arrive in the finger PPG signal

To calculate blood pressure by the second method (SBP_2 and DBP_2), the values of the coefficients a_2 and b_2 of equations (4) and (6) were calculated based on the experimentally determined values of the signals ptt_S = 640 ms, ptt_D = 140 ms (reference values) and the values of the so-called normal blood pressure SBP_{normal} = 120 mmHg, and DBP_{normal} = 80 mmHg according to the data of the American Heart Association [7, 33] through the solution of the system of equations

$$\begin{cases} SBP_{normal} = \frac{a_2}{ptt_S} + b_2 \\ DBP_{normal} = \frac{a_2}{ptt_D} + b_2 \end{cases} \quad (7)$$

Having solved the system of equations (7) taking into account the known values of normal pressure, the numerical values of the coefficients a_2 = -7170.87 and b_2 = 131.22 were obtained, which were used to process the obtained values of the measured ECG and PPG signals from several healthy people of different ages for determination of blood pressure by formulas (4) and (6) using the MATLAB package.

Based on the principles of dependence of systolic pressure on diastolic pressure, which are detailed in [9, 17, 22, 31] and taking into account the expressions for determining the value of the mean blood pressure (MBP)

$$MBP = SBP/3 + 2 \cdot DBP/3 \quad (8)$$

as well as the aforementioned normal blood pressure values SBP_{normal} = 120 mmHg, and DBP_{normal} = 80 mmHg according to the American Heart Association data [7, 33], which can be used to determine the normal mean blood pressure value: $MBP_{normal} = SBP_{normal}/3 + 2 \cdot DBP_{normal}/3$ = 93.33 mmHg.

Given the normal mean blood pressure value MBP_{normal} from expression (8), an equation was obtained to determine the diastolic pressure DBP through the values of the systolic pressure SBP

$$DBP = 1.5 \cdot (MBP_{normal} - SBP/3) \quad (9)$$

Measurements of blood pressure were carried out and, according to the experimental data obtained above, the values of systolic pressure were determined in two ways – according to the formula (3) and according to the formula

$$SBP_2 = a_2/ts + b_2 \quad (10)$$

Based on certain values of systolic pressure (3) and (10), taking into account equation (9), the values of diastolic pressure were determined in two ways

$$DBP_1 = 1{,}5(MBP_{normal} - [a_1 \cdot ts + b_1]/3) \quad (11)$$

$$DBP_2 = 1{,}5(MBP_{normal} - [a_2/ts + b_2]/3) \quad (12)$$

Based on the above mathematical models (1) – (12) for determining blood pressure, experimental data were obtained for indirect measurement of systolic and diastolic pressure of two patients aged 42 and 66 years. The calculations were carried out in the MATLAB environment and the results of the studies are presented in the next section.

3. Results of measurements of blood pressure

The experimental blood pressure values of two patients of different ages, which are calculated by formulas (3) and (5) are method 1. They are presented in table 1. And those calculated by formulas (4) and (6) are method 2. They are presented in table 2.

Characteristics of changes in systolic and diastolic pressure in two patients of different ages, obtained based on the use of two different methods for determining blood pressure from ECG and PPG signals, are shown in Fig. 4.

The results of a study of another approach in determining blood pressure using linear (3) and non-linear (10) models for determining systolic pressure, as well as mathematical models (11) and (12) for determining diastolic pressure determined based on the mean arterial pressure (MBP) are presented in table 3 and 4.

Table 1. The experimental data to determine the blood pressure of two people – the first method

Method 1			
Patient 1		Patient 2	
SBP [mmHg]	DBP [mmHg]	SBP [mmHg]	DBP [mmHg]
112.5	84	104.5	84
111.75	89	112.5	84
117	86.5	112.25	91.5
119.5	56.5	112.5	96.5
113	86.5	112	81.5
162.25	84	114.5	76.5
112.75	84	111.5	99
113.25	89	112.5	89
112.5	89	112	89
111.5	99	113.25	89
113.75	76.5	111.75	89
113.25	84	112	81
112.75	94	111.25	81.5
112.25	91.5	111.75	84
113	86.5	112	79
113.25	84	111.5	84
117.5	76.5	112	81.5
118	64	111.75	79
117	66.5	111	86.5
117.75	64	112	81.5
Mean	**Mean**	**Mean**	**Mean**
116.7	81.75	111.73	85.35
u_A(SD)	**u_A(SD)**	**u_A(SD)**	**u_A(SD)**
10.99	11.12	1.86	5.84

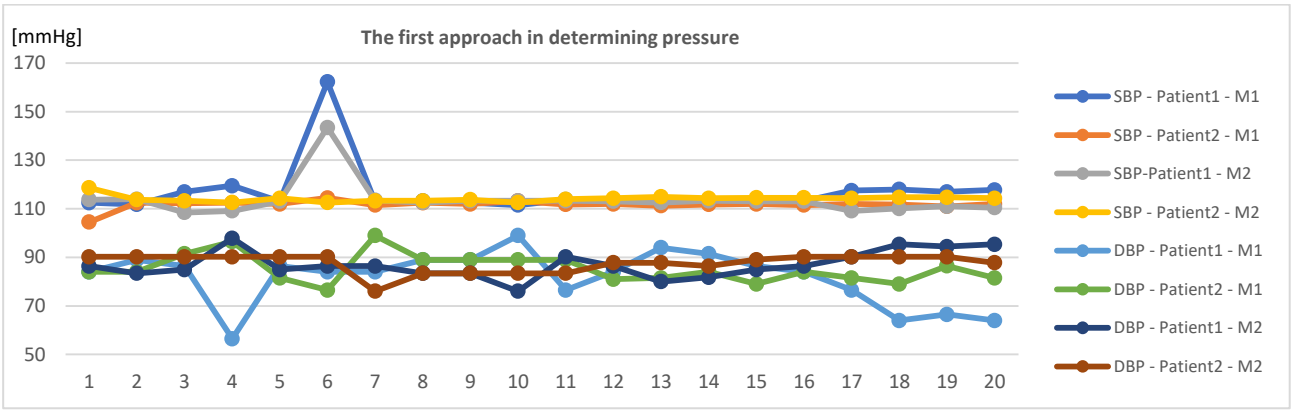


Fig. 4. Characteristics of changes in systolic and diastolic pressure in two patients of different ages

Table 2. The experimental data to determine the blood pressure of two people – the second method

Method 2			
atient 1		Patient 2	
SBP [mmHg]	DBP [mmHg]	SBP [mmHg]	DBP [mmHg]
113.73	86.4	118.64	90.24
113.94	83.41	113.73	90.24
108.46	84.96	113.29	90.24
109.16	97.87	112.59	90.24
113.07	84.96	114.35	90.24
143.48	86.4	112.59	90.24
113.51	86.4	113.29	76.06
112.59	83.41	113.29	83.41
113.29	83.41	113.73	83.41
113.29	76.06	112.59	83.41
113.29	90.24	113.94	83.41
113.07	86.4	114.35	87.76
112.59	80	114.92	87.76
113.29	81.77	114.35	86.4
113.07	84.96	114.54	89.04
113.07	86.4	114.54	90.24
109.16	90.24	114.35	90.24
110.13	95.37	114.74	90.24
111.02	94.45	114.74	90.24
110.43	95.37	114.35	87.76
Mean	**Mean**	**Mean**	**Mean**
113.68	86.92	114.15	87.54
u_A(SD)	**u_A(SD)**	**u_A(SD)**	**u_A(SD)**
7.22	5.54	1.29	3.82

Characteristics of changes in systolic and diastolic pressure in two patients of different ages, obtained on the basis of using a different method for determining blood pressure using the mean arterial pressure (MBP) are shown in Fig. 5.

Based on the results of experimental studies of systolic and diastolic pressure, the uncertainty of type A measurements [1, 4, 10, 12, 16, 18, 19, 21, 23–30] was calculated using the formula

$$u_A = \left[\frac{\sum_i^n (BP_i - \overline{BP})^2}{n(n-1)}\right]^{0.5} \tag{13}$$

Table 3. The experimental data to determine the blood pressure of two people another approach – models (3) and (10)

Method 1			
Patient 1		Patient 2	
SBP [mmHg]	DBP [mmHg]	SBP [mmHg]	DBP [mmHg]
112.5	83.75	112	84
111.75	84.12	111.25	84.37
110	85	111.75	84.12
113	83.5	112	84
112.5	81.5	112.25	83.88
113.25	83.37	112	84
112.75	83.62	112	84
113.25	83.37	112.5	83.75
111.25	84.37	111.75	84.12
112.5	83.75	110	85
111.5	84.25	112.25	83.88
113.75	83.13	112	84
115.25	82.38	111.5	84.25
113.25	83.37	113	83.5
112.5	83.75	112.5	83.75
112.75	83.62	113.75	83.13
112.25	83.88	112.75	83.62
113	83.5	113.25	83.37
113.25	83.37	112	84
112.25	83.88	113.25	83.37
114	83	111.25	84.37
114.75	82.62	112.25	83.75
117.75	81.12	111.5	84.25
Mean	**Mean**	**Mean**	**Mean**
113	83.4	112.12	83.93
u_A (SD)	**u_A(SD)**	**u_A(SD)**	**u_A(SD)**
1.52	0.86	0.79	0.4

Comparing the obtained values of type A measurement uncertainties (standard deviations SD in table 1 – 4), which are obtained using two different approaches to determine systolic and diastolic pressures (table 1 – 4), it can be seen that the deviations are much smaller when using the second approach to determine systolic and diastolic pressures (table 3 and 4).

That is, when determining the systolic pressure SBP through the values of *ts* (the number of readings between the maxima of the ECG and PPG signals) and determining the value of the diastolic pressure DBP based on it, taking into account the normal mean blood pressure value MBP_{normal}.

As the research results showed, when using the first approach to determine blood pressure (Table 1 and 2) according to formulas (3) and (5), (4) and (6), the second method of calculating blood pressure was more accurate - when using formulas (4) and (6). The type A measurement uncertainty in the first patient was 7.22 mmHg for SBP and 5.54 mmHg for DBP. In the second patient, this measurement uncertainty was 1.29 mmHg for SBP and 3.82 mmHg for DBP.

And as the research results showed, when using the second approach to determine blood pressure (Table 3 and 4) using formulas (3) and (11), (10) and (12), the first method for calculating blood pressure was more accurate - when using formulas (3) and (11). The type A measurement uncertainty in the first patient was 1.52 mmHg for SBP and 0.86 mmHg for DBP. In the second patient, this measurement uncertainty was 0.79 mmHg for SBP and 0.4 mmHg for DBP.

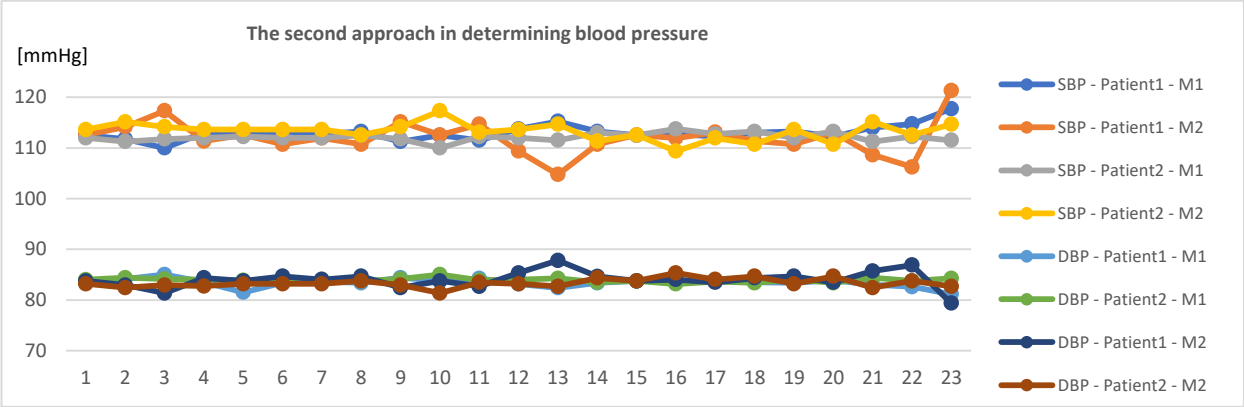


Fig. 5. Characteristics of changes in systolic and diastolic pressure in two patients of different ages using the mean arterial pressure (MBP)

Table 4. The experimental data to determine the blood pressure of two people another approach - models (11) and (12)

Method 2			
Patient 1		Patient 2	
SBP [mmHg]	DBP [mmHg]	SBP [mmHg]	DBP [mmHg]
112.54	83.73	113.64	83.18
114.16	82.92	115.14	82.43
117.32	81.34	114.16	82.92
111.34	84.33	113.64	82.73
112.54	83.73	113.64	83.18
110.71	84.65	113.64	83.18
111.95	84.02	113.64	83.18
110.71	84.65	112.54	83.73
115.14	82.43	114.16	82.92
112.54	83.73	117.32	81.34
114.66	82.67	113.1	83.45
109.35	85.33	113.64	83.18
104.75	87.78	114.66	82.67
110.71	84.64	111.34	84.33
112.54	83.73	112.54	83.73
111.95	84.02	109.35	85.33
113.1	83.45	111.95	84.02
111.34	84.33	110.71	84.65
110.71	84.65	113.64	83.18
113.1	83.45	110.71	84.65
108.63	85.69	115.14	82.43
106.24	86.88	112.54	83.73
121.33	79.34	114.66	82.67
Mean	**Mean**	**Mean**	**Mean**
112.06	83.98	113.28	83.34
u_A(SD)	**u_A(SD)**	**u_A(SD)**	**u_A(SD)**
3.39	1.71	1.72	0.87

4. Conclusions

Studies of indirect determination of systolic and diastolic blood pressure from ECG and PPG signals using the MAX86150 Evaluation System were carried out. The proposed algorithms for determining blood pressure can be used to expand the functionality of an electronic stethoscope. The studies were carried out in the laboratory on two patients of different ages using two different methods for determining blood pressure using the MATLAB pack-age.

Considering the peculiarities of using the MAX86150 Evaluation System for measuring ECG and PPG signals, our group developed mathematical models for indirect determination of systolic and diastolic pressure using fingers on the hand, which were tested in the MATLAB environment. A database of ECG and PPG signals was obtained from two patients aged 42 and 66 years. Based on the proposed mathematical models, ECG and PPG signals were processed in the MATLAB package and the results of indirect measurement of blood pressure were presented (table 1 and 2). The algorithm for determining blood pressure using mathematical models (3) and (11) gives the highest accuracy. In this case, diastolic pressure was determined based on the mean blood pressure.

References

[1] Asgharnezhad H., Shamsi A., Bakhshayeshi I., Alizadehsani R., Chamaani S., Alinejad-Rokny H.: Improving PPG Signal Classification with Machine Learning: The Power of a Second Opinion. In IEEE 24th International Conference on Digital Signal Processing (DSP), 2023, 1–5.

[2] Chao P. C. P., Wu C. C., Nguyen D. H., Nguyen B. S., Huang P. C., Le V. H.: The machine learnings leading the cuffless PPG blood pressure sensors into the next stage. IEEE Sensors Journal 21(11), 2021, 12498–12510.

[3] Chiu Y. C., Arand P. W., Shroff S. G., Feldman T., Carroll J. D.: Determination of pulse wave velocities with computerized algorithms. American heart journal 121(5), 1991, 1460–1470.

[4] Dutt D., Shruthi S.: Digital processing of ECG and PPG signals for study of arterial parameters for cardiovascular risk assessment. In IEEE International conference on communications and signal processing (ICCSP), 2015, 1506–1510.

[5] Fortino G., Giampà V.: PPG-based methods for non invasive and continuous blood pressure measurement: an overview and development issues in body sensor networks. IEEE International Workshop on Medical Measurements and Applications, Ottawa, ON, Canada, 2010, 10–13.

[6] Gómez-Quintana S., Schwarz C. E., Shelevytsky I., Shelevytska V., Semenova O., Factor A., Popovici E., Temko A.: A framework for AI-assisted detection of patent ductus arteriosus from neonatal phonocardiogram. In Healthcare 9(2), 2021, 169.

[7] Haque C. A., Kwon T.-H., Kim K.-D.: Cuffless Blood Pressure Estimation Based on Monte Carlo Simulation Using Photoplethysmography Signals. Sensors 22, 2022, 1175.
[8] Kachuee M., Kiani M. M., Mohammadzade H., Shabany M.: Cuffless blood pressure estimation algorithms for continuous health-care monitoring. IEEE Transactions on Biomedical Engineering 64(4), 2016, 859–869.
[9] Kao Y. H., Chao P. C. P., Wey C. L.: Design and validation of a new PPG module to acquire high-quality physiological signals for high-accuracy biomedical sensing. IEEE J. Sel. Top. Quantum Electron 25, 2019, 18159167.
[10] Liang Y., Chen Z., Ward R., Elgendi M.: Hypertension assessment via ECG and PPG signals: An evaluation using MIMIC database. Diagnostics 8(3), 2018, 65.
[11] Man P. K., Cheung K. L., Sangsiri N., Shek W. J., Wong K. L., Chin J. W., Chan T. T., So R. H. Y.: Blood Pressure Measurement: From Cuff-Based to Contactless Monitoring. In Healthcare 10(10), 2022, 2113.
[12] Morresi N., Casaccia S., Sorcinelli M., Arnesano M., Revel G.: Analysing performances of Heart Rate Variability measurement through a smartwatch. In 2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA), 2020, 1–6.
[13] Mukkamala R., Hahn J. O., Inan O. T., Mestha L. K., Kim C. S., Töreyin H., Kyal S.: Toward Ubiquitous Blood Pressure Monitoring via Pulse Transit Time: Theory and Practice. In IEEE Transactions on Biomedical Engineering 62(8), 2015, 1879–1901.
[14] Payne R. A., Symeonides C. N., Webb D. J., Maxwell S. R.: Pulse transit time measured from the ECG: An unreliable marker of beat-to-beat blood pressure. J. Appl. Physiol. 100, 2006, 136–141.
[15] Pour Ebrahim M., Heydari F., Wu T., Walker K., Joe K., Redoute J. M., Yuce M. R.: Blood pressure estimation using on-body continuous wave radar and photoplethysmogram in various posture and exercise conditions. Scientific Reports 9(1), 2019, 1–13.
[16] Rundo F., Petralia S., Fallica G., Conoci S.: A nonlinear pattern recognition pipeline for PPG/ECG medical assessments. In Convegno Nazionale Sensori, 2018, 473–480.
[17] Samimi H., Dajani H. R.: Cuffless Blood Pressure Estimation Using Calibrated Cardiovascular Dynamics in the Photoplethysmogram. Bioengineering 9(9), 2022, 446.
[18] Semenov A., Osadchuk O., Semenova O., Bisikalo O., Vasilevskyi O., Voznyak O.: Signal Statistic and Informational Parameters of Deterministic Chaos Transistor Oscillators for Infocommunication Systems. 2018 International Scientific-Practical Conference Problems of Infocommunications Science and Technology, 2019, 8632046, 730–734.
[19] Shabaan M., Arshid K., Yaqub M., Jinchao F., Zia M., Bojja G., Iftikhar M., Ghani U., Ambati L., Munir R.: Survey: smartphone-based assessment of cardiovascular diseases using ECG and PPG analysis. BMC medical informatics and decision making, 2020, 1–6.
[20] Sharma M., Barbosa K., Ho V., Griggs D., Ghirmai T., Krishnan S. K., Hsiai T. K., Chiao J. C., Cao H.: Cuff-less and continuous blood pressure monitoring: a methodological review. Technologies 5(2), 2017, 21.
[21] Trishch R., Nechuiviter O., Dyadyura K., Vasilevskyi O., Tsykhanovska I., Yakovlev M.: Qualimetric method of assessing risks of low quality products. MM Science Journal 2021(4), 2021, 4769–4774.
[22] Tseng T. J., Tseng C. H.: Cuffless blood pressure measurement using a microwave near-field self-injection-locked wrist pulse sensor. IEEE Trans. Microw. Theory Tech 68, 2020, 4865–4874.
[23] Vasilevskyi O. M., Yakovlev M. Y., Kulakov P. I.: Spectral method to evaluate the uncertainty of dynamic measurements. Technical Electrodynamics 4, 2017, 72–78.
[24] Vasilevskyi O. M.: A frequency method for dynamic uncertainty evaluation of measurement during modes of dynamic operation. International Journal of Metrology and Quality Engineering 6(2), 2015, 202.
[25] Vasilevskyi O. M.: Assessing the level of confidence for expressing extended uncertainty: a model based on control errors in the measurement of ion activity. Acta IMEKO 10(2), 2021, 199–203.
[26] Vasilevskyi O. M.: Calibration method to assess the accuracy of measurement devices using the theory of uncertainty. International Journal of Metrology and Quality Engineering 5(4), 2014, 403.
[27] Vasilevskyi O. M.: Metrological characteristics of the torque measurement of electric motors. International Journal of Metrology and Quality Engineering 8, 2017, 7.
[28] Vasilevskyi O., Koval M., Kravets S.: Indicators of reproducibility and suitability for assessing the quality of production services. Acta IMEKO 10(4), 2021, 54–61.
[29] Vasilevskyi O., Kulakov P., Kompanets D., Lysenko O. M., Prysyazhnyuk V., Wójcik W., Baitussupov D.: A new approach to assessing the dynamic uncertainty of measuring devices. Proc. SPIE 10808, 2018, 728–735.
[30] Vasilevskyi O., Voznyak O., Didych V., Sevastianov V., Ruchka O., Rykun V.: Methods for Constructing High-precision Potentiometric Measuring Instruments of Ion Activity. In 2022 IEEE 41st International Conference on Electronics and Nanotechnology (ELNANO), 2022, 247–252.
[31] Wang H. S. J., Yeh M. H., Chao P. C. P., Tu T. Y., Kao Y. H., Pandey R.: A fast chip implementing a real-time noise resistant algorithm for estimating blood pressure using a non-invasive, cuffless PPG sensor. Microsyst. Technol 26, 2020, 3501–3516.
[32] Zhang Q., Zeng X., Hu W., Zhou D.: A Machine Learning-Empowered System for Long-Term Motion-Tolerant Wearable Monitoring of Blood Pressure and Heart Rate With Ear-ECG/PPG. In IEEE Access 5, 2017, 10547–10561.
[33] American Heart Association. [https://www.heart.org/en/] (access 08/07/2023).
[34] AnalogDevices Homepage [https://www.analog.com/media/en/technical-documentation/data-sheets/MAX86150EVSYS.pdf] (access 2023/08/07).
[35] THINKLABS Homepage [https://www.thinklabs.com/] (access 2023/08/07).

D.Sc. Eng. Oleksandr Vasilevskyi
e-mail: oleksandr.vasilevskyi@austin.utexas.edu

Doctor of Technical Sciences, professor, senior research of the Department of Mechanical Engineering, University of Texas at Austin (USA). The badge "For scientific and educational achievements" (2020). Laureate of the Prize of the Verkhovna Rada of Ukraine for young scientists (2019). Excellence in Education of Ukraine. Scholarship holder of the Cabinet of Ministers of Ukraine for young scientists (2018–2020). Academician of the Academy of Metrology of Ukraine, an official representative from Ukraine in the IMEKO. Author of more than 215 publications, including 3 monographs, 3 collective monographs, 17 textbooks, 5 patents for inventions, more than 110 scientific articles in peer-reviewed journals, 25 of them in the Scopus (h-index = 13).

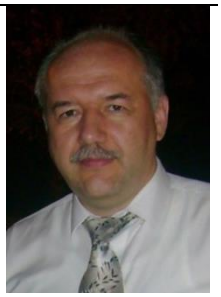
http://orcid.org/0000-0002-8618-0377

Dr. Emanuel Popovici
e-mail: e.popovici@ucc.ie

Ph.D., associate professor of the Department of Electrical and Electronic Engineering at University College Cork (UCC). His research interests include low power embedded systems design for efficient, reliable and secure computing and communications. He has published more than 176 papers in the Scopus (h-index = 26) and co-authored more than 10 papers distinguished by the IEEE, IET, MIDAS, and IARIA. His students achieved more than 20 awards and distinctions. His research is sponsored by SFI, Entersprise Ireland, IDA/Synopsys, NSF, Invest-NI, EU-FP7, Industry, IRCSET/IRC.

http://orcid.org/0000-0001-6813-5030

Dr. Volodymyr Sarana
e-mail: vsarana@ucc.ie



Ph.D., associate professor, Post-Doctoral Researcher of the Department of Electrical and Electronic Engineering at University College Cork (UCC). Author of more than 20 publications, including 2 textbooks and more than 15 scientific articles in professional journals, of which 5 are in scientometric databases Scopus and Web of Science.

http://orcid.org/0000-0002-7778-3176

http://doi.org/10.35784/iapgos.5319 *received: 24.07.2023 | revised: 06.09.2023 | accepted: 27.09.2023 | available online: 30.09.2023*

SEGMENTATION OF CANCER MASSES ON BREAST ULTRASOUND IMAGES USING MODIFIED U-NET

Ihssane Khallassi, My Hachem El Yousfi Alaoui, Abdelilah Jilbab
Mohammed V University in Rabat, Electronic Systems Sensors and Nanobiotechnology, National School of Arts and Crafts, Rabat, Morocco

Abstract. *Breast cancer causes a huge number of women's deaths every year. The accurate localization of a breast lesion is a crucial stage. The segmentation of breast ultrasound images participates in the improvement of the process of detection of breast anomalies. An automatic approach of segmentation of breast ultrasound images is presented in this paper, the proposed model is a modified u-net called Attention Residual U-net, designed to help radiologists in their clinical examination to determine adequately the limitation of breast tumors. Attention Residual U-net is a combination of existing models (Convolutional Neural Network U-net, the Attention Gate Mechanism and the Residual Neural Network). Public breast ultrasound images dataset of Baheya hospital in Egypt is used in this work. Dice coefficient, Jaccard index and Accuracy are used to evaluate the performance of the proposed model on the test set. Attention residual u-net can significantly give a dice coefficient = 90%, Jaccard index = 76% and Accuracy = 90%. The proposed model is compared with two other breast segmentation methods on the same dataset. The results show that the modified U-net model was able to achieve accurate segmentation of breast lesions in breast ultrasound images.*

Keywords: convolutional neural network, segmentation, u-net, residual neural network

SEGMENTACJA MAS NOWOTWOROWYCH NA OBRAZACH ULTRASONOGRAFII PIERSI Z UŻYCIEM ZMODYFIKOWANEGO MODELU U-NET

Streszczenie. *Każdego roku rak piersi powoduje ogromną liczbę zgonów kobiet. Dokładna lokalizacja zmiany piersi jest kluczowym etapem. Segmentacja obrazów ultrasonograficznych piersi przyczynia się do poprawy procesu wykrywania nieprawidłowości piersi. W tym artykule przedstawiono automatyczne podejście do segmentacji obrazów ultrasonograficznych piersi, proponowany model to zmodyfikowany U-net, nazwany Attention Residual U-net, zaprojektowany w celu wspomagania radiologów podczas badania klinicznego, w celu odpowiedniego określenia zasięgu guzów piersiowych. Attention Residual U-net jest połączeniem istniejących modeli (konwolucyjną siecią neuronową U-net, Attention Gate Mechanism i Residual Neural Network). W tym badaniu wykorzystano publiczny zbiór danych obrazów ultrasonograficznych piersi szpitala Baheya w Egipcie. Do oceny wydajności zaproponowanego modelu na zbiorze testowym wykorzystano współczynnik Dice'a, indeks Jaccarda i dokładność. Attention Residual U-net może znacznie przyczynić się do uzyskania współczynnika Dice'a równego 90%, indeksu Jaccarda równego 76% i dokładności równiej 90%. Proponowany model został porównany z dwoma innymi metodami segmentacji piersi na tym samym zbiorze danych. Wyniki pokazują, że zmodyfikowany model U-net był w stanie osiągnąć dokładną segmentację zmian piersiowych na obrazach ultrasonograficznych piersi.*

Słowa kluczowe: konwolucyjna sieć neuronowa, segmentacja, u-net, rezydualna sieć neuronowa

Introduction

Breast cancer is the first causes of death by cancer in women, known as the most detected cancer in the world by recent years, it appears most of the time as a mass in the breast [5, 13]. According to the world health organization early diagnosis of breast cancer can have a high chance of cure and reduce the mortality rate by 40% [7].

For the diagnosis of breast anomalies ultrasound images are often used to create a clear and precise representation of the breast, and they have shown important efficacy in the detection of breast cancer [2, 15].

The purpose of Lesion segmentation in breast ultrasound images is to localize, detect breast tumors in ultrasound images and to assist specialists in their clinical examination instead of taking more time on a single patient because manual analysis for breast ultrasound images needs more works and experience. Manual segmentation depends on manual tracing or visual inspection of the abnormal region requires various medical experts to select exactly the abnormal regions, particularly for large-scale breast ultrasound images which leads to a subjective interpretation between observers that makes the breast segmentation process fastidious and time-consuming task. Automatic segmentation that does not need human intervention which is often used to help radiologists to detect breast masses, reduce errors, increase precision and time control.

Due to the interest in the medical field, the research on automatic segmentation on breast ultrasound images has taken the main focus and has been enhanced during the last decades. Diverse algorithms have been proposed for the segmentation of breast ultrasound images: classical segmentation (edge-based segmentation, threshold-based segmentation, region-based segmentation), machine learning segmentation (unsupervised segmentation, supervised segmentation) [6].

Deep learning segmentation has proven to be very useful for in medical image processing, it is known for its ability to extract high level features by following layer by layer from a raw input, and it leads to excellent segmentation performances. In this paper I will perform automatic segmentation of lesions in breast ultrasound images. The proposed architecture to achieve this goal is based on the popular models including U-Net [16], RESNET [8], Attention Gate [14] which are more used in medical imaging.

The content of the article is structured as bellow, section 1 dedicated to the presentation of the public dataset of breast ultrasound images, the detailed structure of proposed model for segmentation, training and evaluation, section 2 is dedicated to present the results and compare them with two other models. In section 3 the conclusion is presented.

1. Materials and method

1.1. Data collection

The dataset used is the first publicly available breast ultrasound dataset [1] collected in DICOM format at the Baheya hospital, Cairo, Egypt. The data contains 780 images with an average image size of 500*500, acquired in 2018 from 600 female patients between 25 and 75 years old. the images are in PNG format, each images contains its ground truth performed carefully by radiologists. The images are classified into three categories: normal, benign and malignant. This data set is usually in gray level, the tools used in the procedure of digitization are the LOGIQ E9 ultrasound scanner and the LOGIQ E9 agile ultrasound scanner.

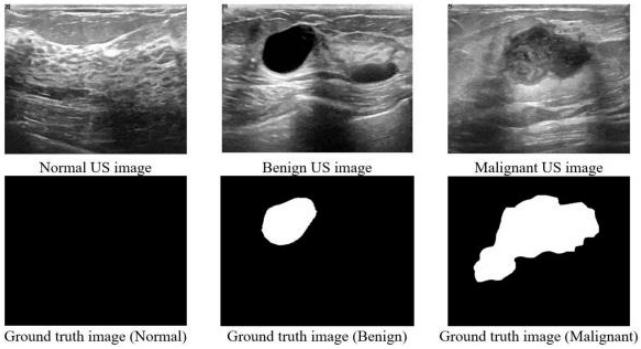


Fig. 1. Breast ultrasound images samples and their ground truth images

1.2. Method

The proposed model for lesion segmentation in breast ultrasound images inspired on mostly common and popular architectures in medical imaging U-Net [16], Attention Gate [14], Residual neural network [8].

1.2.1. Deep convolutional network

U-net

U-net is considered as a convolutional neural network (CNN) most used for semantic segmentation, designed by Olaf Ronneberger et al. in 2015 [16]. Having proven itself in terms of accuracy and speed especially in the medical imaging field. It consists of three sections: the contraction path or encoder, the bottleneck and the expansion path or decoder

The contraction path (encoder) which is used to capture the context and extract the spatial features in the input image, it follows the architecture typical of a convolution network, it is composed of four blocks, each block is composed of two convolution layers (3×3), each layer is followed by a rectified linear unit (Relu), Then a maximum pooling operation of (2×2) with a stride of 2 is applied for down-sampling and keeping the important features that better describes the context of the image. The contraction phase increases the contextual information that defines the nature of an object. During this contraction path the number of feature channels is doubled while the spatial dimension is reduced by half to decrease the number of network parameters.

The expansion path (decoder) is the second part of the U-Net architecture that is used to semantically project the discriminative features learned by the encoder onto the pixel space, it retrieves the object details and recovers the initial image size and localizes the lesion.

The expansion path starts with an up-sampling of the feature map generated by the bridge. The up-sampling operation reduces the number of feature channels by half and increases the size of the image, it also considers as concatenation part, using U-Net skip connections concatenates a feature part of the encoder with the decoder, for each block in the encoder the result of the convolution operation before maximum pooling is transferred to the decoder symmetrically.

Each decoder block receives the learned feature representation from the encoder and concatenates it with the output of the up-sampling operation, followed by two convolution layers (3×3), each of them is reinforced by an activation function (Relu).

The concatenation helps the expansion feature to recover the location information of the respective object and acquire the general information that combines the context and location. It is also useful to capture the possible features lost by the maximum pooling.

At the output of the decoder a convolution (1×1) with an activation function (Sigmoid) is applied. During the expansion phase the image size gradually increases while the number of channels gradually decreases.

The bridge or bottleneck connects the encoder to the decoder and completes the information flow, it consists of two convolutional layers (3×3), each layer is followed by an activation function (Relu). This information is important to obtain a segmentation map with high resolution.

The skip connections get the characteristics from the encoder and concatenate them with the decoder, providing the decoder with sufficient context to create an efficient segmentation mask.

The parameters of the U-Net are described as follows:

- the size of the BUS image in the input and in the output of the U-Net is 256×256×1,
- in all convolution operations the ""same"" mode is used place of ""valid"" to have the same input size at the output,
- the activation function (Relu) is applied in all convolution layers, except the final convolution layer the activation function (Sigmoid) is applied,
- the number of filters for the first layer in the encoder is 64, and the number of filters in the contraction phase is doubled after each maximum pooling operation.

Attention gate

A model that contains attention gate units in its architecture allows to focus attention on target regions and to remove insignificant regions from an input image. Attention gate units can be easily implemented in the skip connections of the U-net model while increasing model sensitivity and prediction accuracy. The results show that the attention gate units improve the performances of the U-net model while maintaining the computational efficiency [14].

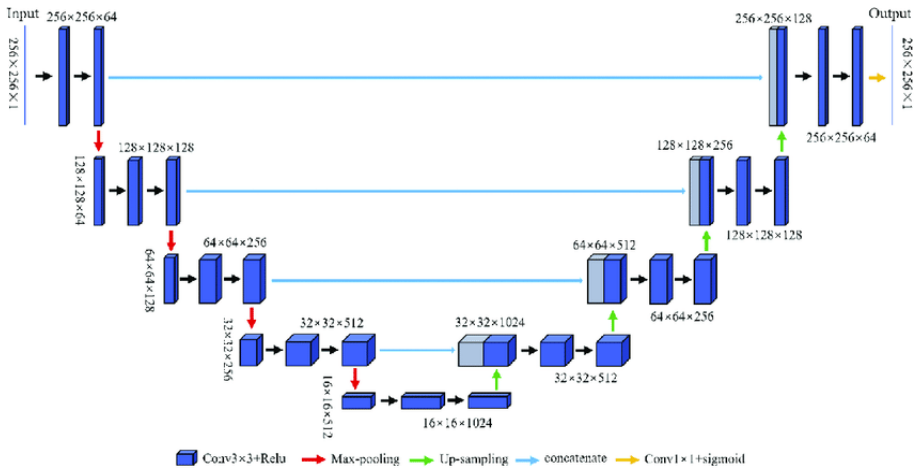


Fig. 2. U-net architecture

In this paper the attention gate mechanism integrated with the skip connections of the U-net model architecture to focus the concentration on the target breast lesion regions and remove the interference of insignificant objects on the generated segmentation map.

The typical architecture of an attention gate unit can be modelled as shown in Fig. 3.

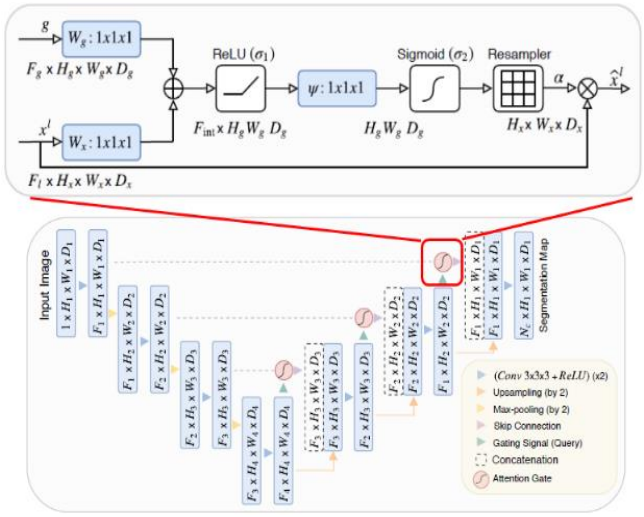


Fig. 3. Schematic of attention gate unit

Residual neural network

Residual network constitutes residual blocks proposed in 2015 by Kaiming He et al [8]. The growth of the depth of a deep neural network can hinder its training process and cause the problem of precision degradation which influence negatively of the performance of the model.

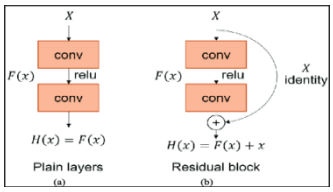


Fig. 4. Residual convolutional block

Replace the basic convolution blocks (plain layer) in the primitive U-Net architecture with residual convolution blocks (Fig. 4) by adding a skip connection between two convolution layers (3×3) helps to build a deeper network without worrying about the problem of gradient loss.

The skip connections used in the network contribute to a better flow of information between the different layers which contributes to a better gradient flow during the network formation.

1.2.2. Architecture of the proposed model

The proposed model for lesion segmentation in breast ultrasound images can be modelled with a combination of the popular existing models U-Net [16], Residual neural network [8] and t the Attention Gate Module [14].

To achieve more exciting performance of lesion segmentation in breast ultrasound images, we integrate attention gates and residual convolution blocks into the U-Net architecture to give the "Residual Attention U-Net" model that that can extract more accurate information about dense features and efficiently restore spatial information and location details.

The input size for the network is 256×256×1, where the size of each image is 256×256 and the number of channels is 1. The contracted path is formed by 4 residual blocks that replace the simple convolution blocks at the architecture of the U-Net, and each residual convolution block contains two convolution layers (3×3) reinforced by a rectified linear unit (Relu), then a skip connection adds the input of the block with the convolution output, followed by a rectified linear unit (Relu).

The size of the first layer is 256×256×64 in the contraction path. During the contraction phase the number of feature map channels is doubled while the feature map size is decreased by half. The bridge connects the decoder to the encoder.

The expansion path contains 4 residual blocks to replace the simple convolution blocks in U-Net and 4 attention gates to improve the information about salient features while disambiguating the responses of irrelevant and noisy features.

The gating signal and feature map obtained from the fourth residual block are combined via the attention gate unit, obtaining more relevant tumor location information.

Before the individual residual block in the expansion path, there is an up-sampling operation that reduces the number of feature channels by half and doubles the size of the image, then a concatenation performed between the output of up-sampling and the output of the attention gate.

At the last layer of the expansion path using a convolution (1×1) with the activation function (sigmoid). The output size for the network is 256×256×1.

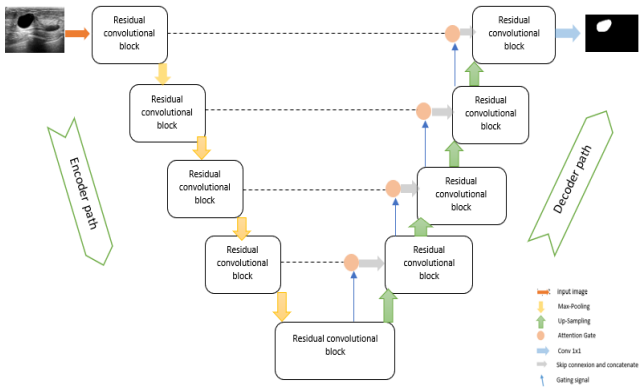


Fig. 5. Attention Residual U-net

1.2.3. Training

The model is trained on the public data set [1] which are partitioned respectively into training data represent 80% and test data which represent 20%. In the training of the network the dimension of the image and its corresponding mask were a 256×256 pixels, the Adam optimizer [10] was used to train the network, using (Accuracy, DICE coefficient [13], Jaccard index [9]) as measures of the accuracy of the segmentation procedure. While (DICE loss [12]) was used as a loss function that was repropagated through the network. The batch size was fixed at eight and the model was trained in 100 epochs and the learning rate was initially 6.1284e-8. The training and evaluation are executed in python using tensorflow [11] and Keras [3] libraries. The number of trainable parameters is 34,540,867.

1.2.4. Data augmentation

Since the number of medical images is limited, data augmentation techniques were used by applying different deformation on the training data using ImageDataGenerator.

The application of these deformation to the original image does not change the target region but only provides a new perspective. The image augmentation technique helps to increase the accuracy and robustness of the proposed model.

The deformation operations used in this work are: rotation by 50 degrees, flipping and shifting.

1.2.5. Experiment

This experiment was performed on a game computer equipped with a NIVIDA GPU GTX 1080 TI with 11GB and up to 24GB of GDDR5 memory intel core i7 7700K CPU @5.00 GHz overclocked processor, 32GB with 3200MHZ frequency of DRR4 Ram, running under Windows 10.

1.2.6. Evaluation metrics

The model performance was evaluated using the three-evaluation metrics commonly used to validate the segmentation task of medical images.

Dice coefficient

It considers as spatial overlap index between two segmentations; it presents the degree of similarity between the segmented output and the ground truth.

$$dice\ coefficient = \frac{2*|X \cap Y|}{|X| + |Y|} \quad (1)$$

We can also notice that the Dice Loss is formulated as:

$$\text{Dice Loss} = 1\text{-Dice Coefficient} \quad (2)$$

X refers to the set of pixels in the lesion region manually delineated by the radiologist, while Y refers to the lesion region automatically generated by the model.

Jaccard index or intersection-over-union

IOU is used to compare the similarity between two se segmentations, presents the ratio of the intersection and union of the ground truth and the segmented output of the model.

$$Iou = \frac{|X \cap Y|}{|X| \cup |Y|} \quad (3)$$

X refers to the set of pixels in the lesion region manually delineated by the radiologist, while Y refers to the lesion region automatically generated by the model.

Accuracy

Common evaluation index for medical image segmentation, it considers as a ratio between the number of correctly predicted pixels and the total number of pixels in the image.

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \quad (4)$$

TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative.

2. Results and discussion

The 3 models (U-Net, Attention U-Net, Residual Attention U-Net) are trained by the same dataset and parameters training and evaluated by the same evaluation metrics.

The Attention Res U-net model achieves exiting performance in lesion segmentation in breast ultrasound images, and it allows to focus attention on significant and important salient regions and disambiguate feature responses and irrelevant regions. The segmentation map generated by the U-net model imports a lot of unnecessary background information that interferes with the target object because at the beginning of the encoder path of the u-net network the feature representation is quite poor.

Table 1. Comparison of the performance of the three models after 100 epochs

model	dice%	iou%	accuracy%
Attention-Residual U-net	90%	76%	90%
Attention U-net	70%	63%	90%
U-net	52%	37%	90%

According to the table above, the best performance was obtained by the Attention Res U-Net model.

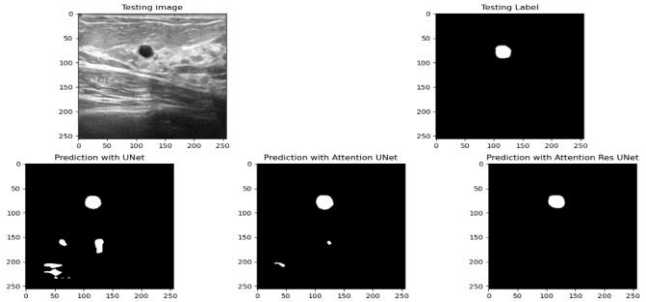


Fig. 6. The difference between ground truth mask with the prediction output of the three models

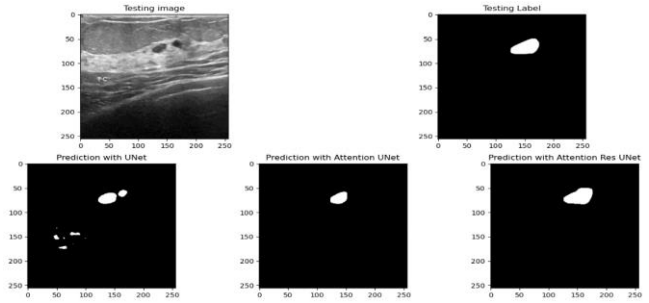


Fig. 7. The difference between ground truth mask with the prediction output of the three models

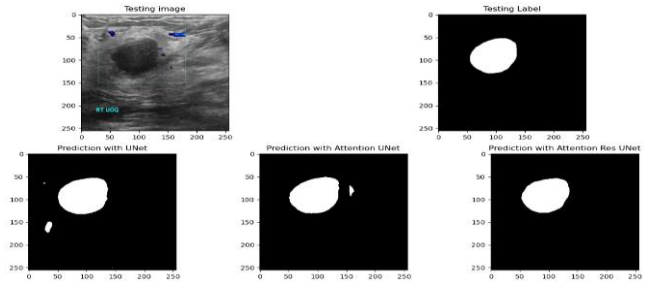


Fig. 8. The difference between ground truth mask with the prediction output of the three models

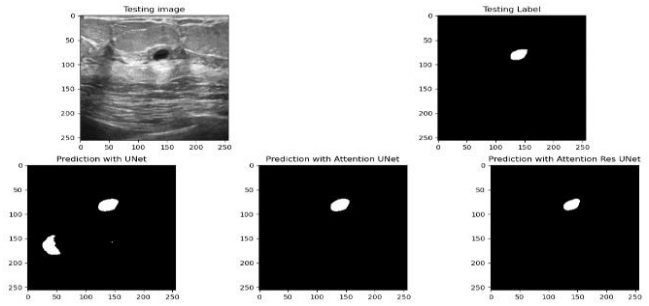


Fig. 9. The difference between ground truth mask with the prediction output of the three models

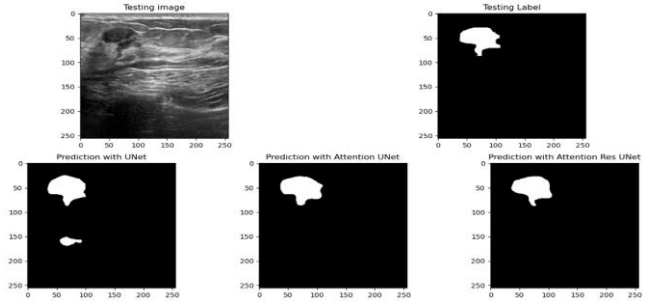


Fig. 10. The difference between ground truth mask with the prediction output of the three models

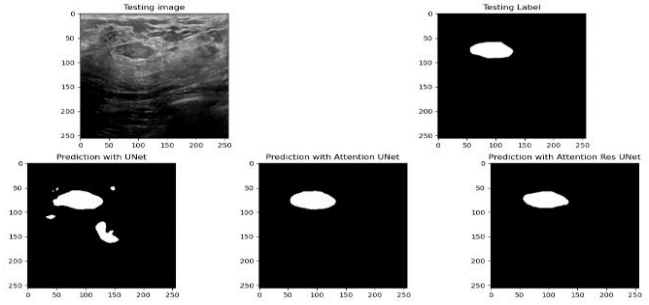


Fig. 11. The difference between ground truth mask with the prediction output of the three models

2.1. Attention residual U-net learning curve

A learning curve is simply a graph showing the progression over time of a specific learning-related metric during the training of a machine learning model. It is simply a mathematical representation of the learning process.

From the graphs of accuracy and loss, we are able to see that in each learning strategy, the model was quickly trained, as the trend in accuracy is up and the loss trend is down.

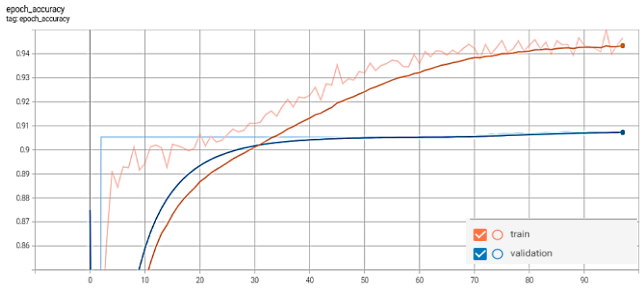


Fig. 12. Learning curves with accuracy over epochs

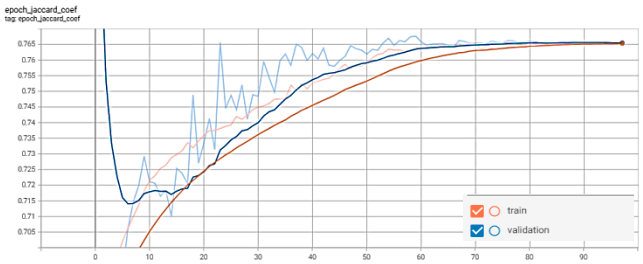


Fig. 13. Learning curves with jaccard index over epochs

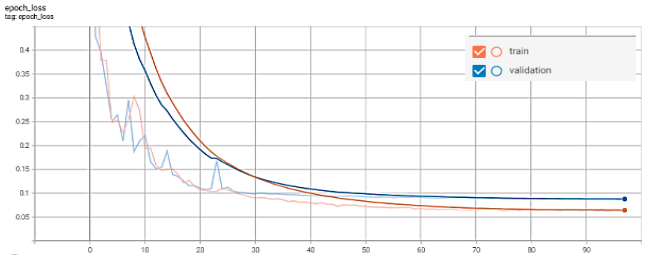


Fig. 14. Learning curves with loss over epochs

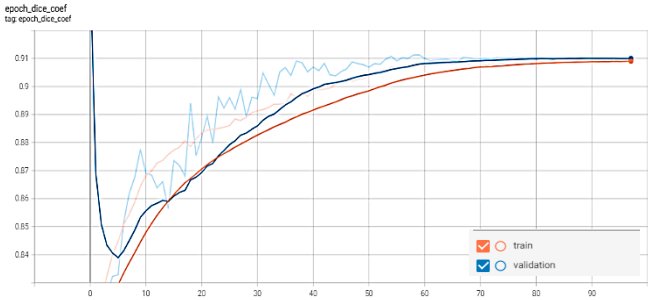


Fig. 15. Learning curves with dice over epochs

3. Conclusion

In this paper, an automated approach for lesion segmentation in breast ultrasound images is presented. The proposed method directly takes a breast ultrasound image to obtain lesion segmentation map.

The proposed model based on a combination of existing popular models U-Net, Residual neural network and Attention Gate module . To achieve more promising performance of lesion segmentation in breast ultrasound images, we integrate attention gates and residual convolution blocks into the U-Net architecture to produce the "Residual Attention U-Net" model that can extract more accurate dense feature information and well restore spatial information and location details.

The proposed model Residual Attention U-Net achieved 90% in terms of DICE and 76% in terms of Jaccard index and 90% in terms of Accuracy. Residual Attention U-Net model has been compared to two models (U-Net and Attention U-Net). Even though the proposed model shows good performance, there is certain points that can be developed in the future, also the proposed model can be applied on other types of medical images for segmentation.

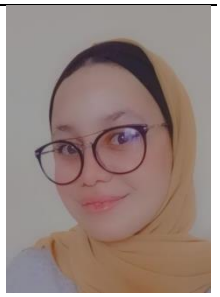
References

[1] AI-Dhabyani W. et al.: Dataset of breast ultrasound images. Data in Brief 28, 2020, 104863.
[2] Balagalla B et al.: A Review On Ultrasound Image Pre-Processing, Segmentation And Compression For Enhanced Image Storage And Transmission. 11-th International Research Conference General, Sir John Kotelawala Defence University, 2018, 106–111.
[3] Chollet F.: Kears. https://github.com/fchollet/keras, 2015.
[4] Dice L. R.: Measures of the Amount of Ecologic Association Between Species. Ecology 26(3), 1945, 297–302.
[5] El Fouhi M. M. et al.: Profil épidémiologique et anatomopathologique du cancer de sein au CHU Ibn Rochd, Casablanca. Pan Afr Med J. 37, 2020, 41.
[6] Epimack M. et al.: Breast Cancer Segmentation Methods: Current Status and Future Potentials. BioMed Research International 2021, 9962109.
[7] Fondation lalla salma, détection précoce 2022.
[8] He K. et al.: Deep Residual Learning for Image Recognition. 2016, arXiv:1512.03385.
[9] Jaccard P.: The Distribution of the Flora in the Alpine Zone. The New Phytologist 11(2), 1912, 37–50.
[10] Kingma D. P., Ba J. L.: Adam: A Method For Stochastic Optimization. 3rd International Conference for Learning Representations, San Diego, 2015, arXiv:1412.6980.
[11] Mart´ın A. et al.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. 2016, arXiv:1603.04467.
[12] Milletari F. et al.: V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. 2016, arXiv:1606.04797.
[13] Nations Unies, ONU Info: Le cancer du sein est désormais le plus fréquent au monde, 2021 [https://news.un.org/fr/story/2021/02/1088502].
[14] Oktay O. et al.: Attention U-Net: Learning Where to Look for the Pancreas, 2018, arXiv:1804.03999.
[15] Prasad S. N., Houserkova D.: A Comparison Of Mammography And Ultrasonography In The Evaluation Of Breast Masses. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub. 151(2), 2007, 315–322 [http://doi.org/10.5507/bp.2007.054].
[16] Ronneberger O. et al.: U-Net: Convolutional Networks for Biomedical Image Segmentation, 2015, arXiv:1505.04597.

M.Sc. Ihssane Khallassi
e-mail: ihssanekhallassi639@gmail.com

Received the master's degree in Electrical Engineering from the National School of Arts and Crafts of Rabat, Mohammed V University, Rabat, Morocco, in 2022. She is a research student of sciences and Technology of the engineer in Ecole National Supérieure d'Informatique et d'Analyse des Systèmes ENSIAS. She is part of E2SN research team at ENSAM RABAT. She is interested in the segmentation of cancer masses on breast ultrasound images using modified u-net.

http://orcid.org/0009-0006-7965-0269



Prof. Abdelilah Jilbab
e-mail: a_jilbab@yahoo.fr

Abdelilah Jilbab graduated in electronic and industrial computer aggregation in 1995. He received the Ph.D. degree in computer and telecommunication from Mohammed V-Agdal University, Rabat, Morocco, in 2009. He is a Professor at ENSAM, Mohammed V University, Rabat, Morocco. Since 2003, he is a member of the laboratory LRIT (unit associated with the CNRST, FSR, Mohammed V University, Rabat, Morocco). His domains of interest include signal processing and embedded systems.

http://orcid.org/0000-0002-1577-9040

Prof. My Hachem El Yousfi Alaoui
e-mail: h.elyousfi@um5r.ac.ma

He is a Professor at ENSAM, Mohammed V University, Rabat, Morocco. He is a member of Electromechanical, Diagnosis and Control Research Team – Electrical Engineering Laboratory – ENSAM, Mohamed V University in Rabat, Morocco. He is a member of E2SN research team at ENSAM RABAT.

http://orcid.org/0000-0003-4285-0540

http://doi.org/10.35784/iapgos.3685 *received: 29.05.2023 | revised: 19.07.2023 | accepted: 25.09.2023 | available online: 30.09.2023*

CLASSIFICATION OF PARKINSON'S DISEASE AND OTHER NEUROLOGICAL DISORDERS USING VOICE FEATURES EXTRACTION AND REDUCTION TECHNIQUES

Oumaima Majdoubi, Achraf Benba, Ahmed Hammouch
Mohammed V University in Rabat, National School of Arts and Crafts, Electronic Systems Sensors and Nanobiotechnology, Rabat, Morocco

Abstract. *This study aimed to differentiate individuals with Parkinson's disease (PD) from those with other neurological disorders (ND) by analyzing voice samples, considering the association between voice disorders and PD. Voice samples were collected from 76 participants using different recording devices and conditions, with participants instructed to sustain the vowel /a/ comfortably. PRAAT software was employed to extract features including autocorrelation (AC), cross-correlation (CC), and Mel frequency cepstral coefficients (MFCC) from the voice samples. Principal component analysis (PCA) was utilized to reduce the dimensionality of the features. Classification Tree (CT), Logistic Regression, Naive Bayes (NB), Support Vector Machines (SVM), and Ensemble methods were employed as supervised machine learning techniques for classification. Each method provided distinct strengths and characteristics, facilitating a comprehensive evaluation of their effectiveness in distinguishing PD patients from individuals with other neurological disorders. The Naive Bayes kernel, using seven PCA-derived components, achieved the highest accuracy rate of 86.84% among the tested classification methods. It is worth noting that classifier performance may vary based on the dataset and specific characteristics of the voice samples. In conclusion, this study demonstrated the potential of voice analysis as a diagnostic tool for distinguishing PD patients from individuals with other neurological disorders. By employing a variety of voice analysis techniques and utilizing different machine learning algorithms, including Classification Tree, Logistic Regression, Naive Bayes, Support Vector Machines, and Ensemble methods, a notable accuracy rate was attained. However, further research and validation using larger datasets are required to consolidate and generalize these findings for future clinical applications.*

Keywords: voice analysis, Parkinson's disease, MFCC, PCA, naive Bayes kernel, machine learning

KLASYFIKACJA CHOROBY PARKINSONA I INNYCH ZABURZEŃ NEUROLOGICZNYCH Z WYKORZYSTANIEM EKSTRAKCJI CECH GŁOSOWYCH I TECHNIK REDUKCJI

Streszczenie. *Przedstawione badanie miało na celu różnicowanie osób z chorobą Parkinsona (PD) od osób z innymi zaburzeniami neurologicznymi poprzez analizę próbek głosowych, biorąc pod uwagę związek między zaburzeniami głosu a PD. Próbki głosowe zostały zebrane od 76 uczestników przy użyciu różnych urządzeń i warunków nagrywania, a uczestnicy byli instruowani, aby wydłużyć samogłoskę /a/ w wygodnym tempie. Oprogramowanie PRAAT zostało zastosowane do ekstrakcji cech, takich jak autokorelacja (AC), krzyżowa korelacja (CC) i współczynniki cepstralne Mel (MFCC) z próbek głosowych. Analiza składowych głównych (PCA) została wykorzystana w celu zmniejszenia wymiarowości cech. Jako techniki nadzorowanego uczenia maszynowego wykorzystano drzewa decyzyjne (CT), regresję logistyczną, naiwny klasyfikator Bayesa (NB), maszyny wektorów nośnych (SVM) oraz metody zespołowe. Każda z tych metod posiadała swoje unikalne mocne strony i charakterystyki, umożliwiając kompleksową ocenę ich skuteczności w rozróżnianiu pacjentów z PD od osób z innymi zaburzeniami neurologicznymi. Naiwny klasyfikator Bayesa, wykorzystujący siedem składowych PCA, osiągnął najwyższy wskaźnik dokładności na poziomie 86,84% wśród przetestowanych metod klasyfikacji. Należy jednak zauważyć, że wydajność klasyfikatora może się różnić w zależności od zbioru danych i konkretnych cech próbek głosowych. Podsumowując, to badanie wykazało potencjał analizy głosu jako narzędzia diagnostycznego do rozróżniania pacjentów z PD od osób z innymi zaburzeniami neurologicznymi. Poprzez zastosowanie różnych technik analizy głosu i wykorzystanie różnych algorytmów uczenia maszynowego, takich jak drzewa decyzyjne, regresja logistyczna, naiwny klasyfikator Bayesa, maszyny wektorów nośnych i metody zespołowe, osiągnięto znaczący poziom dokładności. Niemniej jednak, konieczne są dalsze badania i walidacja na większych zbiorach danych w celu skonsolidowania i uogólnienia tych wyników dla przyszłych zastosowań klinicznych.*

Słowa kluczowe: analiza głosu, choroba Parkinsona, MFCC, PCA, naiwne jądro bayesowskie, uczenie maszynowe

Introduction

Parkinson's disease, which was initially described by Dr. James Parkinson in 1817, is a progressive nervous system condition that deteriorates over time and affects physical movements, including speech [23]. It is the second most prevalent neurological disorder, surpassed only by Alzheimer's, Multiple System Atrophy (MSA), brain tumors, epilepsy, and other neurodegenerative conditions. The disease is caused by the degradation of pigment cells in the basal ganglia, resulting in a deficiency of dopamine and disruption of neurotransmission in the Substantia Nigra region of the midbrain, which is responsible for regulating motor function. Motor problems result from dopamine insufficiency. Symptoms include bradykinesia, postural instability, tremors, and hypokinetic movement abnormalities. Although this condition is straightforward to diagnose in its advanced stages, effective therapy is difficult [22]. There is presently no curative medicinal therapy for Parkinson's disease.

The study of voice abnormalities associated with neurological conditions, which can be caused by various factors that affect muscle tone [29], has significantly advanced. These conditions are categorized as either hypotonia, which refers to low muscle tone, or hypertonia, which refers to high muscle tone. Hypotonia is characterized by reduced loudness, changes in fundamental frequency, and voice instability, while hypertonia is characterized by vocal pauses, voice instability [3], and changes in voice quality [27]. Neurological conditions such as vocal tremors, spasmodic dysphonia, and vocal cord paralysis can all impair the voice. Moreover, Parkinson's disease can impact speech and voice, leading to diminished pitch and loudness fluctuations, decreased overall loudness, and a breathy quality in the voice due to insufficient closure of the vocal cords. Subjective approaches are routinely used by clinicians and vocal pathologists to evaluate speech impairments in Parkinson's disease patients, which can have a substantial impact on communication and quality of life. These approaches utilize acoustic characteristics including fundamental frequency, sound intensity level, tremor, shimmer, ratio of low-frequency to high-frequency components, cephalic peak prominence, and harmonics-to-noise ratio within the signal for assessment [1, 20, 31]. In contrast, recent research has shifted towards utilizing acoustic parameters derived from time-based data, along with spectrum and cepstral measurements, for more objective evaluations [2]. These evaluations provide a more precise understanding of the severity and characteristics of speech disorders in Parkinson's disease (PD) patients and are crucial for effective management. PD commonly results in voice weakening in about 90% of patients [21], typically occurring in those over 60 years old, although it can also affect younger individuals, albeit rarely. Additionally, gender appears to play a role in PD prevalence, with men being more affected than women [10]. Recent research has focused on developing objective methods to diagnose vocal problems by measuring voice quality in the temporal, spectral, and cepstral domains, including parameters such as fundamental frequency (F0), absolute sound pressure level, jitter, shimmer, and harmonicity [18, 19, 25].

In [12], the authors proposed a method for diagnosing PD using speech signals. They collected a dataset of speech recordings from individuals with PD and healthy controls. The method involved extracting resonance and time-frequency-based features from the speech signals, but the paper lacked specific details on the algorithms and parameters used. The extracted features were fused using an undisclosed technique, and machine learning classifiers were employed for PD diagnosis. However, the paper did not specify the classifiers used or provide information on the training and evaluation procedures. The paper also lacked crucial details regarding the dataset, feature extraction, fusion technique, classifiers, and performance evaluation metrics. A comprehensive discussion of the results and their interpretation was also missing. Addressing these limitations would improve the study's scientific rigor and impact.

The authors in [24] presented an approach to diagnose PD by utilizing cepstral features extracted through MFCC and dimensionality reduction techniques. However, the paper did not provide sufficient information regarding the dataset, feature extraction parameters, dimensionality reduction methods, SVM classifier parameters, and performance evaluation metrics. Furthermore, the study lacked in-depth discussions and interpretations of the results, which diminished its overall strength. To improve the research's scientific rigor and impact, it is crucial to address these limitations and provide comprehensive details and analysis.

The primary focus of the study described in [26] is the investigation of feature extraction and classification techniques specifically designed for dysphonic speech disorder in individuals with PD. The main objective is to identify effective methods for extracting pertinent features from dysphonic speech signals and to compare different classification techniques to achieve accurate diagnosis of PD-related dysphonia. The paper thoroughly examines various feature extraction techniques and provides detailed information on the algorithms and parameters employed. Furthermore, it assesses different classification methods using performance metrics. The findings of this study contribute to enhancing our understanding of dysphonic speech in PD and offer valuable insights into the diagnosis of PD-related dysphonia.

While most studies focus on distinguishing between patients with Parkinson's disease and healthy individuals, our research stands out by focusing on classifying Parkinson's disease and other neurological disorders. The objective of our study is to provide an objective method for diagnosing vocal problems in Parkinson's disease patients and differentiating them from patients with other neurodegenerative diseases. By collecting voice recordings from patients with Parkinson's disease, Multiple System Atrophy, and other neurological disorders, we built a database of 76 voice samples. Through the utilization of cepstral domains, we extracted acoustic features from each voice sample, and principal component analysis (PCA) was employed to select the most relevant features. We then employed various supervised machine-learning techniques for classification. The combination of the Naive Bayes kernel with linear classification and 7 components of PCA achieved a maximum classification accuracy of 86.84%. This machine learning method, along with artificial intelligence techniques, can aid in the detection of Parkinson's disease, Multiple System Atrophy, and other neurological diseases. Once the algorithm is trained to recognize these patterns, it can effectively classify patients based on their health status, distinguishing Parkinson's disease patients from those with similar symptoms but other neurodegenerative disorders. Our study has the potential to enhance the accuracy of diagnosis, leading to improved management and treatment of Parkinson's disease.

Unlike [12, 18, 19, 25], the significance of evaluating acoustic characteristics for Parkinson's disease diagnosis is emphasized in this study, regardless of whether dimension reduction is employed. Certain features have exhibited high accuracy in distinguishing Parkinson's disease from other disorders. Additionally, another study utilizing dimension reduction has also shown promising outcomes.

In contrast to previous studies [24, 26], the comparison of various machine learning classifiers such as Classification Tree (CT), Logistic Regression, Naive Bayes (NB), Support Vector Machines (SVM), and Ensemble methods reveals their effectiveness in accurately detecting and distinguishing between patient groups, thereby assisting in clinical diagnosis.

To the best of our knowledge, this study represents the only existing research focused on classifying patients with Parkinson's disease from other neurological diseases with similar symptoms.

Furthermore, our study goes beyond previous research by considering a larger and more diverse dataset, allowing for a more comprehensive analysis of the classification between patients with Parkinson's disease and those with other neurological disorders.

The paper is organized as follows: Section 1 presents the methodology and database used in this study, Section 2 discusses the evaluation metrics, Section 3 presents the obtained results and discussion, and Section 4 provides the conclusion.

1. Methodology

In this research, we utilized the cepstral technique to extract cepstral coefficients from acoustic recordings. We then employed PRAAT software to analyze and extract features, specifically Mel-frequency cepstral coefficients (MFCC), from these samples. This section provides a comprehensive overview of the complete procedure involved in these three approaches, along with their categorization.

1.1. Dataset

All participants in the study provided self-reported diagnoses of Parkinson's disease, MSA, or other ND. Voice recordings were collected using two methods: some participants used their own devices to record their voices, while others were recorded by physicians using a smartphone microphone. The study utilized a database obtained from a previous study [4], which consisted of 76 voice samples. Out of these, 56 samples were collected from Parkinson's disease patients, who were divided into three groups based on the recording method: the first group used smartphones (10 females and 22 males), the second group used tablets (5 females and 3 males), and the third group used a computer (6 females and 10 males). Additionally, voice recordings were obtained from 20 patients with various other neurological disorders, including Multiple System Atrophy 9 patients (2 females and 7 males), Functional Neurological Disorders 5 patients (3 females and 2 males), Essential Tremor (1 female), Dystonia (2 males and 1 female), Cervical Dystonia (1 female), and Somatization (1 female). All voice samples were saved in WAV format and recorded in mono-channel mode.

1.2. Feature extraction

Some individuals with vocal cord pathology may struggle to maintain stable phonation when pronouncing a sustained vowel /a/ [5]. Voice recordings and pre-processing alone may not be sufficient to accurately assess voice disorders. Therefore, we opted to utilize the acoustic software PRAAT to extract a comprehensive set of information on speech characteristics. We employed two methods: autocorrelation (all parameters) to pitch (ac), and cross-correlation method to pitch (cc). A total of 73 parameters were extracted for speech analysis, including 10 jitter parameters (dc and ac) [jitter (local), jitter (local, absolute), jitter (rap), jitter (ppq5), jitter (ddp)], 12 speckle parameters (dc and ac) [(speckle (local), speckle (local, dB), speckle (apq3), speckle (apq5), speckle (apq11), speckle (dda)], 6 harmonicity parameters (dc and ac) (mean autocorrelation, mean noise/harmonic ratio, mean harmonic/noise ratio), 10 pitch parameters (cc and ac) (median pitch, mean pitch, standard deviation, minimum pitch, maximum pitch), 8 pulse parameters (cc and ac) (number of pulses, number of periods, mean period, standard deviation of period), 6 vocalization parameters (fraction of locally unvoiced

pitch frames, number of voicing breaks, degree of voicing breaks), frequency of formants 1, 2, and 3, maximum, minimum, and mean intensity, as well as 15 first coefficient of MFCC.

Jitter and Shimmer are significant metrics used in the evaluation of vocal cord pathology in patients who are affected. Jitter measures the variation in the fundamental frequency (F0) of the voice from cycle to cycle [16, 17], while Shimmer measures the variation in voice cycle width [32]. The algorithm used to calculate Jitter and Shimmer involves computing the mean of the disparities in duration or amplitude or amplitude of consecutive periods. These measurements are useful in aiding the diagnosis and monitoring of vocal cord pathology, as well as evaluating the effectiveness of treatments.

PCA (Principal Component Analysis) is a popular statistical approach for reducing the dimensionality of voice data and extracting the most important information. After applying PCA to the speech data, we obtained 13 principal components, ranging from the first principal component, which accounts for the highest variance, to the 13th principal component, which captures the least variance. The basic idea behind PCA is to describe the variance in a multidimensional dataset using a collection of uncorrelated variables that are linear combinations of the original variables. These new variables, dubbed "principal components", are arranged in decreasing order of significance, with the first representing the most variance in the original data. Subsequent components are chosen to have the least amount of variance while staying uncorrelated with the prior components.

1.2.1. Jitter measurement

Jitter is a parameter that measures micro-perturbation of the fundamental period and has several parameters [28]:

- Jitter (relative or local) refers to the timing variations or fluctuations of a signal relative to a reference signal or a clock, expressed as the difference in timing between the expected timing and the actual timing of signal transitions with respect to a reference signal [11]. Is calculated as follows:

$$\text{Jitter (relative)} = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|T_i - T_{i+1}|}{\frac{1}{N}\sum_{i=1}^{N} T_i}$$

Where Ti represents the period lengths of the extracted fundamental frequency $F0$ s and N is denotes the total number of extracted $F0$ periods.

Table 1. Time-frequency-based features given by Praat acoustic analysis software

Groups	Features
Pitch parameters	Median pitch (Hz)
	Mean pitch (Hz)
	Standard deviation (Hz)
	Minimum pitch (Hz)
	Maximum pitch (Hz)
Pulses parameters	Number of pulses
	Number of periods
	Mean period (s)
	Standard deviation of period (s)
Voicing parameters	Fraction of locally unvoiced frames (%)
	Number of voice breaks
	Degree of voice breaks (%)
Jitter parametrs	Jitter (local) (%)
	Jitter (local. absolute) (s)
	Jitter (rap) (%)
	Jitter (ppq5) (%)
	Jitter (ddp) (%)
Shimmer parameters	Shimmer (local) (%)
	Shimmer (local. dB) (dB)
	Shimmer (apq3) (%)
	Shimmer (apq5) (%)
	Shimmer (apq11) (%)
	Shimmer (dda) (%)
	Mean autocorrelation
Harmonicity parameters	Mean noise-to-harmonics ratio
	Mean harmonics-to-noise ratio
Frequence formant	F1 F2 F3 (Hz)
Intencity	Max inten Min inten mean inten

- Jitter (absolute) is the change in fundamental frequency F0 [25] from cycle to cycle, represented as [5]:

$$\text{Jitter (absolute)} = \frac{1}{N-1}\sum_{i=1}^{N-1}|T_i - T_{i+1}|$$

- Jitter in Praat is a measure of how much the period (or pitch) of speech signals varies or perturbs. Praat offers different variants of jitter, such as Jitter (RAP), Jitter (PPQ5), and Jitter (ddp), which differ in the number of points used for calculation and the way they are expressed [11]. These variants provide different perspectives on the variation in pitch, allowing for more nuanced analysis of speech signals.

1.2.2. Shimmer measurement

Shimmer is a parameter that identifies micro-perturbations of the signal amplitude, and is measured using the parameters listed below [32]:

- Shimmer (relative) is used to describe a measure of the variation or amplitude perturbation in the intensity or loudness of speech signals. It is calculated as the relative change in amplitude between consecutive speech frames, expressed as a percentage [25].

$$\text{Shimmer (relative)} = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|A_i - A_{i+1}|}{\frac{1}{N}\sum_{i=1}^{N} A_i}$$

where A_i represents the extracted peak-to-peak amplitude data, and N denotes the number of extracted fundamental frequency periods F0.

Shimmer (dB) is a speech and voice analysis measure that quantifies the amplitude or intensity variation in a vocal signal. It is expressed in decibels (dB) [11], it can be expressed as:

$$\text{Shimmer (dB)} = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|20\log\left(\frac{A_{i+1}}{A_i}\right)\right|$$

- Shimmer, also referred to as Amplitude Perturbation Quotient (APQ), is a parameter used to assess the perturbation or fluctuation in the amplitude of speech signals. Various versions of shimmer exist, such as Shimmer (APQ3), Shimmer (APQ5), Shimmer (APQ11), and Shimmer (dda), which vary in terms of the number of data points utilized in the calculation and the manner in which they are expressed [11].

1.2.3. Harmonicity

The Mean Harmonics-to-Noise Ratio (HNR), which measures the proportion of harmonic components in the signal [7], and the Mean Noise-to-Harmonics Ratio (NHR), which evaluates the ratio of noise to harmonic energy in different frequency bands [8]. These measures are useful for analyzing the periodicity and noisiness of speech signals in various applications.

1.3. MFCC process

MFCCs (Mel-frequency cepstral coefficients) are a widely used feature extraction technique in audio processing and speech recognition. The Mel-scale is created by stacking triangular filters, usually ranging from 15 to 30, that are linearly spaced up to 1 kHz and logarithmically spaced above 1 kH [4]. The following stages are commonly included in computing MFCCs.

1.3.1. Farming

To account for the non-stationary character of speech waveforms over longer durations, a short-term analysis approach is utilized to efficiently analyze speech signals. Because the movement of speech articulators is physically restricted, frames of 10-30 ms are deemed stable for study. The analysis is carried out in regular time periods or frames, each with a fixed duration. This entails separating the voice signal into N-sample frames with an M-sample gap between consecutive frames (where M is fewer than N) [9].

1.3.2. Pre-emphasis

In this step, we amplify the energy in the voice stream by emphasizing higher frequencies. This is accomplished by solving the following first-order difference equation for the samples $\{s_n, n = 1, \dots, N\}$ [9]:

$$s'_n = s_n - k * s_{n-1}$$

The value of k in the equation represents the pre-emphasis coefficient, which is required to be within the range of $0 \leq k < 1$. according to reference [9]. For our investigation, we used a pre-emphasis coefficient of $k = 0.97$.

1.3.3. Fast Fourier Transform (FFT)

The audio frames that have undergone pre-processing are then subjected to the Fourier transform, which converts the signal from the time domain to the frequency domain. This process results in a representation of the audio signal in terms of its spectral content, which is defined for the set of N samples $\{s_n, n = 0,1,2, \dots, N-1\}$ as follow [4, 13]:

$$s_n = \sum_{k=0}^{N-1} s_k * e^{-2\pi jkn/N}$$

1.3.4. Mel-frequency wrapping

The Mel-filterbank is applied to the spectral representation of the audio signal using a set of triangular filters uniformly spaced in the Mel-scale, which mimics the frequency perception of the human ear. This filterbank is used to compute the energy in each Mel-frequency band. The Mel scale is logarithmic above 1000 Hz and linear below 1000 Hz, with a reference tone of 1 KHz at 40 dB above the perceptual hearing threshold set at 1000 mels. An approximate formula is used to determine the mels of a given frequency in Hz [13, 30].

$$Mel(f) = 2595 * log_{10}\left(1 + {}^{f}/_{700}\right)$$

1.3.5. Cepstrum

The compressed Mel-filterbank energies are subjected to a Discrete Cosine Transform (DCT), which serves to decorrelate the energies and capture cepstral features that represent the spectral shape of the audio signal. In order to transform the logarithm of the Mel spectrum into the time domain, the cepstral representation of the speech spectrum is utilized, achieved by applying the DCT to the log filter bank amplitudes using a specific formula. The MFCCs are valuable for frame analysis in speech processing, as they effectively capture the local spectral features of the speech signal [30].

$$c_i = \sqrt{\frac{2}{N}} \sum_{j=1}^{N} m_j * \cos\left(\frac{\pi i}{N}(j - 0.5)\right)$$

The value of N corresponds to the total number of filter bank channels employed in the computation.

1.3.6. Liftering

Liftering is a technique in audio and speech signal processing that modifies MFCCs by applying a window function, such as a raised cosine window, to emphasize or attenuate specific frequency bands. It is commonly used to adjust spectral features in audio signals. One potential issue is that higher-order cepstral coefficients can become very small, which may pose challenges. To address this, cepstral liftering is employed, which involves normalizing the amplitudes of the cepstral coefficients using a specific formula [4, 9].

$$c'_n = (1 + \frac{L}{2} * \sin\left(\frac{\pi * n}{L}\right) * c_n$$

The value of L in the equation corresponds to the cepstral sine lifter parameter, and we employed $L = 22$ in this work.

2. Machine learning classifiers

After the feature selection process, the subsets of features are utilized in various machine learning algorithms to differentiate between PD patients and healthy individuals. We have employed three classifiers, namely Classification Tree (CT), Logistic Regression, Naive Bayes (NB), Support Vector Machines (SVM), and Ensemble methods.

The Classification Tree (CT) algorithm is an approach that uses decision tree principles to recursively split data based on feature values. It builds a tree structure where internal nodes represent features, and leaf nodes indicate class labels. The splits are determined by criteria like the Gini index or information gain, aimed at optimizing the separation between classes. This decision tree is a flowchart-like structure, where internal nodes symbolize functions or attributes, branches represent decision rules, and leaf nodes signify outcomes. The root node, situated at the top, decides the partitioning based on attribute values. Recursive partitioning is a technique for repeatedly dividing the tree. This flowchart-like representation aids in decision-making and closely mirrors human thinking processes. Consequently, decision trees are easily understandable and intuitive [15].

Logistic Regression is a popular mathematical method for predicting binary outcomes, where the class variable y takes values of 0 or 1. It differs from linear regression, which is more suitable for continuous outcomes, as logistic regression is specifically designed for categorical outcomes. It employs the standard logistic function, which is an S-shaped curve given by the equation:

$$f(x) = \frac{1}{1 + e^{-x}}$$

to calculate the probability of belonging to a particular class. The logistic regression model assumes a linear relationship between the input features and the logarithm of the odds of the outcome. By estimating the probability and applying a predefined threshold, the algorithm assigns the appropriate class label to each instance [15].

Naive Bayes (NB) is a probabilistic classifier that applies Bayes' theorem with the assumption of feature independence. It calculates the probability of each class given the input features and assigns the class label with the highest probability. Despite its "naive" assumption, Naive Bayes has shown good performance in many classification tasks [15].

Naive Bayes classifier is a probabilistic machine learning model that is used for binary (two-class) and multi-class classification problems. The classifier is based on the Bayes theorem:

$$P(y|X) = \frac{P((X|y)P(y)}{P(X)}$$

The class variable y and the parameter/feature variables represented by $X = (x1, x2, \dots, xn)$ are involved in Naive Bayes. The Naive Bayes classifier assumes that the attributes are independent of each other. Hence, it assumes that the presence or value of one attribute does not affect the presence or value of other attributes.

$$P(y|X) = \frac{P(x_1|y)P(x_2|y) \dots P(x_n|y)P(y)}{P(x_1)P(x_2) \dots P(x_n)}$$

For Gaussian Naive Bayes, the conditional probability is derived from a normal distribution, similar to a Gaussian distribution.

$$P(x_i|y) = \frac{1}{\sigma_y\sqrt{2\pi}} e^{-{(x_i-\mu_y)^2}/{2\sigma^2 y}}$$

Support Vector Machines (SVM) is a supervised machine learning technique utilized for solving classification and regression problems. It distinguishes itself by identifying a hyperplane that effectively separates the classes. The hyperplane is determined by maximizing the margin, which is the distance between the hyperplane and the nearest data points from each class. In cases where the data is not linearly separable, the kernel trick is employed. The kernel function allows for transforming the input space from a lower dimension to a higher dimension, enabling the resolution of non-linear separable problems. In our research, we utilized the sigmoid function as the kernel for SVM [15].

Ensemble methods combine multiple individual classifiers to make predictions. Techniques such as Random Forest, Bagging, or Boosting can be used for ensemble methods. These methods create an ensemble of classifiers and aggregate their predictions to make a final decision. Ensemble methods often improve classification accuracy and help mitigate overfitting.

3. Evaluation metrics

In order to evaluate the effectiveness of our classifiers in distinguishing between PD patients and patients with different ND, we used many performance criteria, including accuracy, sensitivity, and specificity, to assess the usefulness of our classifiers in differentiating between PD patients and patients with other ND [4, 6].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

True positives (TP) are Parkinson's disease patients successfully diagnosed by the classifier. True negatives (TN) are patients with ND who have been appropriately classified. False positives (FP) are ND patients who were wrongly categorized. False negatives (FN) are Parkinson's disease individuals who were misclassified [6, 30]. Accuracy quantifies the classifier's performance in discriminating between the two groups, sensitivity measures the accuracy of PD patient recognition, and specificity gauges the accuracy of detecting patients with other neurodegenerative diseases [6, 9]. Additionally, we included two more variables along with accuracy, sensitivity, and specificity. Matthews' correlation coefficient (MCC) and Probability Excess (PE). MCC is a reliable measure of the quality of binary classification and ranges from 0 (for random algorithms) to 1 (for perfect algorithms). PE, on the other hand, also ranges from 0 (for random prediction) to 1 (for perfect prediction) [4].

$$MCC = \frac{TP \times TN - FN \times FP}{\sqrt{((FN + TP)(FP + TN)(FP + TP)(FN + TN))}}$$

$$PE = \frac{TP \times TN - FN \times FP}{(FN + TP)(FP + TN)}$$

Table 2. Confusion matrix used in the analysis

		Result	Of classification
		ND	PD
Diagnosed	ND	TP	FP
	PD	FN	TN
		Sensitivity	Specificity

4. Results and discussion

In our research, we utilized cepstral analysis to extract features and coefficients from voice samples in order to classify patients into two categories: those with Parkinson's Disease (PD) and those without (ND). We employed various classifiers and kernels, such as Classification Tree (CT), Logistic Regression, Naive Bayes (NB), Support Vector Machines (SVM), and Ensemble methods, to accurately differentiate between the two groups and gain insights into the potential use of voice analysis in PD diagnosis. To address the issue of overfitting, we con-ducted cross-validation experiments with different numbers of folds, including 5, 10, 20, and 30, on our dataset which com-prised 76 samples with 75 features each. After careful analysis of the results, we determined that using 20 folds for cross-validation provided accurate performance estimation.

The table 3 presents the classification performance of various acoustic features for differentiating Parkinson's Disease (PD) from other neurodegenerative diseases (ND) without employing dimensionality reduction techniques such as PCA. The results demonstrate the accuracy, sensitivity, specificity, true positive (TP), true negative (TN), false positive (FP), false negative (FN), Matthewscorrelation coefficient (MCC), and classification error rate (PE) for each feature. Overall, the table shows that some features achieved high accuracy of 73.7% or 75% in differentiating PD from ND. However, it is important to note that several features, such as median pitch, mean pitch, standard deviation of pitch, minimum pitch, maximum pitch, mean period, standard deviation of period, jitter parameters, shimmer parameters, mean autocorrelation, and harmonicity parameters, all had an accuracy of 73.7% without distinguishing between PD and ND. These features did not provide sufficient discriminatory power for accurate classification. On the other hand, certain features exhibited higher accuracy and sensitivity values, such as the number of pulses (77.63%), number of periods (78.94%), fraction of locally unvoiced frames (73.68%), number of voice breaks (75%), degree of voice breaks (77.63%), jitter (ppq5) (75%), formant frequencies (F1, F2, F3) (77.5%), and the first 15 coefficients of MFCC (75%). These features showed promise in differentiating PD from other ND with reasonable accuracy. However, it is crucial to consider the limitations of the study and the interpretation of these results. The sample size used in this study was relatively small, with only 56 PD patients and 20 patients with other ND. Therefore, the generalizability of the findings to a larger population should be carefully considered. Additionally, the specific classification algorithms and parameter settings used for the machine learning models are not provided in the table, making it difficult to assess the reproducibility and reliability of the results.

In conclusion, the table highlights the classification performance of different acoustic features in distinguishing PD from other ND without employing dimensionality reduction techniques. Some features exhibited higher accuracy and sensitivity, while others showed limited discriminatory power. Further research with larger and more diverse datasets, along with detailed information on classification algorithms and parameter settings, is necessary to validate and improve the effectiveness of these features for accurate PD classification.

The performance of various classifiers evaluated using linear PCA can be assessed by selecting the best 4 numbers from the 13th principal components. In Table 4, it was observed that Kernel NB, with numeric components 2, 4, 6, and 7, consistently demonstrated superior performance in terms of accuracy (ranging from 80.3% to 86.84%), sensitivity (ranging from 72.77% to 89.65%), specificity (ranging from 77.77% to 89.65%), and MCC (ranging from 0.5837 to 0.6743). Gaussian NB also showed relatively good performance with accuracy ranging from 75% to 78.94% and sensitivity ranging from 50% to 68.75%. The number of numeric components used in linear PCA appeared to impact the performance of certain classifiers. SVM and ensemble (Bugged trees) showed consistent accuracy of 73.7% across all numeric component values, indicating that the performance was not influenced by the number of components. However, Rus Boosted trees displayed a significant drop in accuracy from 34.21% to 28.57% as the number of components increased from 2 to 4. Sensitivity, also known as recall or true positive rate, measures the ability of a classifier to accurately identify positive cases, while specificity, also known as true negative rate, measures the ability to correctly identify negative cases. Higher sensitivity and specificity values are generally desired for a reliable classifier, as they indicate better capability in correctly classifying both positive and negative cases. MCC is a metric that assesses the quality of binary (two-class) classification, taking into account true positives, true negatives, false positives, and false negatives. A higher MCC value indicates better classification performance. PE measures the average prediction error of the classifier, with lower values indicating better performance.

Table 3. Classification Performance of Acoustic Features for Differentiating Parkinson's Disease and Other Neurodegenerative Diseases without Dimensionality Reduction

Groups	Features	Accuracy (%)	Sensitivity (%)	Specificity (%)	TP	TN	FP	FN	MCC	PE
Pitch parameters	Median pitch	73.7	0	100	0	56	20	0	0	0
	Mean pitch	73.7	0	100	0	56	20	0	0	0
	Standard deviation	73.7	0	100	0	56	20	0	0	0
	Minimum pitch	73.7	0	100	0	56	20	0	0	0
	Maximum pitch	73.7	0	100	0	56	20	0	0	0
Pulses parameters	Number of pulses	77.63	63.63	80	7	52	13	4	0.3486	0.2823
	Number of periods	78.94	63.63	80.30	7	53	13	3	0.3891	0.4643
	Mean period	73.7	0	100	0	56	20	0	0	0
	Standard deviation of period	73.7	0	100	0	56	20	0	0	0
Voicing parameters	Fraction of locally unvoiced frames	73.68	5.3	75.71	3	53	17	3	0.1574	0.2571
	Number of voice breaks	75	10	75.34	2	55	18	1	0.1857	0.42
	Degree of voice breaks	77.63	80	77.46	4	55	16	1	0.3235	0.5746
Jitter parametrs	Jitter (local)	73.7	0	100	0	56	20	0	0	0
	Jitter (local. absolute)	73.7	0	100	0	56	20	0	0	0
	Jitter (rap)	73.7	0	100	0	56	20	0	0	0
	Jitter (ppq5)	75	3.44	75.67	2	56	18	0	0.275	0.7567
	Jitter (ddp)	73.7	0	100	0	56	20	0	0	0
Shimmer parameters	Shimmer (local)	73.7	0	100	0	56	20	0	0	0
	Shimmer (local. dB)	73.7	0	100	0	56	20	0	0	0
	Shimmer (apq3)	73.7	0	100	0	56	20	0	0	0
	Shimmer (apq5)	73.7	0	100	0	56	20	0	0	0
	Shimmer (apq11)	73.7	0	100	0	56	20	0	0	0
	Shimmer (dda)	73.7	0	100	0	56	20	0	0	0
Harmonicity parameters	Mean autocorrelation	73.7	0	100	0	56	20	0	0	0
	Mean noise-to-harmonics ratio	73.7	0	100	0	56	20	0	0	0
	Mean harmonics-to-noise ratio	73.7	0	100	0	56	20	0	0	0
Frequence formant	F1 F2 F3	77.5	60	72.22	6	52	14	4	0.2977	0.3878
Intencity	Max inten, Min inten, mean inten	73.7	0	100	0	56	20	0	0	0
Coefficient MFCC	15 firsts Coeff	75	3.44	75.67	2	52	18	0	0.275	0.7567

Table 4. Performance Comparison of Different Classifiers with Varying Number of Numeric Components

Classifiers		Number numeric of components	Acc (%)	Sens (%)	Spec (%)	TP	TN	FP	FN	MCC	PE
CT		2 4 6 7	73.7	0	100	0	56	20	0	0	0
LR		2 4 6 7	73.7	0	100	0	56	20	0	0	0
Navies Bayes	Gaussian NB	2	75	53.85	79.37	7	50	13	6	0.2839	0.3321
		4	78.9	68.75	0.845	11	49	9	5	0.4934	0.5323
		6	75	53.85	79.37	7	50	13	6	0.2839	0.3289
		7	78.94	66.66	86.66	8	52	12	4	0.3968	0.4792
	Kernel NB	2	80.3	77.77	80.59	7	54	13	2	0.4283	0.5837
		4	82.9	73.33	85.24	11	52	9	4	0.5295	0.5858
		6	84.21	72.77	87.93	13	51	7	5	0.5808	0.6015
		7	86.84	77.77	89.65	14	52	6	4	0.6510	0.6510
SVM		2 4 6 7	73.7	0	100	0	56	20	0	0	0
Ensemble	Bugged trees	2 4 6 7	73.7	0	100	0	56	20	0	0	0
	Rus Boosted trees	2 4 6 7	34.21	28.57	100	20	6	56	0	0.6732	0.2857

In summary, after analyzing the outcomes of linear PCA on various classifiers, it can be concluded that Kernel NB consistently demonstrates superior performance across different numeric component values, followed by Gaussian NB. However, the choice of the best classifier may also depend on specific requirements and goals of the classification task, and other factors like computational efficiency, interpretability, and ease of implementation should also be taken into consideration. It is crucial to thoroughly evaluate the performance of different classifiers using appropriate evaluation metrics and select the one that aligns with the specific needs of the task at hand.

5. Conclusion

The study aimed to assess speech impairments in patients with Parkinson's disease and other neurodegenerative diseases by anlyzing acoustic measurements over time, as well as spectral and cepstral measurements. The researchers extracted acoustic features and MFCCs from voice samples using cepstral domains, and then performed feature selection through PCA to differentiate between PD patients and those with other neurodegenerative diseases. The results showed that using PCA with the Naive Bayes kernel and linear classification, the maximum classification accuracy achieved was 86.84%. This study demonstrated the potential of using objective assessments and machine learning techniques to accurately identify and differentiate speech disorders in PD patients, which could lead to improved management and treatment of these disorders, ultimately enhancing communication and quality of life for patients. The researchers utilized various classifiers and kernels, along with cross-validation, to address overfitting and classify patients with PD and other neurodegenerative diseases based on voice samples. The Naive Bayes kernel with PCA yielded the most favorable results in terms of classification accuracy. However, when selecting the best classifier, specific task requirements and other factors such as computational efficiency, interpretability, and ease of implementation should be considered. Thorough evaluation of classifier performance using appropriate metrics is crucial in choosing the most suitable classifier for the task at hand.

Future work should consider incorporating deep learning and convolutional neural networks (CNN) to assess the severity of the disease. This can improve the accuracy and specificity of speech disorder classification in Parkinson's disease (PD) and other neurodegenerative diseases. Furthermore, the inclusion of larger datasets and exploration of alternative machine learning algorithms can provide valuable insights and enhance the overall performance of classification models. These advancements have the potential to enhance our understanding of speech impairments in neurodegenerative diseases and contribute to personalized treatment approaches.

References

[1] Abramson E. L. et al.: Physician experiences transitioning between an older versus newer electronic health record for electronic prescribing. International journal of medical informatics 81(8), 2012, 539–548.
[2] Andersen T. et al.: Designing for collaborative interpretation in telemonitoring: Re-introducing patients as diagnostic agents. International journal of medical informatics 80(8), 2011, e112–e126.
[3] Baker K. K. et al.: Thyroarytenoid muscle activity associated with hypophonia in Parkinson disease and aging. Neurology 51(6), 1998, 1592–1598.
[4] Benba A. et al.: Discriminating between patients with Parkinson's and neurological diseases using cepstral analysis. IEEE transactions on neural systems and rehabilitation engineering 24(10), 2016, 1100–1108.
[5] Benba A. et al.: Hybridization of best acoustic cues for detecting persons with Parkinson's disease. Second World Conference on Complex Systems (WCCS), IEEE, 2014, 622–625.
[6] Benba A. et al.: Using RASTA-PLP for discriminating between different neurological diseases. International Conference on Electrical and Information Technologies (ICEIT), IEEE, 2016, 406–409.
[7] Boersma P. et al.: Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. Proceedings of the institute of phonetic sciences, 1993, 97–110.
[8] De Colle W.: Voce & Computer – analisi acustica digitale del segnale verbale. Omega, 2001.
[9] Gales M. et al.: The application of hidden Markov models in speech recognition. Foundations and Trends in Signal Processing 1(3), 2008, 195–304.
[10] Gillies G. E., et al.: Sex differences in Parkinson's disease. Frontiers in neuroendocrinology 35(3), 2014, 370–384.
[11] Gorris C. et al.: Acoustic analysis of normal voice patterns in Italian adults by using Praat. Journal of Voice 34(6), 2020, 961.e9–961.e18.
[12] Goyal J. et al.: A hybrid approach for Parkinson's disease diagnosis with resonance and time-frequency based features from speech signals. Expert Systems with Applications 182, 2021, 115283.
[13] Gupta S. et al.: Feature extraction using MFCC. Signal & Image Processing. An International Journal 4(4), 2013, 101–108.
[14] Holi M. S. et al.: Automatic detection of neurological disordered voices using mel cepstral coefficients and neural networks. IEEE Point-of-Care Healthcare Technologies (PHT), IEEE, 2013, 76–79.
[15] Islam M. R., Nahiduzzaman M.: Complex features extraction with deep learning model for the detection of COVID19 from CT scan images using ensemble based machine learning approach. Expert Systems with Applications 195, 2022, 116554.
[16] Lieberman P.: Perturbations in vocal pitch. The Journal of the Acoustical Society of America 33, 1961, 597–602.
[17] Lieberman P.: Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. The Journal of the Acoustical Society of America 35(3), 1963, 344–353.
[18] Little M. et al. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. Nature Precedings, 2008, 1–1.
[19] Little M. et al.: Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. Nature Precedings, 2007, 1–1.
[20] Mandal I., Sairam N.: Accurate telemonitoring of Parkinson's disease diagnosis using robust inference system. International journal of medical informatics 82(5), 2013, 359–377.
[21] O'Sullivan S. B., Schmitz T. J.: Parkinson disease. Physical Rehabilitation 5th ed., F. A. Davis Company, Philadelphia 2007, 856–894.
[22] Pahuja G., Nagabhushan T. N.: A comparative study of existing machine learning approaches for Parkinson's disease detection. IETE Journal of Research 67(1), 2021, 4–14.
[23] Parkinson J.: An Essay on Shaking Palsy. Whittingham and Rowland Printing, London 1817.
[24] Rahman A. et al.: Parkinson's disease diagnosis in cepstral domain using MFCC and dimensionality reduction with SVM classifier. Mobile Information Systems 2021, 1–10.
[25] Rahn III. et al.: Phonatory impairment in Parkinson's disease: evidence from nonlinear dynamic analysis and perturbation analysis. Journal of Voice 21(1), 2007, 64–71.
[26] Sharanyaa S. et al.: An Exploration on Feature Extraction and Classification Techniques for Dysphonic Speech Disorder in Parkinson's Disease. Inventive Communication and Computational Technologies: Proceedings of ICICCT 2021, Springer Singapore, 2022, 33–48.
[27] Stelzig Y. et al.: Laryngeal manifestations in patients with Parkinson disease. Laryngo-rhino-otologie 78(10), 1999, 544–551.
[28] Teixeira J. P. et al.: Vocal acoustic analysis–jitter, shimmer and hnr parameters. Procedia Technology 9, 2013, 1112–1122.
[29] Teston B.: L'évaluation objective des dysfonctionnements de la voix et de la parole; 2e partie: les dysphonies. Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA) 20, 2001, 169–232.
[30] Tiwari V.: MFCC and its applications in speaker recognition. International journal on emerging technologies 1(1), 2010, 19–22.
[31] Van Den E. et al.: Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity. American Journal of Epidemiology 157(11), 2003, 1015–1022.
[32] Wendahl R. W.: Laryngeal analog synthesis of jitter and shimmer auditory parameters of harshness. Folia Phoniatrica et Logopaedica 18(2), 1966, 98–108.

M.Sc. Oumaima Majdoubi
e-mail: oumaima_majdoubi@um5.ac.ma

Received the master's degree in Electrical Engineering from the National School of Arts and Crafts of Rabat, Mohammed V University, Rabat, Morocco, in 2022. She is a research student of sciences and Technology of the Engineer in Ecole Nationale Supériure d'Informatique et d'Analyse des Systémes ENSIAS. She is part of E2SN research team at ENSAM-RABAT. She is interested in the detection of neurological diseases based on artificial intelligence and signal processing.

http://orcid.org/0009-0000-2968-7975

Prof. Achraf Benba
e-mail: achraf.benba@ensam.um5.ac.ma

Received the Ph.D. degree in electrical engineering from ENSAM, Mohammed V University, Rabat, Morocco, in 2017. He is a member of E2SN teams at ENSAM, Mohammed V University. His interests are in speech processing for detecting people with neurological disorders and Cardiac.

http://orcid.org/0000-0001-7939-0790

Prof. Ahmed Hammouch
e-mail: ah.hammouch@gmail.com

Received the M.S. degree and the Ph.D. degree in automatic, electrical, electronic by the Haute Alsace University, Mulhouse, France, in 1993, and the Ph.D. degree in signal and image processing by the Mohammed V University, Rabat, Morocco, in 2004. From 1993 to 2013, he was a professor in the Mohammed V University, Rabat, Morocco. Since 2009 he manages the Research Laboratory in Electronic Engineering. He is an author of several papers in international journals and conferences. His domains of interest include multimedia data processing and telecommunications. He is with National Center for Scientific and Technical Research, Rabat, Morocco.

http://orcid.org/0009-0005-8691-6662

http://doi.org/10.35784/iapgos.3745 received: 14.06.2023 | revised: 18.07.2023 | accepted: 26.09.2023 | available online: 30.09.2023

DEVELOPMENT OF THE POWER SUPPLY AND CONTROL SYSTEM FOR THE HEMODIALYSIS MACHINE

Volodymyr Yaskiv, Anna Yaskiv
Ternopil Ivan Puluj National Technical University, Faculty of Applied Information Technologies and Electrical Engineering, Ternopil, Ukraine

Abstract. *The article describes new approaches to creating an autonomous compact system with automatic control for hemodialysis. It is proposed to organize a closed circuit for cleaning the dialysis solution using an electrolytic regenerator as a function of the concentration of urea in it. The functional diagram of the created system is presented and described. To power the regenerator, ensure thermal stabilization of the solution, and power auxiliary electronic and electrical equipment, a multi-channel power supply and control system for the hemodialysis machine based on high-frequency magnetic amplifiers has been developed and researched. The advantages of power switches based on high-frequency magnetic amplifiers in comparison with transistor switches, including in the construction of controlled power sources, are given. The principle of operation of the voltage regulator on high-frequency magnetic amplifiers is described. Theoretical and experimental oscillograms are given. Photographs of the experimental unit as well as an industrial sample of the multi-channel power supply and control system of the hemodialysis machine are provided. Their main technical characteristics are given. Conclusions to the conducted work are formulated. Carrying out the regeneration of the dialysis solution significantly reduces its costs – 2 liters of solution, which is suitable for use for 6 months, is enough for the operation of the device. Existing hemodialysis machines are a stationary open system using a single-use dialysis solution at a rate of up to 35 l/h, which ties the machine to stationary clinical conditions. Introducing feedback on the concentration of urea in the dialysis solution allows you to automate the blood purification procedure, as well as automatically complete hemodialysis at the necessary time, and also eliminates the dependence of the device on the conditions of the hospital.*

Keywords: hemodialysis, power supply and control system, rectangular hysteresis loop, high-frequency magnetic amplifier

ROZWÓJ UKŁADU ZASILANIA I STEROWANIA URZĄDZENIA DO HEMODIALIZY

Streszczenie. *Do zasilania regeneratora, stabilizacji termicznej roztworu oraz zasilania pomocniczych urządzeń elektronicznych i elektrycznych opracowano i zbadano wielokanałowy układ zasilania i sterowania aparatu do hemodializy oparty na wzmacniaczach magnetycznych wysokiej częstotliwości. Przedstawiono zalety wyłączników mocy opartych na wzmacniaczach magnetycznych wysokiej częstotliwości w porównaniu z wyłącznikami tranzystorowymi, w tym w budowie sterowanych źródeł prądu. Opisano zasadę działania regulatora napięcia we wzmacniaczach magnetycznych wysokiej częstotliwości. Podano oscylogramy teoretyczne i eksperymentalne. Przedstawiono fotografie jednostki doświadczalnej oraz próbkę przemysłową wielokanałowego układu zasilania i sterowania aparatu do hemodializy. Podano ich główne parametry techniczne. Sformułowano wnioski z przeprowadzonej pracy. Przeprowadzenie regeneracji płynu do dializy znacząco obniża jego koszty – do pracy urządzenia wystarczają 2 litry płynu, który wystarcza na 6 miesięcy. Istniejące aparaty do hemodializy to stacjonarne systemy otwarte wykorzystujące roztwór do dializy jednorazowego użytku z szybkością do 35 l/h, co wiąże urządzenie ze stacjonarnymi warunkami klinicznymi. Wprowadzenie informacji zwrotnej o stężeniu mocznika w płynie dializacyjnym pozwala zautomatyzować procedurę oczyszczania krwi, a także automatycznie zakończyć hemodializę w wymaganym czasie, a także eliminuje zależność urządzenia od warunków panujących w szpitalu.*

Słowa kluczowe: hemodializa, układ zasilania i sterowania, prostokątna pętla histerezy, wzmacniacz magnetyczny wysokiej częstotliwości

Introduction

In medicine, there is a problem of ensuring the functioning of the body in case of complete or partial failure of vital organs. Kidney functions in such extreme situations in clinical institutions are performed by a biotechnical system that provides artificial hemodialysis – an "artificial kidney" device. Hemodialysis treatment of patients with acute or chronic renal failure had a dramatic effect on reducing the mortality rate of these patients. The main component of the hemodialysis machine is the dialyzer. The dialyzer contains semipermeable membranes that filter the blood. In the process of hemodialysis, the blood is cleaned from metabolic products due to their diffusion into the dialysis solution. At the same time, the patient's blood (up to 3 litres) enters the dialyzer, which is separated by a special membrane. A flowing dialysis solution enters the second section of the dialyzer. By means of a biochemical reaction, there is a transition of metabolic products from the patient's blood into the dialysis solution through the membrane – a transition from an environment with a higher concentration to an environment with a lower one. Contaminated dialysis solution requires appropriate disposal. Although there are known cases of its direct discharge into the sewer.

In addition, the hemodialysis procedure itself is a complex biotechnical system, the functioning of which requires solving a number of difficult problems. Despite the progress made, this membrane therapy is still an incomplete renal replacement, as the mortality and morbidity rates remain unacceptably high. An important characteristic of the membrane is its pore size, porosity and thickness. Further research is underway to improve the mechanical behavior of thin membranes [10]. The solution to the problem of patient safety, in particular electrical safety, requires further improvement [3]. Hemodialysis machines are devices with a risk of infection. Disinfection and sterilization are the main parts of injection control in hemodialysis units [12]. During a hemodialysis cycle, which can last several hours (up to 6 hours in critically ill patients), the actual dialysis machine does not display the water displacement ratio. Therefore, another problem that is currently being solved is obtaining information about excessive hydration or dehydration of patients [16].

But the main drawback of the existing hemodialysis machines, which are widely used today in clinical practice, is that they are a stationary open system. Quality control is carried out indirectly according to the patient's condition, and the use of a single-use dialysis solution (and in large quantities – up to 35 l/h) ties the device to stationary clinical conditions.

1. Formulation of the problem

Thus, the disadvantages of existing systems for hemodialysis are:

- their use only in hospital conditions, which significantly limits the number of operating places;
- large consumption of dialysis solution;
- the need for the used dialysis solution disposal;
- lack of automatic control of the hemodialysis process.

Taking into account these shortcomings, the task of creating an autonomous compact system with automatic control for hemodialysis arises.

Its solution became possible with the appearance of a sensor for the concentration of urine in the dialysis solution. The next step is the organization of the regeneration process of this solution. For this purpose, it is necessary to create a closed loop of circulation of the dialysate solution, in which the regenerator is sequentially switched on. The regenerator works from the applied power as a function of the concentration of urine in the dialysis solution. In this way, it is possible to obtain a closed

system of regulation and control of the concentration of urine in the dialysis solution. However, at the same time, the problem of creating a controlled power source for the regenerator arises. The use of existing methods of building pulse power sources based on modern semiconductor elements is limited by a number of factors:

- inability to implement multi-channel sources with equivalent and independent output channels;
- semiconductor commutation elements can switch only constant voltage [5, 1];
- the complexity of the switch control system, which also requires its own power supply;
- impossibility power source and control functions implementation in one converter on the same switch.

To achieve this goal, it is proposed to use the methods developed by the authors for building sources and power supply systems based on high-frequency magnetic amplifiers (MagAmp). The magnetic cores of MagAmps are made of an amorphous alloy with a rectangular hysteresis loop [9, 11, 14].

2. Principles of MagAmp operation

In the controllable power supplies it is suggested to use high-frequency magnetic amplifiers (MagAmps) as the switching elements [1, 2, 4, 6-8, 15, 17-20].

MagAmp switch is just a coil wound on a core with a relatively square B-H characteristic as shown in Fig. 1. The MagAmp is operated in two operating regions, either along the steep or along the shallow slopes of the B-H curve [19]. In the vicinity of operating point R the MagAmp core is unsaturated. Here, the high permeability of the core causes the MagAmp winding to present a high inductance to the circuit, which allows only a trace current to flow. MagAmp requires a certain volt-sec, which is the integral of voltage over time, to be applied to its terminals for the magnetic flux to build up in the core and reach the saturation level. The stronger the reset field, the more volt-secs the MagAmp can withstand until saturated. Once saturated, MagAmp operating point shifts to L, see Fig. 1. Here, the permeability of the core is very low and the inductance of the MagAmp coil has only a negligible value, which allows a large current flowing in the circuit [18].

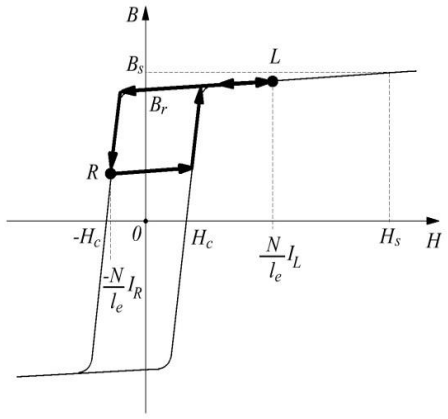


Fig. 1. MagAmp switching trajectory in B-H plain [12]

A MagAmp can be used as a semi-controlled switch that can block and delay the applied voltage. However, MagAmp cannot interrupt the current once started. Hence, MagAmps are used in pulse circuits where they are assisted by diode rectifiers, which cut off the current as the applied voltage changes polarity [18].

Here, the MagAmp post-regulator is implemented as a push-pull full-wave circuit at the secondary of the power transformer, T, as shown in Fig. 2.

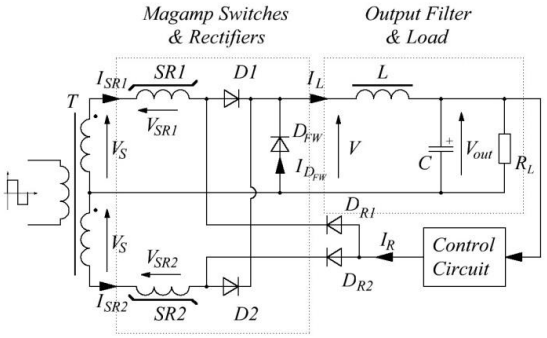


Fig. 2. Push-pull MagAmp regulator circuit [12]

The post-regulator is comprised of a pair of the MagAmp switches SR1 and SR2; power rectifier diodes D1-D2; a free-wheeling diode D_{FW} and a second order LC output filter. As compared to a single ended MagAmp circuit, the push-pull MagAmp configuration has the advantages of the symmetrical operation and doubled dc output voltage.

Such MagAmp switch combined with a rectifying diode and demagnetizing circuit forms a synchronous rectifier, whose control circuit contains just 1-2 active components. Power-width modulation in such switch is achieved through the regulation of the ratio of the time the core is in saturated state to the time the core is unsaturated. This regulation is achieved due to the change of the core demagnetization depth that is a function of the controlled parameters or the regulation algorithm.

Here, both MagAmp switches operate with current reset. The reset current I_R, is provided by the controller. The steering diodes D_{R1} and D_{R2} naturally steer the reset current I_R, towards the device to be reset.

Idealized waveforms of the push-pull MagAmp postregulator are illustrated in Fig. 3.

The upper trace is the transformer secondary winding voltage, V_S. The second and third traces show the MagAmp voltages, V_{SR1}, V_{SR2}. The positive portion of the MagAmp waveform represents the volt-secs blocked by the MagAmp, whereas the negative portion of the waveform is due to the reset process. As a result of the MagAmp blocking, the MagAmp output voltage, V, the upper trace is the transformer secondary winding voltage, V_S. the second and third traces show the MagAmp voltages, V_{SR1}, V_{SR2}. The positive portion of the MagAmp waveform represents the volt-secs blocked by the MagAmp, whereas the negative portion of the waveform is due to the reset process. As a result of the MagAmp blocking, the MagAmp output voltage, V, has a reduced duty cycle and, therefore, has a reduced average value. Fig. 6 also shows the output filter inductor current, I_L, and the currents of the MagAmp switches, I_{SR1} and I_{SR2}. The MagAmp blocking state current is drawn out of proportion for illustration purposes. The free-wheeling diode current, $I_{D_{FW}}$, provides the conduction path for inductor current when the MagAmps are in the blocking state.

MagAmp operation principles, analysis of the processes in a regulator based on MagAmp, comparative analysis of MagAmps and transistor switches are described in detail in the following literature [4-8, 15, 17-20].

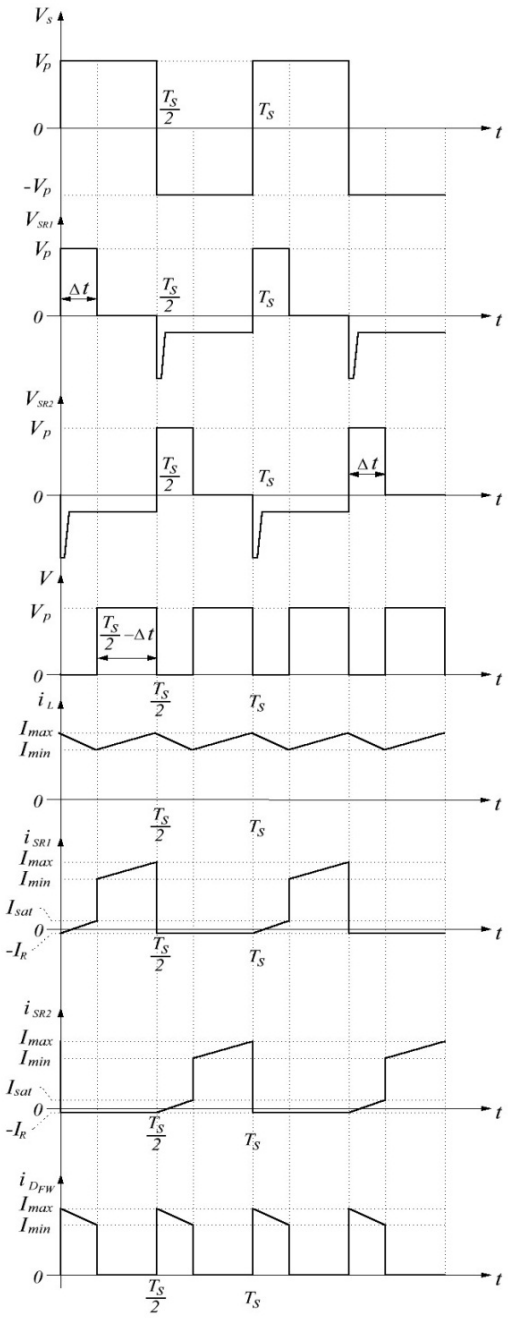


Fig. 3. Theoretical waveforms of a push-pull MagAmp regulator [16]

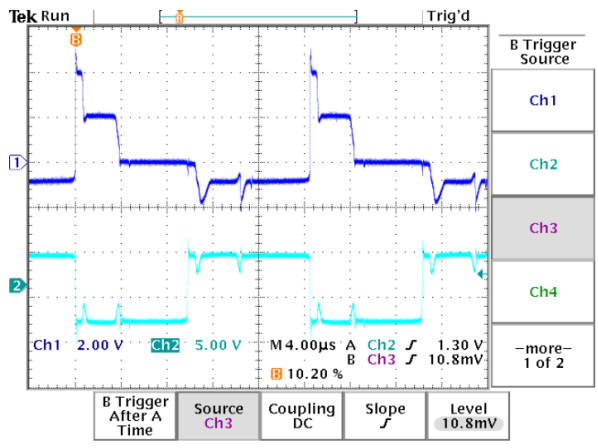


Fig. 4. Real waveforms for the MagAmp voltage regulator (at output parameters 24 V, 4 A): 1 – input voltage; 2 – MagAmp voltage

The real waveforms of the input voltage for the MagAmp voltage regulator (at output parameters 24 V, 4 A) and the voltage directly on the saturation core of MagAmp are presented in Fig. 4.

The advantages of the magnetic switch in comparison with the transistor switch are the following:

- is an AC voltage switch;
- not critical to the form of input voltage;
- gain on a current up to 1000;
- simplicity of the control circuit (1-2 transistors in linear mode);
- does not create electromagnetic interferences;
- is the filter of input interferences (both in non-saturated, and in a saturated condition);
- high efficiency (99 %), the losses do not depend on load current;
- high level of radiation stability and mechanical stability;
- does not require protection (itself serves as a protection device of high-frequency transistor inverter);
- is multifunctional: the power amplifier, power switching device, pulse-width modulator, executes functions of the integrator, comparator, protection device.

It is the advantages of the magnetic switches that determine the advantages of MagAmp power converters compared to traditional transistor power converters. They are the following:

- possibility of implementing multi-channel power converters with equivalent and independent output channels with 100% load current change range;
- allow a wide range of changes in the input voltage;
- high quality of output voltages (no high-frequency peaks and low-frequency component, minimal high-frequency component);
- high level of dynamic characteristics – the disturbance is worked out in a time equal to half a period of a high switching frequency (in the frequency domain, the MagAmp is an inertialess element with a delay of half a period of the operating frequency), the transient process ends when the regulated value (voltage) reaches the first value of its established level [17];
- lower level of electromagnetic interference;
- lower cost;
- a higher level of radiation and mechanical stability;
- a higher level of reliability due to both the physical nature of the MagAmp and the significant simplification of the circuitry;
- high level of the specific power;
- high level of unification – the possibility of using the same standard size of the MagAmp core and the same circuitry solutions to realize of the power converters in a wide range of output parameters.

3. Power supply and control system for the hemodialysis machine

Hemodialysis device, implemented according to the principle of a closed cycle, is intended for cleaning the patient's blood from harmful impurities. The block-diagram of the device is presented in Fig. 5. The essence of its action is reduced to purification by electrochemical decomposition of the dialysate solution, which provides purification of the patient's blood through the appropriate membrane equipment [19].

The patient's blood enters the dialyzer 1, which is separated by a membrane. The second part of it is filled with dialysis solution. In order to clean the dialysis solution from harmful impurities entering it through the membrane from the patient's blood, a closed loop of its circulation is organized. Depending on the concentration of urine in the patient's blood, different levels of direct current are required to maintain the reaction in the electrolyzer at a given speed. Therefore, the power regulator itself is made on the basis of high-frequency magnetic amplifiers. Such

a decision made it possible to simply implement a power source with the function of controlling its output power in the 100% range of its change on one element. The output power of this channel is a function of the urine concentration in the patient's blood.

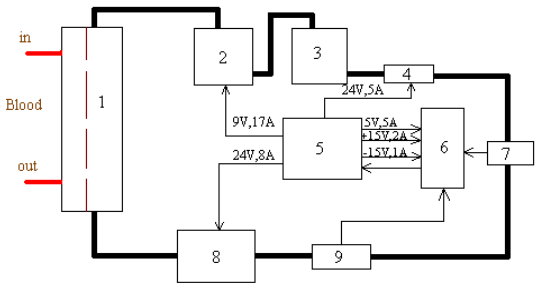


Fig. 5. Block-scheme of the hemodialysis machine (1 – dialyser, 2 – electrochemical regenerator, 3 – block of passive post-cleaning, 4 – pump, 5 – power suply, 6 – control system, 7 – sensor unit of concentration, 8 – heat exchanger, 9 – temperature sensor)

The signal from the urine concentration sensor 7 is sent to the corresponding control system 6, which matches it with the output signal (0–5 V), and also forms a signal for the completion of the hemodialysis procedure. Under the influence of this signal, the MagAmp saturation chokes are remagnetized. The main channel works as follows. When the concentration of urine in the patient's blood increases, the value of the signal sent from the concentration sensor to the control system increases. At the same time, the voltage at its output increases, which leads to a decrease in the demagnetization current. The MagAmp saturation choke is demagnetized to a smaller depth B (Fig. 1). And this corresponds to the mode of operation of the MagAmp saturation choke, in which it will reach its saturation faster in the next working half-cycle. This leads to an increase in the regenerator current. Under its influence, the intensity of decomposition processes increases, and accordingly, the purification of the dialysis solution.

In series with the regenerator 2, a block of passive post-cleaning 3, filled with a sorbent, is turned on. The speed of circulation of the dialysis solution in the circuit is set by the pump 4. Its operation is provided in two modes – automatic and manual. The output stabilized channel with output parameters of 24 V, 5 A is used to power the pump. It is implemented according to the traditional MagAmp scheme of the voltage stabilizer.

Another responsible element of this circuit is the heat exchanger 8. Its task is to maintain the temperature of the dialysis solution in the circuit at the level of the patient's body temperature. It works as a function of the signal from the temperature sensor 9 according to the same algorithm as the main channel. The executive elements of the heat exchanger are elements that use the Peltier effect. To power the heat exchanger and stabilize the temperature, the system implements the MagAmp controlled output channel with output parameters of 24 V, 12 A.

The remaining three output channels of the power supply are designed to power the microprocessor control system. One of them (5 V, 5 A) is designed as the MagAmp voltage stabilizer, the other two are linear voltage stabilizers in an integrated design.

The developed power supply and control system of the hemodialysis machine also meets electrical safety requirements. First of all, this is a double galvanic isolation. It is provided by the use of optocouplers in the control channel and the implementation of the high-frequency power transformer. In addition, the high-frequency power transformer has twice the insulation breakdown voltage value between the primary and secondary windings – instead of 2 MV, 4 MV is provided.

The functional diagram of the main channel of the power supply and control system of the hemodialysis machine is presented in Fig. 6. The method of uniform distribution of the load current between the diodes of the high-frequency rectifier was proposed and investigated, which allows to significantly reduce the time of existence of the short circuit in the circuit of the rectifier diodes during two-stroke rectification at the beginning of each half-cycle. This provides higher efficiency and improves the quality of the output voltage. At the same time, each diode VD3...VD5 (VD6...VD8) is connected to its secondary half-winding of the power high-frequency transformer TV through its MagAmp winding of the saturation choke TS1 (TS2).

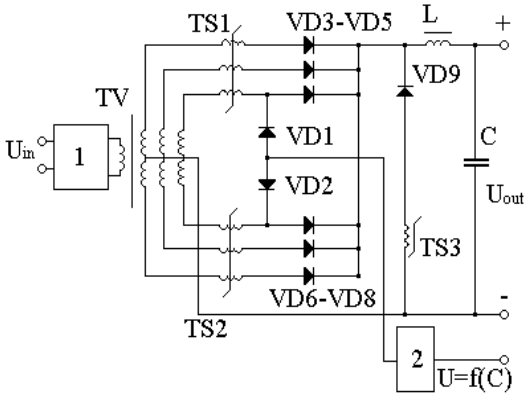


Fig. 6. Main channel of the power supply and control system (1 – nonregulated high frequency transistor inverter, 2 – MagAmp control circuit, C – concentration of the urea)

The uniformity of the distribution of the load current between the rectifier diodes at their parallel operation is determined by the technological spread of the diode parameters, namely the direct voltage drop and its frequency characteristics. The recovery time of the valve properties in the arm of the rectifier is determined by the diodes through which the largest current flows and which have the worst frequency characteristics. In addition, at the beginning of each half-cycle with push-pull rectification, there is a mode in which a short circuit occurs in the circuit of rectifier diodes. This is due to the properties of the p-n junction of the diode, namely: the closing time of the diode is longer than the opening time. The time of existence of a short-circuited circuit is determined by diode with the worst frequency characteristics and through which the largest current flows. The following experiment was conducted: five 2D213 diodes were connected for parallel operation. At an output current of 30A, short-circuited circuit time was almost 4 μs (at a half-cycle of the working frequency of 10 μs).

The following solution was proposed and investigated, which contributes to the uniform distribution of the load current between the diodes of the rectifier and which allows to significantly reduce the time of existence of the short-circuited circuit in the rectifier at the beginning of each half-period of working frquency, and, therefore, to increase the efficiency of the source and improve the quality of the output voltage. For this purpose, the winding of the saturation choke TS and the secondary winding of the power transformer TV of the transistor converter are split. The number of wires of the split winding is equal to the number of diodes in the arm of the rectifier. This method of building an output high-frequency rectifier with an output current of 30 A at an operating frequency of 50 kHz made it possible to reduce the existence time of the short-circuited circuit by several times – it was less than 1 μs. It was this method that was implemented during the construction of output channels with a high level of output current of the power supply and control system of the hemodialysis machine.

As an unregulated transistor inverter the developed power autogenerator with a saturation choke in the positive feedback circuit for the output voltage of the inverter is used (Fig. 7). This inverter was specially created to work with MagAmp voltage regulators (stabilizers). It works as a power generator with feedback from the collector current and the output voltage of the inverter. Its use of bipolar transistors is due to the

possibility of controlling their switching processes. The TS saturation choke is made of high-frequency amorphous alloy with a rectangular hysteresis loop.

The working frequency and the switching trajectory of the power transistors VT1 and VT2 of the inverter are determined by the specially created remagnetization modes of the TS saturation choke with a rectangular hysteresis loop in the positive feedback circuit for the output voltage of the inverter. The time of complete remagnetization of the saturation choke TS in the current source mode, which is formed by the voltage of the feedback winding W_U of the high-frequency power transformer TV1 and the internal resistance of the current generator, with the limitation of the speed of its remagnetization determines the half-cycle time of the inverter working frequency. Limiting the speed of magnetization is carried out by the voltage drop on the diodes VD1...VD4 and VD5...VD8 in the base circuits of the power transistors.

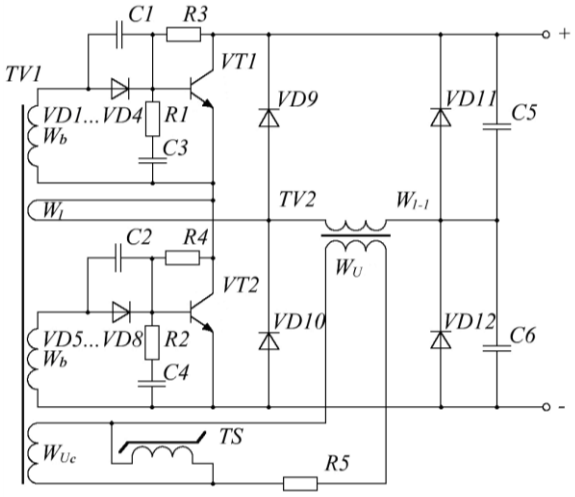


Fig. 7. The unregulated high-frequency transistor inverter

The use of a saturation choke TS with a rectangular hysteresis loop in the positive feedback circuit for the output voltage of the inverter is decisive in the operation of the inverter. Its main functions and advantages of use:

- provides an autogenerator mode of operation of power switches, which significantly simplifies the circuitry of the converter – there is no need for a special control circuit, which, in turn, often requires additional power;
- the moment of saturation of the saturation choke TS determines the moment of re-commutation of the power switches of the autogenerator. Unlike Royer's scheme, this makes it impossible for overcurrents to appear in the power circuit, which ultimately leads to a significant increase in the output power level (up to 500–700 W);
- remagnetization along the full hysteresis loop and the equality of the volt-second integrals of the rectangular hysteresis loop automatically leads to the symmetrization of the power transformer operating modes, and in dynamic modes, the perturbation recovery takes place in half the period of the working frequency;
- the remagnetization time of the saturation choke TS determines the time of the half-cycle of the operating switching frequency;
- ensures the start of the power autogenerator from a high frequency;
- prevents the occurrence of through currents in power switches;
- when the converter is overloaded, only one transistor fails;
- increases the level of electromagnetic compatibility of the converter.

The main advantages of the inverter include:

- high efficiency;
- high reliability;
- low level of the electromagnetic interference;
- high stability of the switching frequency in a wide range of input voltage change and at 100% range of the load current change;
- simplicity of the topology – no transistor control circuit, no additional power source is required for the control circuit;
- soft start with high switching frequency;
- absence of magnetization asymmetry of the power transformer;
- when overloaded, only one transistor fails;
- low cost.

MagAmp voltage regulators (stabilizers) and integrated stabilizers for powering low-power consumers are connected to the secondary circuits of the high-frequency power transformers of these inverters.

Main technical data of the power supply and control system of the hemodialysis machine:

- output power 500 W
- input voltage 220V, 50–60 Hz
- switching frequency 50 kHz
- main regulated output channel 9 V, 17 A
- stabilized output channel 24 V, 5 A
- regulated output channel 24 V, 8 A
- stabilized output channels: 5 V, 5 A
  + 15 V, 1.5 A
  - 15 V, 1 A
- efficiency > 0.85
- specific power 140 W/dm^3

Structurally, the power supply system is made in the form of two boards of exactly the same size, placed in one housing. Each of them has an unregulated transistor inverter. The output channels of the power supply of the regenerator and the pump are placed in the secondary circuits of the first. The remaining power output channels are connected to the second inverter.

An industrial sample of the board of the power supply and control system of the hemodialysis machine, on which the main channel is placed, shown in Fig. 8.

Fig. 8. Industrial sample of the power supply and control system of the hemodialysis machine on the same parameters

The appearance of the experimental prototype of the power supply and control system of the hemodialysis machine, developed and manufactured by the author of the article, is shown in Fig. 9.

Main technical data of the experimental prototype of the power supply and control system of the hemodialysis machine:

- output power 250 W
- input voltage 220V, 50–60 Hz
- switching frequency 50 kHz
- main regulated output channel 12 V, 14 A
- regulated output channel 15 V, 3 A
- stabilized output channels: 5 V, 1.5 A
  + 15 V, 1 A
  - 15 V, 1 A
- efficiency > 0.85
- specific power 120 W/dm^3

Fig. 9. Experimental prototype of the power supply and control system of the hemodialysis machine

The developed hemodialysis system with dialysis solution regeneration was put into production at the firm "B. Braun – Rolitron" (Budapest, Hungary).

Today, the methods of synchronous rectification in MagAmp power converters proposed by the authors of the article allow creating converters with an efficiency of up to 96% when powered from the industrial frequency network [17, 18].

4. Conclusion

The introduction of feedback on the concentration of urea in the dialysate solution made it possible to automate the blood purification procedure and control the moment of termination of the hemodialysis operation. Carrying out the regeneration of the dialysis solution not only significantly reduces its costs (2 litres of solution, suitable for use for 6 months, is enough for the operation of the device), but also eliminates the dependence of the device on the conditions of the hospital. This made it possible to obtain an autonomous system for hemodialysis, suitable for use even at home.

The use of high-frequency magnetic amplifiers in the role of power switching and regulating elements and methods of building multi-channel power sources based on them with controlled and stabilized output channels made it possible to create a system of power supply and control of a hemodialysis machine with a high level of operational characteristics: a high level of quality of output voltages, dynamic characteristics, efficiency, reliability, low level of electromagnetic interference and price.

References

[1] Brkovic M., Cuk S.: Novel single-stage AC-to-DC converters with magnetic amplifiers and high power factor. IEEE APEC Conf. Rec. 1, 1995, 447–453.
[2] Chen C. L., Wen C. C.: Magamp application and limitation for multiwinding flyback converter. IEE Proceedings – Electric Power Applications 152(3), 2005, 517–525.
[3] Costa T. H. et al.: A safety engineering on the design of hemodialysis systems. IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, 2018, 1–2 [http://doi.org/10.1109/ICCE.2018.8326292].
[4] Dyvak M., Yaskiv V., Yaskiv A.: Simulation and Numerical Optimization of Specific Characteristics of the Unified Range of Power Converters. 12th International Conference on Advanced Computer Information Technologies – ACIT 2022, 2022, 13–17 [http://doi.org/10.1109/ACIT54803.2022.9913076].
[5] Galad M. et al.: Comparison of parameter and efficiency of transformerless inverter topologies. IEEE Intenational Conference on Electrical Drives and Power Electronics 2015, 65–68.
[6] Harada K., Nabeshima T.: Applications of magnetic amplifiers to high-frequency dc-to-dc converters. Proc. IEEE 76(4), 1988, 355–361.
[7] Hang L.J., Gu Y.L., Lu Z.Y., Qian Z.M., Xu D.H. Magamp post regulation for LLC series resonant converter with multi-output. 31st Annual Conference of IEEE Industrial Electronics Society – IECON, 2005, 207–213 [http://doi.org/10.1109/IECON.2005.1568977].
[8] Lee J., Chen D. Y., Jamerson C. Magamp post regulators-practical design considerations to allow operation under extreme loading conditions. Proceedings of the IEEE APEC-98, 1988, 368–376.
[9] Mammano B.: Magnetic amplifier control for simple, low-cost, secondary regulation. Unitrode corporation [now Texas Instruments]. Lexington, MA 02173, 2001, SLUP129 [https://www.ti.com/lit/ml/slup129/slup129.pdf].
[10] Saliba J. et al.: Nanostructured porous silicon membrane for hemodialysis. 2nd International Conference on Advances in Biomedical Engineering, Tripoli, Lebanon, 2013, 145–147 [http://doi.org/10.1109/ICABME.2013.6648868].
[11] Saturable cores for mag-amps, – Toshiba, https://pdf.directindustry.com/pdf/toshiba-america-electronics-components/saturable-cores-mag-amps/33679-562725.html#search-en-saturable-cores-mag-amps
[12] Sezdi M., Benli İ.: Disinfection in hemodialysis systems. Medical Technologies National Congress (TIPTEKNO), Antalya, 2016, 1–4 [http://doi.org/10.1109/TIPTEKNO.2016.7863132].
[13] Sharma R.: Soft Switched Multi-Output PWM DC-DC Converter. International Journal of Power Electronics and Drive Systems (IJPEDS) 3(3), 2013, 328–335.
[14] Tape Wound Cores for Magnetic Amplifier Chokes, Nanocrystalline VITROPERM 500 Z, Preliminary Product Leaflet Vacuumschmelze, GmbH &Co. KG. [https://www.vacuumschmelze.de/fileadmin/documents/broschueren/kbbrosch/PKVP500Z_10.pdf].
[15] Wen C. et al.: Magamp Post Regulation for Flyback Converter. Proc. of IEEE Power Electron. Spec. Conf., 2001, 333–338.
[16] Yahya M. B. et al.: Measurement of water transfer changes during hemodialysis cycle. 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Monastir, 2016, 313–318
[17] Yaskiv V. et al.: Synchronous Rectificier in High-Frequency 24V/15A MagAmp Power Converter. IEEE 4th International Conference on Intelligent Energy and Power Systems (IEPS), Istanbul, 2020, 113–117.
[18] Yaskiv V. et al.: Synchronous rectification in High-Frequency MagAmp Power Converters. Proceedings of the International Conference Advanced Computer Information Technologies – ACIT 2018, Ceske Budejovice 2018, 128–131 [http://ceur-ws.org/Vol-2300/] (available 23.03.2020).
[19] Yaskiv V. et al.: System of Power Supply and Control of the Apparatus “the Artificial Kidney”. IEEE International Conference on Modern Problems of Telecommunications, Computer Science and Engineers Training TCSET’2000, Lviv-Slavsko, 2000, 162.
[20] Yaskiv V. et al.: Modular High-Frequency MagAmp DC-DC Power Converter. 9th International Conference on Advanced Computer Information Technologies (ACIT). Ceske Budejovice, 2019, 213–216.

Prof. Volodymyr Yaskiv
e-mail: yaskiv@yahoo.com

Received Doctor of Science Degree in Electrical Engineering in National Technical University "Kharkiv Polytechnic Institute" in 2021. Specialization: Semiconductor electric power converters. He is the author of more than120 scientific publications. Scientific interests: switch mode AC/DC and DC/DC power converters with high-frequency magnetic amplifiers, parallel operation, high level of the load current, resonant converters, power inverters.

http://orcid.org/0000-0003-0043-3909

Ph.D. Anna Yaskiv
e-mail: annyaskiv@gmail.com

In 2021 Anna Yaskiv defended her Ph.D. thesis at Ternopil Ivan Puluj National Technical University, specialty – Mathematical Modeling and Numerical Methods, topic Mathematical Modeling of High-Frequency Magnetic Switches for Secondary Power Supplies. She is the author of more than 30 scientific publications. Her scientific interests include mathematical modeling and computer-aided design for power electronics.

http://orcid.org/0000-0003-3101-7107

http://doi.org/10.35784/iapgos.3663 *received: 20.05.2023 | revised: 31.08.2023 | accepted: 24.09.2023 | available online: 30.09.2023*

VALIDATION OF A THREE-DIMENSIONAL HEAD PHANTOM FOR IMAGING DATA

Jolanta Podolszańska
Częstochowa University of Technology, Częstochowa, Poland

Abstract*. This paper presents the research results on the design of a three-dimensional head phantom for cone beam projection. The head model is based on a Shepp-Logan mathematical head model, which is used to simulate the operation of the CT scanner. The model is then compared with the reference data for structural similarity, reasoning, and shape. The geometric parameters of the obtained images are investigated. The reconstructed image is analyzed using the FDK method. The results show that the geometric parameters directly correlate with the number of projections. A mathematical framework of cone beam 3d reconstruction via the first derivative of the radon transform is presented.*

Keywords: computed tomography, FDK reconstruction, 3D mathematical phantom, Shepp-Logan phantom

WALIDACJA TRÓJWYMIAROWEGO FANTOMU GŁOWY DLA DANYCH OBRAZOWYCH

Streszczenie. *W artykule przedstawiono wyniki badań nad projektem trójwymiarowego fantomu głowy do projekcji wiązki stożkowej. Model głowy jest oparty na matematycznym modelu głowy Shepp-Logan, który jest używany do symulacji działania skanera CT. Model jest następnie porównywany z danymi referencyjnymi pod kątem podobieństwa strukturalnego, rozumowania i kształtu. Badane są parametry geometryczne uzyskanych obrazów. Zrekonstruowany obraz jest analizowany przy użyciu metody FDK. Wyniki pokazują, że parametry geometryczne mają bezpośredni związek z liczbą projekcji. Przedstawiono matematyczne ramy rekonstrukcji 3D wiązki stożkowej za pomocą pierwszej pochodnej transformaty radonowej.*

Słowa kluczowe: tomografia komputerowa, rekonstrukcja FDK, trójwymiarowy matematyczny model fantomu, fantom Shepp-Logan

Introduction

Using phantoms as a standard for head imaging simulation and reconstruction has allowed for examination without interference by the human body, which is crucial to avoid potential damage caused by X-rays. Specithe finally, the Shepp-Logan phantom was used to reconstruct magnetic resonance imaging (MRI) and for k-space simulation. However, the CT version is the only one to include the radiation mitigation properties of the head and brain. The phantom version of the MRI was not adapted to MR physics, so comparisons and validations of the model are not conducted based on various studies. This work focuses exclusively on the CT-adapted phantom and aims to present the existing implementations of the Shepp-Logan model and its modifications for simulation purposes.

1. 3D Cone-Beam Geometry

In CT, spiral scanning uses a specific activation protocol. This involves rotating the gantry simultaneously as the bed moves with the patient. The hypothesis is that the object scanning will remain in place, but this happens only if the source follows the spiral trajectory. In 2D geometry, the source trajectory is represented by vectors.

$$x_{focal} = \begin{pmatrix} D_1 sin\beta \\ -D_1 cos\beta \\ z_{start\_pos} \end{pmatrix}$$

The spiral scanning standard can express the source trajectory by the formula.

$$z_{focal} = \left[z_{start_{pos}}; z_{stop\_pos} = z_{start} \frac{d \cdot \beta_{stop}}{2\pi} \right]$$

For the reconstruction of spiral scanning, interpolation methods are used. A gantry with a four-angular range is required for complete scanning reconstruction to create a complete fan beam sinogram. However, the X-ray fan beam passes through these two segments at the beginning and end of the spiral path with a length of d at each angle at least once.

The inversion of the three-dimensional Radon transform R^3 is obtained using an equation.

$$f(x) = -\frac{1}{8\pi} R^3 \left[\frac{\partial^2}{\partial r^2} r(r\alpha) \right]$$

The application of the Radon transformation equation in 3D has been studied by scientists for decades. Reconstruction of projection data obtained in cone-beam geometry is a priority, as it is the geometry used in volumetric tomography with 2D detectors. A correct reconstruction is only possible if all planes intersect the radiation source's path at least once, later called the Tuy-Smith data sufficiency condition [6]. Based on Tuy's formula, Grangeat proposed a complete solution: to use the radiation derivative of the three-dimensional Radon transformation as an extract from the linear integrals extracted from the geometry of the cone beam. However, this solution is associated with the problem of numerical instability [7].

It is worth mentioning that accurate 3D reconstruction has limited practical application in the real world. The most commonly used source trajectory for most CT scanners is circular motion, which does not satisfy the Tuy-Smith sufficiency condition. To cope with the incompleteness of the data, researchers have worked on various trajectories; among them, only the circle and the spiral with constant pitch and radius have found significant applications. Therefore, approximate rather than exact algorithms are the most widely used in practical cone beam reconstruction. The most commonly used method for approximate cone beam reconstruction was derived by Feldkamp, Davis, and Kress (Feldkamp 1984 [8]).

1.1. FDK Method

In 1984, Feldkamp, Davis, and Kress introduced an algorithm for cone beam circular tomography reconstruction called the approximation method. This method is named so because the reconstructed result will differ slightly from the actual object, no matter how high the measurement resolution is. Despite this, the algorithm is favored for its simplicity and is widely used for cone beam reconstruction. Unlike the other methods, the FDK algorithm is unique in handling truncated data in the longitudinal direction. Initially designed for planar detectors, the FDK algorithm requires pre-weighting factors that depend on fan and cone angles and splicing with a ramp filter.

$$\bar{p}^F(\beta, a, b) = \left(\frac{R}{\sqrt{R^2 + a^2 + b^2}} p^F(\beta, a, b) \right) * g^P(a)$$

The pre-weighted and filtered projections are back-projected to the reconstruction volume.

$$f_{FDK}(x,y,z) = \int_0^{2\pi} \frac{R^2}{U(x,y,\beta)} (\beta, a(x,y,\beta), b(x,y,z,\beta))\, d\beta$$

where:

$$b(x,y,z,\beta) = z\frac{R}{R + x\cos\beta + y\sin\beta}$$

When dealing with discrete cases, the sum of projection angles replaces the integral. The points in spiral geometry share similarities with those in the fan beam, with a longitudinal coordinate added to account for the detector's axial extension. The Feldkamp algorithm operates under the premise that the acquisition geometry of a conical beam should not differ from that of a multiplanar fan beam. This algorithm suits a planar detector system, where projection data can be interpolated onto a Cartesian grid. In contrast, Schaller's 1998 paper suggests that the algorithm should not filter along straight lines but rather along curved curves on a cylindrical detector. This method's calculation is quite similar to the primary FDK method, starting with ramp filtering.

$$\bar{p}(\beta,\gamma,q) = \left(\cos\gamma \frac{R}{\sqrt{R^2+q^2}} p(\beta,\gamma,q)\right) * g(a)$$

The FDK algorithm utilizes a filtered approach to solve reconstruction tasks using computation efficiently. Due to its effectiveness, it has been widely adopted in commercial medical scanners and remains a leading technique in modern CT.

2. Two-dimensional Shepp-Logan mathematical head model

The two-dimensional Shepp-Logan mathematical head model was developed in 1974 [1] to simulate head and brain image reconstruction in CT and projection reconstruction. To mimic the head's geometric and X-ray attenuation properties, the model used ten ellipses (table 1) of varying size (grey levels) and material density.

Table 1. A function describing a phantom as the sum of 10 ellipses inside a 2×2 square

No. ellipse	Distance from the center of the image	x-axis	y-axis	Theta	Greyscale
1	(0, 0)	0.69	0.92	0	2
2	(0, -0.0184)	0.66	0.87	0	-0.98
3	(0.22, 0)	0.11	0.31	-18°	-0.02
4	(-0.22, 0)	0.16	0.41	18°	-0.02
5	(0, 0.35)	0.21	0.25	0	0.01
6	(0, 0.1)	0.046	0.046	0	0.01
7	(0, -0.1)	0.046	0.046	0	0.01
8	(-0.08, -0.605)	0.0446	0.023	0	0.01
9	(0, 0.605)	0.023	0.023	0	0.01
10	(0.06, - 0.605)	0.023	0.046	0	0.01

The original projection number was 180×160, while the current image resolution is 256×256 or 512×512.

Fig. 1. Sheep-Logan mathematical phantom with ten ellipses source: own implementation in Python language

3. Three-dimensional Sheep-Logan mathematical head model

In 1980, the Sheep-Logan phantom was updated with 17 ellipsoids to accommodate three-dimensional acquisition images and included new anatomical structures like ears, eyes, nose, and mouth. However, during the latter part of the 1980s, the 3D phantom was simplified to just ten ellipsoids by removing six anatomical regions and the blood clot region, previously named the subdural hematoma area. In 1994, the phantom was enhanced again by adding two tumor regions, bringing the total ellipsoids to 12.

3D head phantoms in imaging medicine research make it possible to simulate various clinical conditions, directly improving imaging techniques and treatment planning. For phantoms to be helpful in a clinical setting, they must be carefully validated for their similarity to accurate imaging data. Researching the effectiveness of reconstruction algorithms ithout testing them on the patient, directly exposing him to a high radiation dose, is challenging. Algorithms must undergo validation testing before being put into clinical use. Therefore, engineers responsible for reconstruction algorithms work mainly on computer simulations that mimic the operation of an accurate CT scanner [2].

Physical phantoms have been developed that fully provide a comprehensive evaluation of image acquisition. The phantom presented in three-dimensional space is an evolution of the two-dimensional phantom. It has an extension in the form of an additional dimension, represented by mathematical functions (table 2).

Table 2. 3D phantom description function, version with ten ellipses

No. ellipse	x-axis	y-axis	z-axis	a	b	c	α	Greys cale
1	0	0	0	6.9	9.20	9	0	2
2	0	0	0	0.62	8.74	8.8	0	0.98
3	-0.22	0	-0.25	0.41	0.16	0.21	-108	0.02
4	2.2	0	-2.5	3.10	1.10	2.2	-72	0.02
5	0	3.5	-2.5	2.10	2.50	5	0	0.01
6	0	1	-2.5	0.46	0.46	0.46	0	0.01
7	-0.8	-6.5	-2.5	0.46	0.23	0.20	0	0.01
8	0.6	-6.5	-2.5	0.46	0.23	0.20	90	0.01
9	0.6	-1.05	6.25	0.56	0.40	1	90	0.01
10	0	1	6.25	0.56	0.56	1	0	0.01

4. Methods

A set of different methods was used to validate the 3D head phantom. A three-dimensional model of the head phantom was created, then compared with reference data for structural similarity, reasoning, and shape. In addition, a qualitative and quantitative analysis of the phantom was performed by comparing the results obtained using accurate imaging data.

The physical CT scanner collects raw data during a patient's scan. To replicate this process, a three-dimensional Shepp-Logan model was utilized to generate data converted into raw data for a RAW extension. This extension is obtained when the scanner acquires data but has not yet undergone digital processing. A RAW image stores a broader range of dynamic and gray level scales than the final image format, containing most of the original image's information. The simulation involves two essential components: the CT scanner and the console used by the electrobiology technicians. A simulation has been developed to meet these requirements.

5. Results

Researchers utilized a Python-designed simulation environment to conduct a study. They employed a spiral beam reconstruction algorithm and processed data from a virtual head phantom into raw data. This data was then utilized to reconstruct the object further. The simulation aims to replicate the function of a CT scanner, extracting the necessary parameters for image reconstruction.

5.1. Raw data

During a scan, the X-ray beam gathers raw data, including all measured detector signals. These signals are then calibrated to account for fluctuations in lamp power and beam hardening. The attenuation properties of each X-ray signal and its correlation with beam position are also considered. Mathematical procedures like filtered back projection are used to reconstruct the CT images from the raw data. Furthermore, a different filter can reconstruct additional planes and images later. The computer uses the raw data to map local attenuation within the studied section [4].

In the axial (transverse) plane, every image comprises a grid of data points assigned a number indicating the X-ray attenuation at a specific point on the human body. This number is essentially a scaled version of the X-ray attenuation of water, referred to as a Hounsfield unit. The air is also accounted for through normalization. A CT number of -1,000 HU is assigned, and water is posted as 0 HU.

Modern CT scanners typically have a matrix size of 512×512 pixels. The size of each pixel is determined by the reconstructed field of view and the matrix size. In body applications, the pixel size is usually between 0.6–0.8 mm, while for the brain, it is approximately 0.5 mm. The pixel size varies for bone imaging, ranging from 0.3–0.5 mm.

5.2. CT and console simulation in CT

The simulation software is designed to incorporate Feldkamp-type reconstruction from precomputed projection data. It's important to note that the spiral geometry parameters are fixed and cannot be altered during reconstruction. The geometry must remain consistent with the one used for generating the pre-calculated projection data set, which will be loaded through a specific method. To better understand the parameters used during the reconstruction, refer to the image in (figure 2).

```
detector_type: DetectorType.CT
mode: Mode.CONEBEAM
pixels_per_slice_nb: 64
detector_slice_nb: 64
slice_pitch_mm: 4
detector_shape: DetectorShape.PLANAR
sdd_mm: 750
sad_mm: 500
fan_angle_deg: 30
gantry_angles_nb: 360
angular_range_deg: 360
image_matrix_size_mm: [256 256 256]
voxel_size_mm: [2 2 2]
number of projections: 1474560
```

Fig. 2. Simulation parameters in simulation software

For the simulation, each axis of the voxel size measures 2 mm. A coarse reconstruction grid will be utilized to obtain the projection in a reasonable time. The reconstruction begins with a sinogram, a 3D array of line integrals. The shape of the collection is based on the 2D reconstruction, meaning the first two dimensions are used in the 2D sinograms. When using cone beam geometry, the projection data is referred to as radiographs of a 3D object. Expanding the sinogram with an additional dimension is all that's necessary to update it. Volumetric or spatial reconstruction is another term for 3D reconstruction.

5.3. Head model visualization

To properly simulate on the console, we must input the data for the reference object (phantom) and the pre-calculated projection data. Specifically, we are working with a 3D phantom of the Shepp-Logan head, which has been sampled on a 256^3 grid. It is crucial to transpose the data in the correct order to ensure accurate results in the subsequent steps.

For this solution, a Ram-Lak filter was utilized by generating a 3D array that included all the necessary copies of the 1D Ram-Lak filter. This array was then used to filter the projection data in the frequency domain. Like the western beam, weighting factors were computed for the projection data just before implementing the Ram-Lak filter. To accomplish this, a grid of detector coordinate values was established. These weighting factors will be used on the original projection data.

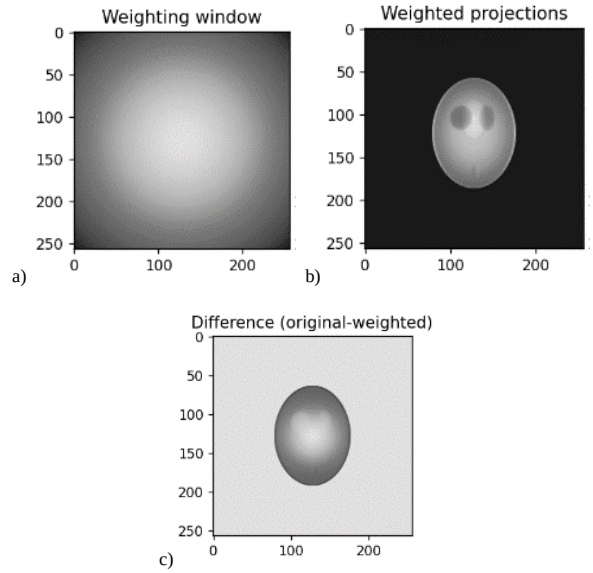


Fig. 3. Filtering: a) weighting window, b) weighted projections, c) difference between original and weighted image

In the next step, ramp filtering of the projections was performed. It is important to remember to properly arrange the input and output arrays so that the projection data is filtered line by line along the radial direction.

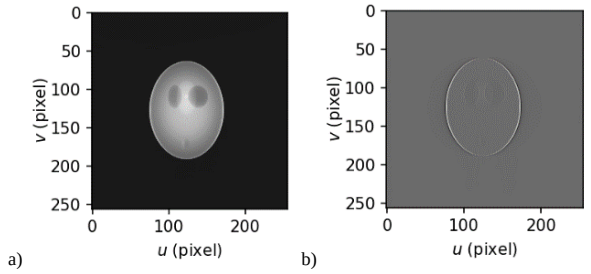


Fig. 4. Filtering: a) original model 3D Shepp-Logan, b) ramp filter applied to the model

A back projection of the cone beam was made. In this way, it will be possible to make a single view of the filtered radiographs. We will first select a single picture from the (filtered) projection data and create the auxiliary variables needed to calculate the w_BP weighting factor. We will perform a two-dimensional interpolation associated with the back projection on the cone beam.

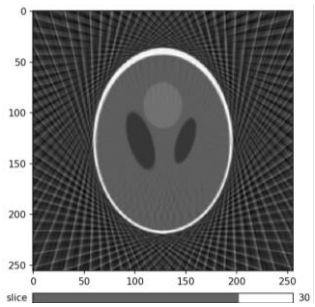


Fig. 5. Reconstruction with a limited number of projections

After a full back projection, we choose a volume with fewer cross-sections so that the reconstruction time does not burden the average computer. The number of cross-sections can be modified if we have a computer with more computing power. Of course, with the change in the number of cross-sections, the coordinate grids and weighting factors should be recalculated to cope with the new volume size.

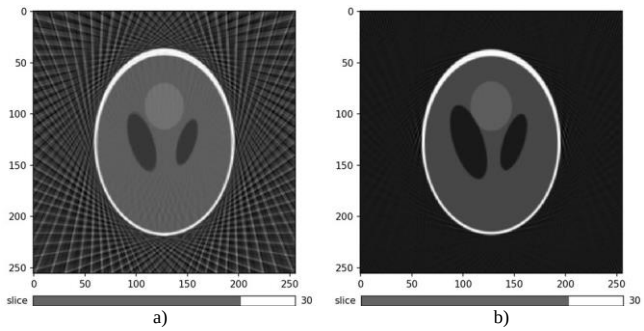


Fig. 6. Effect of reconstruction: a) without filtering, b) with filtering

The image volume was displayed using the matplotlib package (Fig. 4). The artifacts are due to solid angular subsampling. Partial volumes were reconstructed. The actual image reconstruction may take 5 to 10 minutes. The effect of the model reconstruction is shown below (Fig. 6).

6. Conclusions

This paper presents the research results on the design of a three-dimensional phantom for cone beam projection. The research included creating a simulation environment for a three-dimensional Shepp-Logan and reconstructing the projection image using the FDK method. The paper presents a mathematical description of the head model and suggestions for the designed simulation. The experimental relationship between the geometric parameters of the obtained images was established. Various simulation parameters were investigated for the reconstructed image, including a photo with limited projections and a more significant number of forecasts. The simulation environment will be expanded in the future.

Conflict of interests

The author declares no conflicts of interest.

References

[1] Batur A. et al.: Hounsfield unit density in the characterization of bile duct lesions. Polish Journal of Radiology 84, 2019, 397–401.
[2] Dzierżak R., et al.: The influence of the normalization of spinal CT images on the significance of textural features in identifying defects in the spongy tissue structure. Innovations in Biomedical Engineering. Springer International Publishing, 2019.
[3] Hansen P. C. et al.: Computed tomography: algorithms, insight, and just enough theory. Society for Industrial and Applied Mathematics, 2021.
[4] Ilmavirta J., Monkkonen K.: X-ray tomography of one-forms with partial data. SIAM Journal on Mathematical Analysis 53(3), 2021, 3002–3015.
[5] Panetta D., Camarlinghi N.: 3D Image Reconstruction for CT and PET: A Practical Guide with Python. CRC Press 2020, 65–73.
[6] Senchukova A.: Learned image reconstruction in X-ray computed tomography, 2020.
[7] Sun W. et al.: Review of high energy x-ray computed tomography for non-destructive dimensional metrology of large metallic advanced manufactured components. Reports on Progress in Physics 85(1), 2022, 016102.
[8] Tadeusiewicz R.: Komputerowe systemy wizyjne w zastosowaniach przemysłowych. Utrzymanie Ruchu 3, 2019, 14–21.
[9] Withers P. J. et al.: X-ray computed tomography. Nature Reviews Methods Primers 1(1), 2021, 18.
[10] Xu X. et al.: Review of electromagnetic vibration in electrical machines. Energies 11(7), 2018, 1779.
[11] Zuo C. et al.: Transport of intensity equation: a tutorial. Optics and Lasers in Engineering 135, 2020, 106187.

M.Sc. Eng. Jolanta Podolszańska
e-mail: jolanta.podolszanska@pcz.pl

In 2019 she started her doctoral studies at the Doctoral School of the Częstochowa University of Technology in the field of technical informatics and telecommunications. She is a research and teaching assistant at the UJD Mathematics and Computer Science Department.

Research interests: health informatics, artificial intelligence, programming in Python.

http://orcid.org/0000-0002-6032-5654

http://doi.org/10.35784/iapgos.3505 received: 17.02.2023 | revised: 29.04.2023 | accepted: 22.09.2023 | available online: 30.09.2023

OVERLOAD AND TRAFFIC MANAGEMENT OF MESSAGE SOURCES WITH DIFFERENT PRIORITY OF SERVICE

Valerii Kozlovskyi[1], Valerii Kozlovskyi[2], Andrii Toroshanko[2], Oleksandr Toroshanko[3], Natalia Yakymchuk[4]

[1]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Institute of Special Communications and Information Protection, Kyiv, Ukraine, [2]National Aviation University, Faculty of Cyber Security and Software Engineering, Department of Information Protection System, Kyiv, Ukraine, [3]Taras Shevchenko National University of Kyiv, Department of Cyber Security and Information Protection, Kyiv, Ukraine, [4]Lutsk National Technical University, Faculty of Computer and Information Technologies, Department of Electronics and Telecommunications, Lutsk, Ukraine

Abstract. *The scheme of dynamic management of traffic and activity of message sources with different priority of service is considered. The scheme is built on the basis of the neuroprognostic analysis model and the gradient descent method. For prediction and early detection of overload, the apparatus of the general theory of sensitivity with indirect feedback and control of activity of message sources is used. The control algorithm is started at the bottleneck of the network node. It uses a recursive prediction approach where the neural network output is referred to as many steps as defined by a given prediction horizon. Traffic with a higher priority is served without delay using the entire available bandwidth. Low-priority traffic will use the remaining bandwidth not used by higher-priority traffic. An algorithm for estimating the maximum available bandwidth of a communication node for traffic with a low service priority has been developed. This approach makes it possible to improve the efficiency of channel use without affecting the quality of service for high-priority traffic.*

Keywords: telecommunication network, overload prediction, sensitivity function, neural network, gradient descent method, service priority

PRZECIĄŻENIE I ZARZĄDZANIE RUCHEM ŹRÓDEŁ WIADOMOŚCI O RÓŻNYCH PRIORYTETACH USŁUG

Streszczenie. *Rozważono schemat dynamicznego zarządzania ruchem i aktywnością źródeł komunikatów o różnym priorytecie obsługi. Schemat zbudowany jest w oparciu o model analizy neuroprognostycznej oraz metodę gradientu. Do prognozowania i wczesnego wykrywania przeciążenia wykorzystuje się aparaturę ogólnej teorii wrażliwości z pośrednim sprzężeniem zwrotnym i kontrolą aktywności źródeł komunikatów. Algorytm sterowania jest uruchamiany w wąskim gardle węzła sieci. Wykorzystuje metodę predykcji rekurencyjnej, w której dane wyjściowe sieci neuronowej są odnoszone do tylu kroków, ile określono w danym horyzoncie predykcji. Ruch o wyższym priorytecie jest obsługiwany bez opóźnień z wykorzystaniem całej dostępnej przepustowości. Ruch o niskim priorytecie będzie wykorzystywał pozostałą przepustowość niewykorzystaną przez ruch o wyższym priorytecie. Opracowano algorytm szacowania maksymalnej dostępnej przepustowości węzła komunikacyjnego dla ruchu o niskim priorytecie usługi. Takie podejście umożliwia poprawę efektywności wykorzystania kanałów bez wpływu na jakość obsługi ruchu o wysokim priorytecie.*

Słowa kluczowe: sieć telekomunikacyjna, predykcja przeciążenia, funkcja czułości, sieć neuronowa, metoda opadania gradientu, priorytet usługi

Introduction

The quality of service provision in the telecommunications network is largely determined by routing algorithms, traffic management and operation in overload conditions.

In the theory of telecommunications, overload is defined as the loss of information during transmission caused by an increase in network load [3, 6, 12]. The task of overload management is to develop appropriate algorithms to prevent or reduce such loss and, first of all, each user should be provided with mechanisms for determining and receiving data from the network. For example, if a user wishes to obtain low-latency mass service, the system must provide a mechanism to achieve the goal. If the network is unable to prevent the loss of user data, then it is necessary to try to limit the loss as much as possible, and, subsequently, to try to be fair to all affected users.

The effectiveness of the traffic and overload management system largely depends on the routing methods used and the speed of the computing facilities, which should ensure minimal data delay in the network, as well as avoiding or minimizing the probability of overload [4, 13]. The accuracy of overload prediction and feedback control of traffic and data flows in high-speed computer networks is highly dependent on data transmission delays between network communication nodes.

As a result, the responses of the control commands will take effect within the network after some delay, and the control information received at the data sources or network access points may turn out to be outdated [1, 18].

There is a distinction between active and reactive overload management, which coexist in most networks. In a strict active scheme, the overload management mechanism is the reservation of network resources. In a reactive scheme, data sources need to monitor and respond to changes in the network state to prevent overload. Both management methods have their strengths and weaknesses. In systems with active management, users can be guaranteed lossless data delivery, but the number of active users must be limited. Reactive management allows much greater flexibility in the allocation of resources, but the probability of overloading increases.

The key performance indicators of the telecommunications network and the quality of user service largely depend on the traffic and overload management mechanisms used. The above conditions the relevance and necessity of research in this direction.

1. Literature review

In [18], a control method based on the sensitivity function of the performance of the telecommunications network is considered for traffic management. The sign of the performance sensitivity function provides the optimal direction for adjusting the data source speed.

In [7] proposed a traffic and overload management algorithm for mass service systems with uniform time distribution (Least Favorable Distribution, LFD). In modern telecommunication networks, flow distributions have self-similar properties, so the obtained asymptotic estimates will lead to unreasonably optimistic conclusions.

The results obtained in [7, 19] are of a general nature, their use requires new non-traditional approaches to solving the problem as a whole. Realistic estimates can be obtained by applying neural network models that must adapt to load spikes and variations in the probability distributions of application flows [18].

Increasing the memory volume of input buffers to prevent overload causes the problem of buffer bloat [19]. The number of unprocessed packets in the buffers increases, which leads to delays in their processing and can cause an avalanche process of buffer overflow and packet loss.

Works [5, 8] present a detailed classification of modern methods of overload management, identified advantages and disadvantages of their application in certain conditions of network operation. In particular, the RED (Random Early Detection) method, which is used in the TCP protocol to detect

and prevent overload, is considered. However, the mentioned works are, to a certain extent, review in nature. Analytical expressions and quantitative comparative evaluations of the RED method and other methods, in particular, Tail Drop, WRED (Weighted Random Early Detection), etc., are not provided.

In works [6, 9], the TCP Veno algorithm, which is quite common in overload management systems, was investigated. The use of this algorithm is quite effective in wireless networks with a high rate of lost packets. It tries to isolate non-overload losses so as not to include overload control algorithms where it is not needed. Meanwhile, the very method of recognizing the nature of losses according to the TCP Veno algorithm is, in fact, quite trivial (linear classifier). But it requires a detailed analysis of statistics of a sufficiently large volume, which limits its use.

2. Formulation of the problem

In works [13, 18], a model of dynamic neuroprognostic analysis using the sensitivity function is proposed for optimal traffic management and overloading of the telecommunications network [15, 17]. The algorithm is based on traffic management and the activity of message sources with different service priorities. Traffic with a higher priority is served without delay using the available part of the bandwidth. Low-priority traffic will use the remaining bandwidth not used by higher-priority traffic.

Fig. 1. shows a model of an overload control system with joint processing of traffic with different priorities [16].

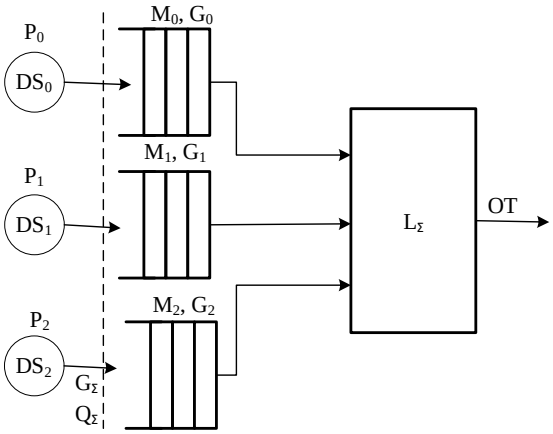


Fig. 1. Traffic management with joint processing in systems with different priorities

The vertical dashed line represents the joint control of the packet arrival rate taking into account the generalized threshold value Q_Σ for traffic with priority classes P_0, P_1 and P_2. For such a model, the total outgoing traffic L_ε is formed from the condition of analysis and taking into account the total capacity G_Σ of the output link of the node.

Fig. 2 shows a model of an overload control system with isolated traffic processing L_2 with a lower priority class [16].

A corresponding queue is organized for each type of traffic on the network node. Class P_0 has the highest priority, P_2 – the lowest.

In Fig. 1 and Fig. 2 the following designations are accepted (indexes indicate the corresponding priority):

- DS_0, DS_1, DS_2 – sources of input data;
- M0, M1, M2 – blocks of buffer memory for incoming packets;
- M_0, M_1, M_2 – the number of incoming packets with the corresponding priority, expected to be issued at time t;
- G_Σ – the total number of packets of all priorities expected to be issued at time t;
- Q_Σ, Q_2 – the threshold (limit) value of the queue length of total traffic and traffic with priority P_2,
- OT – outgoing traffic;
- HPOT – higher priority outgoing traffic;
- LPOT – lower priority outgoing traffic.

This approach allows for improvement in the efficiency of using the channel without affecting the quality of service for high-priority traffic. When the residual bandwidth changes over time, it is necessary to constantly adjust the traffic speed only of low-priority sources. This simplifies the construction of the incoming traffic control scheme. At the same time, the task of estimating the maximum available bandwidth for low-priority traffic arises.

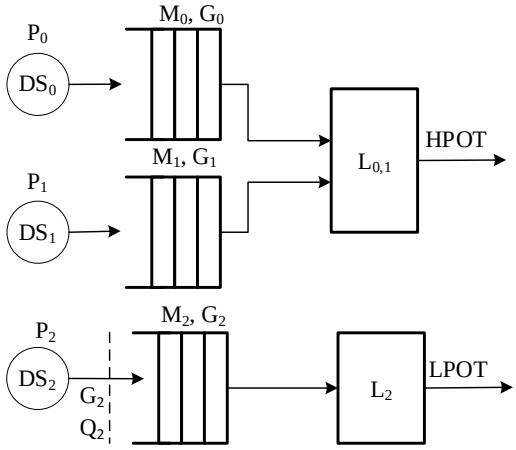


Fig. 2. Traffic management with separate processing in systems with different priority

3. Estimate available bandwidth for low-priority traffic

The structural diagram of the traffic management, forecasting and overload detection system is shown in Fig. 3. The scheme is built on the basis of a neural network using the sensitivity function of the key parameter of the network's efficiency – productivity [15, 16]. On the IR (Input Regulator) device the input traffic is regulated by the feedback control signal U.

The output queue length G is controlled by the *OR* (Output Regulator) device, taking into account the current G and the predicted $G^\wedge$ output queue length.

The deviation E of the current value of the queue size from its threshold value $Q(t)$ is carried out by the *CQ* (Control Queue) device. As a rule, the parameter Q is a constant value, i.e. $Q(t) = Q$.

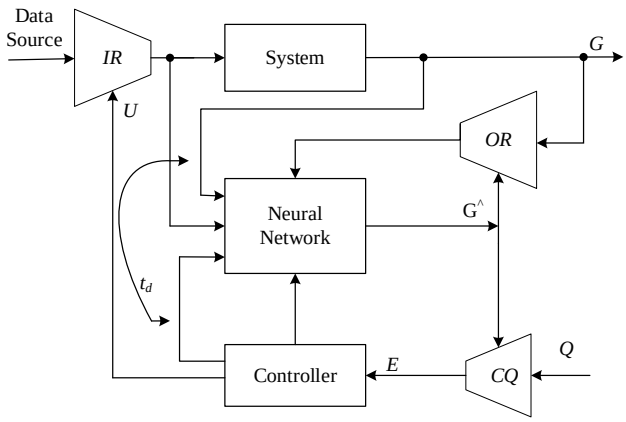


Fig. 3. Traffic control scheme

In [16], a mathematical model was developed for a transmission system to a network buffer with one data source and a fixed circular delay t_d.

We will evaluate the available bandwidth for low-priority traffic.

Let's mark $Y(t)$ as the available bandwidth at time t for low-priority traffic P_2. Available bandwidth $Y(t)$ is a function of node

usage as well as traffic $L_{0,1}$ with higher priority P_0 i P_1 (Fig. 2). The expected available rate $Y^{\wedge}(t)$ will be equal to the bandwidth not used by the high-priority traffic plus the bandwidth required to fill the buffer capacity to a given threshold value Q at a given time *t*.

A nonlinear network system can be described by the following discrete expression [10, 11]:

$$Y(t) = F\begin{bmatrix} y(t-1), \dots, y(t-1), u(t-t_d), \\ \dots, u(t-t_d-m) \end{bmatrix} \quad (1)$$

where $Y(t)$ – scalar output (i.e., service delay, queue length, etc.); $u(t)$ – scalar input (feedback control signal U); $F[*]$ is an unknown nonlinear function generated by a neural network; $1 \dots m$ are orders of functions $y(t)$ and $u(t)$; $t_d \geq 1$ – circular delay.

The task of the control algorithm is to generate the control signal $u(t)$ so that the output of the system $Y(t)$ is as close as possible to the threshold value Q. The neural model for the unknown system (1) can be expressed as

$$Y^{\wedge}(t+1) = F^{\wedge}\begin{bmatrix} y(t), \dots, y(t-B+1), u(t-t_d+1), \\ \dots, u(t-t_d-m+1) \end{bmatrix} \quad (2)$$

where $Y^{\wedge}(t+1)$ – the predicted one-step (one-step) value of the output of the neural network $F^{\wedge}$; $B > d$ – prediction horizon.

If the neural network is properly configured, then the squared error $(Y(t+1) - Y^{\wedge}(t+1))^2 = \varepsilon$ is insignificant, so the value of $Y^{\wedge}(t+1)$ is close to the predicted output systems (1). As a result, the control signal $u(t-d_t+1)$ can be chosen so that the value of $Y^{\wedge}(t+1)$ at the next step $t+1$ is as close as possible to the value of Q.

Since the neural model is an asymptotic control system, it can be used to predict the next R values of the system output at the observation interval T:

$$Y^{\wedge} = [y(t+d_t), y(t+d_t+1), \dots, y(t+B)]^T \quad (3)$$

and the error vector can be obtained as:

$$E = [e(t+d_t), e(t+d_t+1), \dots, e(t+L)]^T$$

$$e(t+i) = r(t+i) - y^{\wedge}(t+1), d_t \leq i \leq L$$

where

$$R = [r(t+d_t), r(t+d_t+1), r(t+L)]^T$$

is a further predicted values of the system output at the observation interval T.

Given the above objective function for calculating the remaining bandwidth for traffic with a lower priority can be determined through the deviation index J from the predicted value of the controlled parameter as follows:

$$J = \frac{1}{2}\sum_{i=d_t}^{L} [r(t+i) - y^{\wedge}(t+1)]^2.$$

Let us define feedback control signals U as

$$U = [u(t), u(t+1), \dots, u(t+B-1]^T.$$

Then the task of finding the optimal value of the remaining bandwidth for traffic with a lower priority is to find the values of the feedback control signals U such that the deviation index J is minimal:

$$U^{i+1} = u^i - \frac{\partial J}{\partial U^k}; \frac{\partial J}{\partial U^k} = -\frac{\partial Y^{\wedge k}}{\partial U^k} E^k$$

where $U^k(t+i)$ denotes the k-th iteration of control signal determination $u(t+i)$.

To solve this problem, you can use the following rule of gradient descent [16, 17]:

$$u^{k+1}(t+1) = u^k(t+i) + \frac{\partial Y^{\wedge k}}{\partial U^k} E^k$$

$$u^0(t+i) = u(t-1) + i;\ i = 0,1, \dots, L-1$$

k – iteration index (1, 2, …); $u^k(t+i)$ – k-th iteration of the control signal formation $u(t+i)$.

Control sequence

$$u^{k-1}(t-i), i = 0,1, \dots, B-1$$

is used to determine the predicted sequence

$$y^{\wedge k}(t+j), j = d_t, \dots, B$$

For the first iteration ($k = 1$), the control sequence starts with the value $u^0(n+i)$, which was determined in the last period. Then the deviation index J is minimized iteratively until its extremum is found.

According to the principle of receding control horizon [2, 14], only the first control signal is used to determine the control sequence U:

$$u(t+1) = u(t) + \sum_{i=d_t}^{B} e(t+i) \frac{\partial y^{\wedge}(t+i)}{\partial u(t)}$$

One iteration is performed in each period.

Thus, the algorithm for calculating the maximum available bandwidth of a communication node for traffic with a low service priority involves the following steps:

1. Determination of the prediction horizon using the developed analytical expression and based on the specified performance requirements and other key parameters of the network.
2. Determination of the threshold values of the main network parameters to ensure the specified quality of service. The described method can be used both for systems with a fixed threshold and, if necessary, for dynamic systems with time-varying thresholds.
3. Selection of feedback control functions U that minimize the difference between the expected and actual value of the controlled output of the system. Minimization is performed by a neural network based on the considered gradient descent rule.
4. Use the first element of the control function U as the next control input and repeat the entire process in the next cycle.

4. Conclusions

Dynamic neural models of traffic management in a telecommunications network with different service priorities are considered. The proposed algorithm for forecasting and prevention of node overload in a network bottleneck. It is shown that an attempt to combat overload by simply increasing the buffer capacity does not lead to a solution to the problem, but on the contrary, leads to bufferbloat and an unacceptable increase in service delays.

An algorithm and analytical expressions are developed for calculating the maximum available bandwidth of a communication node for traffic with a low service priority. The algorithm involves the following 4 steps: definition of the horizon based on the requirements for ensuring the key parameters of network efficiency; determination of threshold values of the main parameters of the network to ensure the given quality of service; the selection of control functions of the feedback *U*, which ensure the minimum difference between the expected and the actual value of the controlled output of the system; using the first element of the control function *U* as the next control input, repeat the entire process in the next cycle.

The principle of traffic separation and processing according to the established priority classes makes it possible to use the bandwidth of the network in the most optimal way without losing the quality of user service. Traffic with a higher priority is served without delay using the entire available bandwidth. Low-priority traffic will use the remaining bandwidth not used by higher-priority traffic.

The algorithm in question uses a recursive prediction approach, where the output of the neural network refers to itself to go as far as the prediction horizon requires. Another approach is a single model in which all predictions are made at the same time at different nodes in the output layer of the neural network.

For prediction and early detection of overload, the apparatus of the general theory of sensitivity with indirect feedback and control of the activity of message sources is used. The resulting solutions allow you to significantly save channel and computing resources of the network.

References

[1] Bonaventure O.: Computer Networking: Principles, Protocols and Practices. Release. 2018.
[2] Golmohammadi A.: Prioritizing Service Quality Dimensions: A Neural Network Approach. World Academy of Science, Engineering & Technology 42, 2010, 602–605.
[3] Göransson P. et al.: Software Defined Networks: A Comprehensive Approach, 2nd ed. Morgan Kaufmann, 2017.
[4] Klymash M. M., Strykhaliuk B. M., Kaidan M. V.: Teoreticheskiye osnovy telekommunikatsionnykh setyei. LAP LAMBERT Academic Publishing, Saarbrücken 2014.
[5] Korolkova A. V., Kulyabov D. S., Tchernoivanov A. I.: On the Classification of RED Algorithms. Bulletin of the Russian Peoples' Friendship University 3, 2009, 34–46.
[6] Kurose J. F., Keith W. R.: Computer Networking: A Top-Down Approach, 7th Ed. Pearson Education, Inc., 2017.
[7] Lu Z. et al.: Overload Control for Signaling Congestion of Machine Type Communications in 3GPP Networks. PLOS ONE, 2016. [http://doi.org/10.1371/journal.pone.0167380].
[8] Maximov V. V., Chmykhun S. O.: Classification of algorithms of controlling networks congestions. Scientific proceeding of Ukrainian Research Institute of Communication 5(33), 2014, 73–79.
[9] Maxymov V. V., Chmykhun S. O.: Research of the algorithm of controlling congestion TCP Veno. Telecommunication and Information Technologies 4, 2015, 30–36.
[10] Shooman M. L.: Reliability of Computer Systems and Networks – Fault Tolerance, Analysis, Design. JohnWiley&Sons, Inc., NewYork 2002.
[11] Snarskyy A. A., Lande D. V.: Modelyrovanye slozhnыkh setey. Kyiv 2015.
[12] Stallings W.: Foundations of Modern Networking: SDN, NFV, QoE, IoT, and Cloud. Pearson Education, Inc., Old Tappan, New Jersey 2016.
[13] Tanenbaum A. S., Wetherall D. J.: Computer Networks. Prentice Hall, Cloth, 2011.
[14] Tasad R., Ruggieri M.: Technology Trends in Wireless Communications. Artech House, Boston – London 2003.
[15] Tkachuk A. et al.: Basic Stations Work Optimization in Cellular Communication Network. D. Cagánová et al. (eds.), Advances in Industrial Internet of Things, Engineering and Management, EAI. Springer Innovations in Communication and Computing, 2021, 1–19.
[16] Toroshanko O. S.: Multi-step model for prognostication and detection of telecommunication network overload. Telecommunication and Information Technologies 2(63), 2019, 35–43.
[17] Toroshanko Ya. I.: Sensitivity analysis of systems of mass service on the base of model of adaptation and regulation of foreign traffic. Herald of Khmelnytskyi national university 6(243), 2016, 171–175.
[18] Vinogradov N. et al.: Development of the Method to Control Telecommunication Network Congestion Based on a Neural Model. Eastern-European Journal of Enterprise Technologies 2(9), 2019, 67–73.
[19] Vynohradov N. A., Drovovozov V. Y., Lesnaya N. N., Zembytskaya A. S.: Analyz nahruzky na sety peredachy dannыkh v systemakh krytychnoho prymenenyya. Zvyazok 1(61), 2006, 9–12.

D.Sc. Valerii Kozlovskyi
e-mail: valerey@ukr.net

Research interests: Microwave devices. Heterogeneous transmission lines.
Author of nearly 200 publications.

http://orcid.org/0000-0003-0234-415X

D.Sc. Valerii Kozlovskyi
e-mail: vv_k@nau.edu.ua

Research interests: cyber security, neural networks, traffic control.
Author of nearly 150 publications

http://orcid.org/0000-0002-8301-5501

Andrii Toroshanko
e-mail: atoroshanko@gmail.com

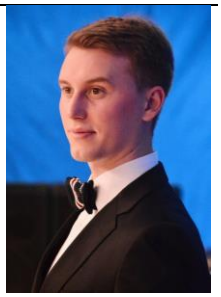
Research interests: heterogeneous networks, self-similar traffic. network congestion.
Author of 7 publications.

http://orcid.org/0000-0002-0816-657X

Ph.D. Oleksandr Toroshanko
e-mail: toroshanko@gmail.com

Research interests: wireless sensor networks, traffic control, cyber security, neural networks.
Author of nearly 30 publications

http://orcid.org/0000-0002-2354-0187



Ph.D. Natalia Yakymchuk
e-mail: n.yakymchuk@lntu.edu.ua

Research interests: diagnostics and control of the telecommunication networks state, end-to-end diagnostics, congestion management.

http://orcid.org/0000-0002-8173-449X

http://doi.org/10.35784/iapgos.3828 *received: 03.07.2023 | revised: 06.09.2023 | accepted: 27.09.2023 | available online: 30.09.2023*

RESEARCH ON CALCULATION OPTIMIZATION METHODS USED IN COMPUTER GAMES DEVELOPMENT

Nataliia Fedotova[1], Maksim Protsenko[1], Iryna Baranova[1], Svitlana Vashchenko[1], Yaroslava Dehtiarenko[2]
[1]Sumy State University, Faculty of Electronics and Information Technologies, Department of Information Technology, Sumy, Ukraine, [2]Lublin University of Technology, Faculty of Electrical Engineering and Computer Science, Lublin, Poland

Abstract. *In the field of computer game development, there are numerous optimization methods that help to significantly reduce the number of calculations that the game system performs while playing the game. In its turn, this allows to display increasingly realistic graphics. The paper presents the performed analysis of general optimization methods used in the game engine Unreal Engine, such as Distance Culling, Occlusion Culling, Frustum Culling, LODs, Level Streaming, and the Nanite System. The main factors such as resource intensity, visual quality, number of objects, and scale have been determined. The research results demonstrate that properly applied optimization methods can improve game performance and reduce the computational load on the system, which is crucial both functionally and aesthetically.*

Keywords: UE5, optimization methods, games development

BADANIE METOD OPTYMALIZACJI OBLICZEŃ STOSOWANYCH W TWORZENIU GIER KOMPUTEROWYCH

Streszczenie. *W dziedzinie tworzenia gier komputerowych istnieje wiele metod optymalizacyjnych, które pozwalają znacznie zredukować liczbę obliczeń, jakie wykonuje system gry podczas renderowania, co z kolei pozwala na wyświetlanie coraz bardziej realistycznej grafiki. W pracy przeprowadzono analizę ogólnych metod optymalizacji stosowanych w silniku gier Unreal Engine, takich jak Distance Culling, Occlusion Culling, Frustum Culling, LODs, Level Streaming oraz Nanite System. Zidentyfikowano główne czynniki, które mają znaczenie: zużycie zasobów, jakość wizualna, liczba obiektów i skalowalność. Wyniki badań pokazują, że prawidłowo zastosowane metody optymalizacji mogą poprawić wydajność gry i zmniejszyć obciążenie systemu obliczeniowego, co jest istotne zarówno pod względem funkcjonalnym, jak i estetycznym.*

Słowa kluczowe: UE5, metody optymalizacji, tworzenie gier

Introduction

A computer game is a complex program that involves a large number of mathematical calculations. Each object in the game has its own set of unique geometric (shape, number of polygons) and physical parameters (weight, size, density) and optical properties of the object's surface, which characterize the its ability to reflect light and shadow. Additionally, sometimes it is necessary to consider the partial or complete destruction of the object. Each of these aspects requires to be calculated at least once for each frame, which are sequentially displayed on the monitor. Depending on the game's characteristics, this can occur at 30, 60, 120, or even 240 frames per second (fps).

In modern games, the number of objects with the same model can reach tens of thousands, and the number of polygons on a single model can reach millions. The resources of gaming devices are not infinite, which necessitates reducing the load on the computer's hardware resources used by the game. At different stages of game development, optimization methods corresponding to that stage are employed. The developer must understand how and which optimization method or their combination affects the game's performance. Performance optimization in games aims to reduce the time required to display a game frame and increase the number of frames per second. This ensures a more convenient and smooth gaming experience for the user and overall comfort during gameplay.

This research aims to examine the performance optimization methods used in the computer game industry and to analyze the most popular and effective of them.

A game prototype using the Unreal Engine 5 (UE5) game engine environment, where optimization methods will be tested, was developed to evaluate the effectiveness of existing methods. The testing will be conducted in two scenarios. The first one will be carried out without utilizing optimization methods. And the second scenario will be performed with implementing the optimization method. Each method will be tested separately.

When analyzing the methods' effectiveness, it is essential to consider the following aspect. Practically all existing or new games contain many geometric objects and their textures, complex shaders for model visualization, and numerous intricate visual effects. However, computers have certain technical limitations, such as available memory capacity for the game, frame rate restrictions, or limits on the number of objects that can be displayed on the screen. Therefore, solving this issue is urgent.

1. Optimization methods

It should be noted that all users' computers significantly differ in their efficiency. However, they are all united by the fact that the fewer calculations will be performed by the device, the better. One of the key tasks during the game development is to create a smooth game process. The game should be uninterrupted, and any delays should only be caused by the plot or game mechanics. Optimization methods are used exactly to achieve such gameplay by minimizing delays in calculations.

Optimization methods depend on the game engine used for the game development. For example, let's consider the most popular game engines as Unity and Unreal Engine. While Unity is more oriented towards simple indie games and mobile applications, Unreal Engine aims to create more labor-intensive games with many calculations and realistic graphics. Further, we are going to discuss optimization methods related exactly to Unreal Engine.

According to research materials [1, 5], the following main optimization methods are defined in Unreal Engine:

- Distance culling;
- Occlusion culling;
- Frustum culling;
- LODs;
- Level streaming;
- Nanite System.

Let's consider each of them in detail.

Distance culling. The essence of this method is that objects stop being displayed when the camera is at a certain distance from them. The distance can be set for each object or a group of objects.

Occlusion culling. In this method objects, which are not within the camera's field of view, are culled. A preload time is set in advance to start loading the objects slightly earlier, ensuring they are loaded by the time the player (camera) looks at them [10]. Figure 1 illustrates an example of disabling objects which are outside the player's field of view for a better understanding of the method.

Frustum culling. The principle of this method is very similar to Occlusion culling' one, but objects, which overlap by other objects, are not drawn on the screen.

In other words, the player does not see particular objects covered by other objects (Fig. 2).

LODs (Level of Detail). This method reduces the number of geometric elements that must be processed for displaying on the screen. Instead of fully loading all the fine object details, its less detailed versions are loaded depending on the distance to the object.

The model becomes more detailed when the player is close to the object and can see it clearly [6, 13] (Fig. 3).

Level streaming. This method is used to reduce memory load. It allows loading the game parts, which are the closest to the player, while other elements, which are far away, are not loaded.

If everything is done correctly, this allows creating the extensive seamless maps where the players can feel themselves like playing in a world which impresses with its scale [14]. The Level Streaming method is represented schematically by Fig. 4.

The Nanite System. This method is based on geometric objects' discretization, which allows rendering highly detailed scenes with billions of polygons in real-time.

During camera zooming out from the object, the Nanite System merges multiple model polygons into one, thereby significantly reduce the system load.

Benefits of Nanite: There is no need to optimize models according to fewer polygons. The system handles it automatically. It became possible directly to import high-quality output models, such as ZBrush sculpts and scanning the photogrammetry. Level of Detail (LOD) is processed automatically and no longer requires manual setting for individual LOD mesh. Quality loss is rare or nonexistent, especially with LOD transitions [15]. Despite the benefits, there are practical limitations to the Nanite System which remain in recent versions. For example, the objects amount with a single model in the scene, triangles amount per the mesh, material complexity, output resolution as well as efficiency should be carefully measured for any combination of content and hardware.

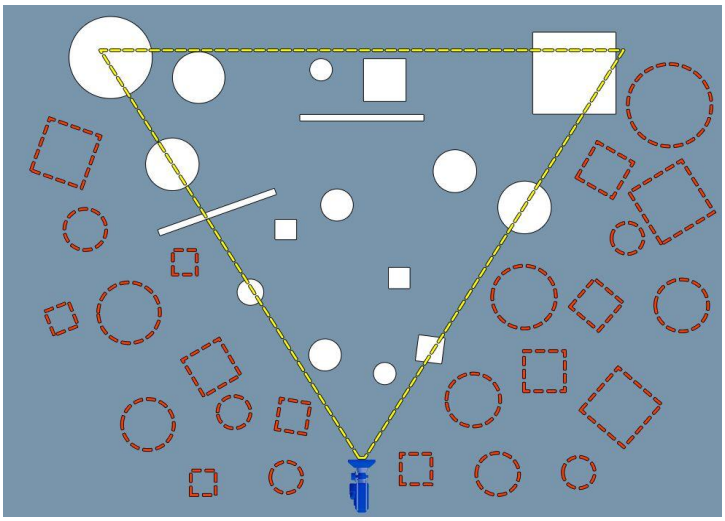


Fig. 1. Occlusion culling principle of operation [12]

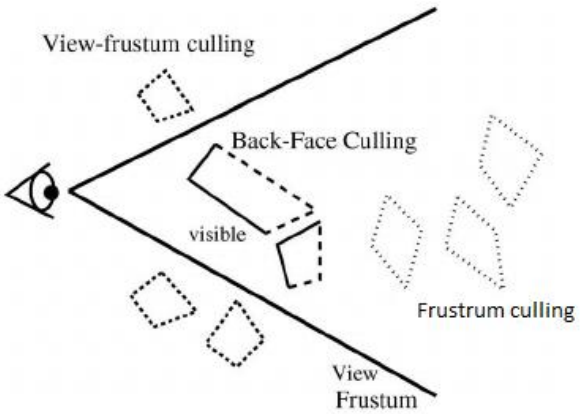


Fig. 2. Frustum culling principle of operation [8]

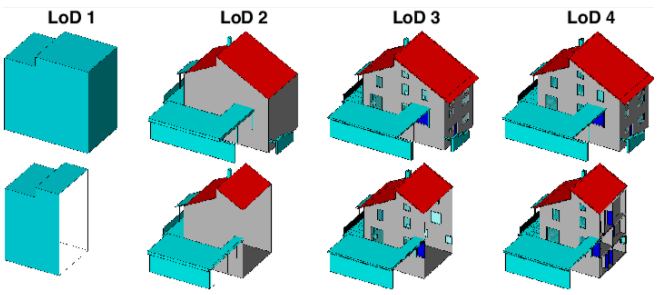


Fig. 3. Using LOD levels [2]

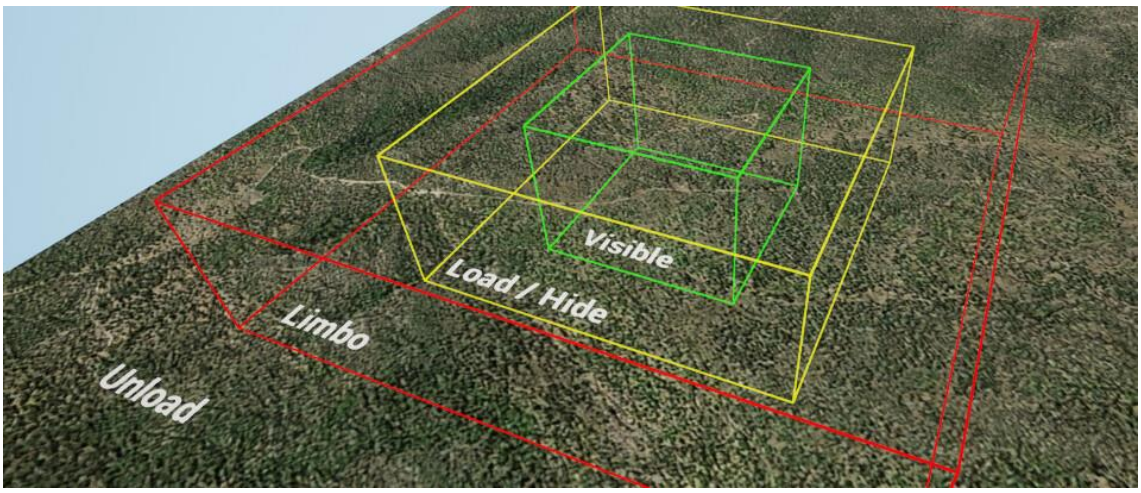


Fig. 4. Level streaming principle of operation [14]

2. Related work

The results of application the optimization methods may differ and strongly depend on the project itself. One team can bet on optimization immediately during the project development. Otherwise, the game may resemble a "dump" that will have to be sorted out for a long time. In this case, while using optimization methods, the system load can be reduced tenfold.

There is still no universal answer regarding which of the existing methods should be used for a specific project. It all depends on the game and the result the developers want to achieve. If the game is exceedingly small, it may not be optimized. If the volume is average, the simplest optimization methods can be used, such as Distance culling, Occlusion culling and Frustrum culling [7]. For large projects with various objects and large maps, it is advisable to utilize a combination of all optimization methods immediately. When dealing with a large number of identical objects in the scene, it is worth thinking about using the Nanite System [9]. In some cases, only this system can increase the number of fps (frames per second) by multiple times.

If we pay attention to a specific situation, in his book Chris Dickinson describes a case where he and his team had to optimize a particular game. In that case, it was possible to increase the average frame rate in the game from 30 to 120 fps (tests were conducted on the same computer with the same settings) [4]. What's interesting is that he didn't use all the optimization methods. He just applied the LODs and set them correctly. Since the project had many high-poly models, the increase in the frames number was fantastic. They also used other optimization methods in the game, but Chris's decision to integrate LODs played a significant role in achieving such a remarkable result.

Intel's research with Unreal Engine 4 shows a 12.2% increase in fps. Optimization was carried out only by optimizing the models and textures [16]. The same optimization principle was applied in Alberto Alvarez's work, where a simple 2D game achieved a 5% improvement in fps [3].

Certain specific games often use the same specific optimization techniques, as described in Hai Xu's article on optimizing a 3D city in Unreal Engine 5. He describes an interesting system where the city models are replaced with a volumetric map for optimization. This map consists of a city image overlaid on a height map, creating a visually appealing representation of the city (as long as the camera is not zoomed in too closely). This approach significantly increased the frame rate for his project (over 224%) [11].

Therefore, this research has been conducted to analyze and compare the effectiveness of existing optimization methods.

3. Research of optimization methods using a UE5 game prototype

3.1. Initial testing conditions

To systematize the results, we used the MSI Afterburner program [17], which provides statistics of the load on the processor, RAM, graphics adapter (video card), and the number of frames per second (fps). Additionally, the Unreal Engine's built-in tools were used to display CPU and GPU load statistics. The computer characteristics, using which the optimization methods were tested, are given in table 1.

To assess the effectiveness of the Occlusion culling, Distance culling, Frustrum culling, and Level Streaming methods, a game level was developed. It consists of a set of simple geometric shapes that have minimal impact on system load and ten fire sources, which heavily stress the graphics card. The result of applying these methods is demonstrated by exactly these fire sources. Testing the Nanite System and LODs methods was conducted on a game level which contains 1000 identical high-detailed objects. The objects of the Nanite System have 38,099 polygons, while the LODs have 960 polygons.

Table 1. Hardware characteristics

CPU	Ryzen r7 5800x
GPU	GTX 1660 super
RAM	32 Gb DDR4 3600 MHz
SSD	MSI M390 – 1Tb
Monitor	Full HD (1920x1080)

3.2. Researching the impact of the Occlusion culling method

Occlusion culling is already implemented and enabled by default in UE5. We can only turn it off and compare the characteristics of the system load with the method on and method off. To disable it, the "Occlusion Culling" checkbox should be deactivated in the project settings in the "Rendering" tab.

3.3. Researching the impact of the Flustrum culling method

Flustrum culling is also implemented and enabled by default in UE5. During the testing of Frustum culling, similar results were obtained compared to the previous tests because the operation principle is identical. The testing was conducted on the same level as previously and involved disabling objects which were not visible to the player.

Based on the increasing in frames per second, we can conclude that the load on the graphics card at least halved by using Frustum culling. Without optimization, in average we had forty frames per second, and with optimization, it increased to 90 frames per second. Since the graphics card was fully utilized by 100%, it was the main factor limited the system from generating more frames per second.

3.4. Researching the impact of the Distance Culling method

Initially, we have checked how the system behaves if the player moves far away from the load source. The system load is slightly lower compared to being close to the fire source.

Distance culling in UE5 can also be implemented, but it needs to be enabled manually. It can be configured specifically for the fire by setting each actor's Desired Max Draw Distance setting.

When we are at the specified distance from the fire, the effect disappears, and the system load returns to the level observed in the previous methods.

3.5. Researching the impact of the LODs method

To use this method, you have to create a set of LODs (Level of Detail) for the model that needs setting. This time, the test level consists of a substantial collection of spheres (2120 units), the existence of which loads the system by the calculations of light, physics, etc. Five levels of LODs were created for the sphere model using the built-in tools of UE5. The first level of the sphere model has 960 polygons, while the last one has 120 polygons.

The engine automatically switches LODs regarding the player's distance from a particular object. It can also be configured within the LOD creation menu using default parameters. For this research the system load has been tested using the default settings.

Comparison the testing results of all the mentioned optimization methods are presented in table 2. As we can see, using the LODs method results in a relatively slight decrease in system load. However, we obtained a significantly higher frame rate. This is because the RAM bandwidth of the device limited us due to using 2120 models. And since using LODs in real-time playback mode, the model's size decreases, and the load on the RAM bandwidth also decreases. This allows the system to utilize its potential better, increasing system load while achieving a higher frame rate.

3.6. Researching the impact of the Level streaming method

To investigate the impact of the Level Streaming method, we have created two sub-levels in Unreal Engine 5. The first sub-level contains elements with a high load on the graphics card, such as fire simulation or complex graphics. The second one has a trigger that disables the previous sub-level after the player reaches a certain point in the game.

This has allowed us to measure the sub-level loading time and the impact of the Level Streaming method on game efficiency. The general view of the scene is shown by figure 5. Sub-level 2 contains fire effects which cause significant resource load on the computer.

The switching of sub-levels is implemented using a trigger. The logic for the switching was described in the Blueprint of the level that contains two sub-levels (Fig. 6).

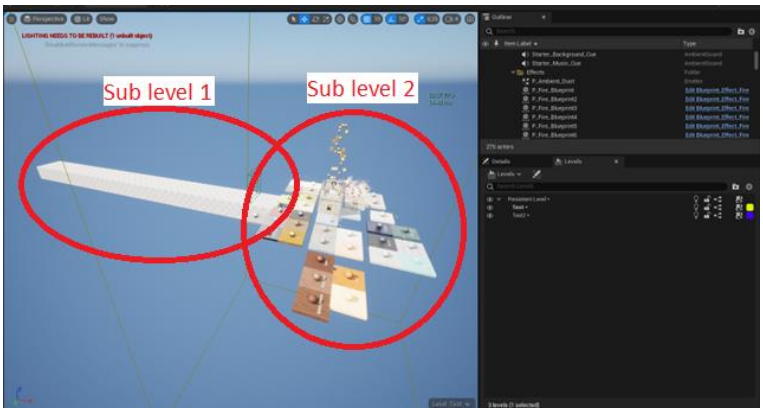


Fig. 5. General view of the level

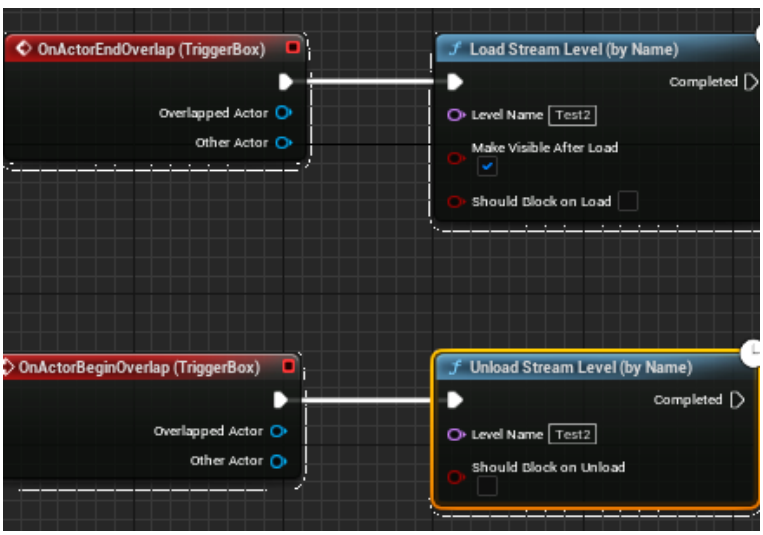


Fig. 6. Level Streaming logic implementation

Table 2. Test results

	Occlusion Culling		Flustrum culling		Distance culling		LODs		Level streaming		Nanite system	
Method usage status	Off	On	Off	On	Off	On	Off	On	Off	On	Off	On
CPU load (%)	2%	3%	2%	3%	2%	3%	5%	7%	3%	1%	3%	2%
Video card load (%)	100%	99%	100%	99%	100%	99%	39%	49%	100%	98%	100%	98%
RAM usage (MB)	17690	17773	17690	17334	17161	16792	15642	17919	17566	17556	19472	20148
The coefficient of load on the graphics card	25,27	12,78	25,27	10,62	15,98	11,22	6,7	11,89	18,7	10,27	14,92	10,66
Average number of fps	39	78	39	93	62	93	36	46	53	92	66	92
Download time (seconds)	0.3	0.3	0.3	0.3	0.3	0.23	2.65	3.05	0.3	0.3	0.33	0.45

The logic implementation involves disabling one sub-level under certain conditions, in our case, when we go beyond its boundaries. The high-load effects are located in the second sub-level, so the load decreases when we switch to the first sub-level.

When any item hits the trigger, sub-level 2 disappears, reducing the system load. Disabling the sub-level with high graphics load (fire effect) has significant impact and reduces the system load.

3.7. Researching the impact of the Nanite system method

To test the Nanite system optimization method, we have used a stone model which has 800 triangles (polygons). The Nanite system is already implemented in the engine.

We have enabled it for the specific model, waited for shaders compilation, and tested the changes. The obtained result, according to terms of frames per second, does not differ from the initial one. However, both load on the graphics card and the frame creation time have significantly decreased.

Therefore, we can conclude that the load on the graphics card has been reduced by half, and the frame rate is limited by the PC's memory or the engine itself (due to the enormous number of objects in the scene).

We have used a high-quality model with 38,099 polygons to simulate a higher load on the graphics card. We have added 100 instances of this model to the scene and observed the results without optimization (Table 2). As we can see both an increase in RAM usage and an increase in the number of frames per second take place. Thus, the Nanite System is a new way to optimize models. Although, it is not sufficiently developed for games.

It works perfectly when using high-poly models with large file sizes on the hard disk. While this may be fine in film production, it poses game challenges. In games, it is still simpler and more practical to use LOD systems, which can be generated easily in just a few clicks in UE5.

4. Discussion

Based on the information described above, game optimization is a complex process which even may not be necessary applied to all games. Game projects with various gameplay designs require developers to create combinations of optimization methods to ensure optimal game operation on different devices and for diverse audiences.

As the tendency shows, optimization is an important aspect of modern game development. It is essential to develop techniques for graphics optimization which allow reducing the load on calculations resources and maintaining game efficiency.

In addition, the increasing scale of game projects and demand changing for several games types lead to need of using the optimization to increase efficiency and provide maximum comfort for players. To achieve these aims, developers use a wide range of optimization techniques, including optimizing the processes of physics processing, reducing the number of interactions with objects in the scene, and utilizing interim software.

5. Conclusions

According to the review of optimization methods applying in the game development, it can be stated that the optimization process requires a comprehensive approach for analyzing the project and finding the most problematic areas. Reviewing popular and effective optimization methods is useful, as game development is. Based on the results of this research, it is possible to conclude the following desired scope of using the optimization methods, which were introduced within the paper framework:

- Occlusion Culling and Frustum Culling should be used in all games which render sprites and models to prevent excessive device load due to rendering these graphics.
- Distance Culling is helpful in games with open-world environments where the player can be far away from particular objects and can be turned off.
- LODs are useful for games which have detailed models.
- Level Streaming is good for large levels which may be problematic to display completely.
- Nanite System is appropriate for large projects with many high-poly models which do not have time for optimization (especially for cinematography).

References

[1] Akenine-Moller T. et al.: Real-Time Rendering. 4th ed., A K Peters/CRC Press, 2018. 1200.
[2] Akmalia R. et al.: TLS for generating multi-LOD of 3D building model. IOP Conference Series: Earth and Environmental Science 2014.
[3] Alvarez A.: Exploring Game Design through Human-AI Collaboration, 2022.
[4] Dickinson C.: Unity 5 Game Optimization. Packt Publishing, 2015.
[5] Gregory J.: Game Engine Architecture. 3rd ed., CRC Press, 2019.
[6] Hasenfratz J.-M. et al.: A survey of Real-Time Soft Shadows Algorithms. Computer Graphics Forum 4(22), 2003, 753–774.
[7] Hogan J. et al.: Analyzing Performance Issues of Virtual Reality Applications. ArXiv 2022, (abs/2211.02013).
[8] Johansson M., Roupé M., Bosch-Sijtsema P.: Real-time visualization of building information models (BIM). Automation in Construction 54, 2015.
[9] Penty C.: Behind the Scenes of The Cavern UE5 Cinematic Visual Tech Test SIGGRAPH '22. New York, USA: Association for Computing Machinery, 2022.
[10] Sekulic D.: Efficient Occlusion Culling Addison-Wesley Professional, 2018.
[11] Xu H. et al.: Efficient visualization of 3D city scenes by integrating the GIS and Unreal Engine. SPIE, 2023. 125510I.
[12] Unreal Engine 5 Documentation: Visibility and Occlusion Culling. https://docs.unrealengine.com/5.1/en-US/visibility-and-occlusion-culling-in-unreal-engine/ (available: 02 12.07.2022).
[13] Unreal Engine 4 Documentation: Creating and Using LODs. https://docs.unrealengine.com/4.26/en-US/WorkingWithContent/Types/StaticMeshes/HowTo/LODs/ (available: 12.07.2022).
[14] Unreal Engine 4 Documentation: Level Streaming Overview. https://docs.unrealengine.com/4.27/en-US/BuildingWorlds/LevelStreaming/Overview/ (available: 20.07.2022).
[15] Unreal Engine 5 Documentation. Nanite Virtualized Geometry. https://docs.unrealengine.com/5.0/en-US/nanite-virtualized-geometry-in-unreal-engine/ (available: 20.07.2022).
[16] Unreal Engine 4 Optimization Tutorial, Part 1: https://www.intel.com/content/www/us/en/developer/articles/training/unreal-engine-4-optimization-tutorial-part-1.html (available: 20.09.2022).
[17] MSI Afterburner: https://ua.msi.com/Landing/afterburner/graphics-cards (available: 20.09.2022).

Ph.D. Nataliia Fedotova
e-mail: n.fedotova@cs.sumdu.edu.ua

Associate professor of the Department of Information Technology, Sumy State University, Ukraine. Author and co-author of more than 30 scientific papers. The author`s research area focuses on e-learning systems, information technologies of design and management in complex systems, 3D modeling, visualization and animation, real-time computer graphics and simulations.

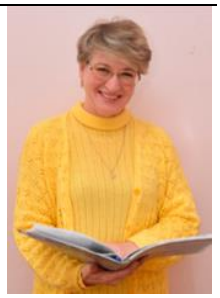
http://orcid.org/0000-0001-9304-1693

M.Sc. Maksim Protsenko
e-mail: protsenko85g@gmail.com

Bachelor and master of the Department of Information Technologies, Faculty of Electronics and Information Technologies, Sumy State University. Middle C++ developer at MoonMana Games.

http://orcid.org/00009-0008-9575-8549

Ph.D. Iryna Baranova
e-mail: i.baranova@cs.sumdu.edu.ua



Associate professor of the Department of Information Technology, Sumy State University, Ukraine. Author and co-author of more than 30 scientific papers. Research interests: 3D modeling, visualization, animation, software application for the development of systems of automated design, e-learning systems.

http://orcid.org/0000-0002-3767-8099

Ph.D. Svitlana Vashchenko
e-mail: s.vashchenko@cs.sumdu.edu.ua

Associate professor of the Department of Information Technology, Sumy State University, Ukraine.
Research interests: software engineering, use of modern information technologies (in particular, system modeling, structural-functional analysis) for the development of application software in various fields of activity, e-learning systems, and visualization.

http://orcid.org/0000-0002-7021-2629

Yaroslava Dehtiarenko
e-mail: s99277@pollub.edu.pl

Student of the Faculty of Electrical Engineering and Computer Science, Lublin University of Technology, Poland. The author`s research area focuses on optimization methods, human-computer interaction, real-time systems, computer graphics and simulations.

http://orcid.org/0009-0004-9455-7092

http://doi.org/10.35784/iapgos.4252 *received: 13.07.2023 | revised: 05.09.2023 | accepted: 27.09.2023 | available online: 30.09.2023*

ANALYSIS OF THE QUALITY OF PRINTED PLA SAMPLES USING VARIOUS 3D PRINTERS AND PRINT PREPARATION PROGRAMS

Karolina Tomczyk, Albert Raczkiewicz, Magdalena Paśnikowska-Łukaszuk
Lublin University of Technology, Faculty of Fundamentals of Technology, Lublin, Poland

Abstract. *The article presents the impact of individual settings of both 3D printers and software dedicated to the preparation of printouts. 3D printing methods are discussed. The individual methods and tools necessary for the implementation of 3D prints are described. Materials that are used in 3D printing technology are also discussed. In addition, the construction of printers on which the samples for the tests described in this work were prepared was discussed. The advantages and disadvantages of using individual slicers for 3D printing are presented. The obtained measurement results of samples made of polylactide (PLA) are presented. The work was summarized with final conclusions.*

Keywords: 3D print, PLA, slicer, 3D modelling

ANALIZA JAKOŚCI WYDRUKOWANYCH PRÓBEK PLA PRZY UŻYCIU RÓŻNYCH DRUKAREK 3D I PROGRAMÓW DO PRZYGOTOWANIA WYDRUKU

Streszczenie. *W artykule przedstawiono wpływ poszczególnych ustawień zarówno drukarek 3D oraz oprogramowania dedykowanego do przygotowania wydruków. Omówiono metody druku 3D. Opisano poszczególne metody i narzędzia niezbędne do realizacji wydruków 3D. Omówiono także materiały, które są wykorzystywane w technologii druku 3D. Ponadto omówiono budowę drukarek, na których zostały przygotowane próbki do badań opisanych w tejże pracy. Przedstawiono wady i zalety korzystania z poszczególnych slicerów do druku 3D. Zaprezentowano otrzymane wyniki pomiarowe próbek wykonanych z polilaktydu (PLA). Pracę podsumowano wnioskami końcowymi.*

Słowa kluczowe: druk 3D, PLA, slicer, modelowanie 3D

Introduction

Three-dimensional objects are created using special 3D printers, which are based on the principle of applying low layers. The surface of the printed model is smoother when the layers are smaller. This can be described as follows: the smaller the layer, the smoother the surface of the 3D print [9].

3D printing technologies include, among others:

- SLS (Selective Lase Sintering) – this is a technology based on selective laser sintering.
- SLM (Selective Laser Malting) – this is a technology that uses the method of selective laser melting [10].
- DPL (Digital Light Processing) – this is a technology that uses the light emitted by the projector to harden the liquid resin.
- EBM (Electron Beam Melting) – this is a technology that uses an electron beam to melt powders [3].
- FDM (Fused Deposition Modeling) – this is a technology based on the use of thermoplastic material and its modeling [9].

This paper focuses on the FDM method [12].

The material in the FDM technology is a filament that is fed into the print head through the extruder [8, 11]. The filament is melted there and then distributed into layers. The printer applies each subsequent layer on top of the previous one and thus creates the entire object. The technology consists in the use of a thermoplastic material, most often plastic.

The concept of 3D modeling is associated with each of these technologies. 3D modeling is a process of creating a mathematical interpretation of the surface of a given object. As a result, a three-dimensional object is obtained [6]. To create it, it need special computer software such as CAD. There are many such programs on the market, among others: Autodesk Inventor, AutoCAD or SolidWorks [12]. Before the object goes to print, a model should be created in such a program, and then the project should be saved in the appropriate format. The standard format is the STL format, which approximates the walls of the designed object to a grid made of triangles, and the software divides the model into layers. This software is a slicer. It is responsible for preparing files for printing. Thanks to it, you can, among other things, change the print speed settings or the thickness of the extruded filament layer. All applied settings will be applied to the processing of the 3D model and dividing it into individual print layers [5]. 3D printing technology is currently the most dynamically developing technology in the manufacturing process.

1. Materials and methods

The quality of the 3D print is affected by many factors, regardless of the technical parameters regarding the parameters of the program, printer, but also environmental factors such as air temperature and air flow.

1.1. ABS i PLA

The most common materials that are used in FDM 3D printing are ABS (Acrylonitrile butadiene styrene) and PLA(polylactide).

ABS is an acrylonitrile-butadiene-styrene copolymer derived from petroleum. Due to its mechanical properties, it is very popular in industry and is used to make all kinds of cases, knobs, handles, etc. This material is found in the creation of various things, e.g. furniture, pipes, car parts, household appliances, washing machines and small kitchen items, whole containers and toys. Lego bricks are the most common use case for ABS. The popularity of this thermoplastic is due to its hardness, impact resistance, wear resistance and high temperature resistance. It retains its properties in the temperature range from -20 to 80°C (some filament manufacturers give temperatures up to 98°C) [12].

PLA or poly(lactide) is especially popular with 3D printer users. This material is a bio-polymer classified as aliphatic polyester. For its production, corn starch and sugar cane. In addition to FDM technology, this material is also used in the production of various types of disposable packaging. Due to its biocompatibility and degradation time within a few years, it can also be used as a scaffold element for absorbable surgical threads, bandages, stents, orthopedic implants and tissue scaffolds. PLA is recyclable. It is believed to be biodegradable, but without special conditions such as composting. It takes a long time to decompose. The main reasons for the material's popularity are that it is easy to use and cheap [12]

Due to the low shrinkage of the material, heating the bed is not required. Does not require a heating chamber. Printing takes place at a relatively low temperature of 190–220°C. It has very good tensile strength and stiffness. All this makes it an excellent choice not only for novice users, but also for those who want to easily obtain large prints with excellent dimensional accuracy. The main disadvantage is the sensitivity to high temperatures – it deforms at 50–60°C. Therefore, items made of this material should not be used in places exposed to high temperatures (boiling water, boiler rooms, dishwashers, etc.). Prolonged exposure to UV

radiation will make it brittle. PLA is a stiff material and sensitive to high temperatures, which can make it difficult to process. This applies, for example, to grinding and drilling [1]. It is also not recommended to use solvents to smooth the print surface. PLA does not dissolve in acetone and is sensitive to substances (such as tetrahydrofuran and dichloromethane) that are not recommended for use outside of a fume cupboard.

1.2. 3D printers

The development of the rapid prototyping technique makes 3D printers more and more popular and common [4]. All models printed for the article were created using a Zortrax M200 Plus and Ender 3 V2 printer.

1.2.1. Zortrax M200 Plus

Zortrax M200 Plus is a device based on Fused Deposition Modeling (FDM) technology. Working parameters:
- technology: LPD (Layer Plastic Deposition)
- layer resolution: 90–390 microns.
- minimum wall thickness: 400 microns.
- dimensional accuracy: +/- 0.2%
- X/Y positioning accuracy: 1.5 microns.
- table leveling: automatic.
- maximum table temperature: 105℃.
- working area of the table: 200 × 200 × 180 mm [7].

Thanks to the technology of printing based on layer deposition, Zortrax M200 Plus (Fig. 1) allows you to create precise and complex models in three dimensions. The printer uses a special Z-ABS filament that is resistant to mechanical damage and chemicals, which makes it ideal for printing utility items. Zortrax M200 Plus also has many functionalities that facilitate work, such as automatic calibration of the printing table, filament sensor or intuitive user interface. In addition, the printer is equipped with a cooling system that prevents overheating of electronic components, which affects its reliability. To sum up, Zortrax M200 Plus is a solid and efficient 3D printer that is perfect for business and educational applications, where high print quality and reliability of the device are required.

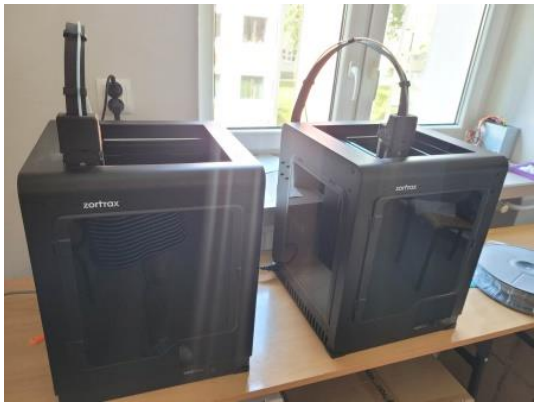


Fig. 1. Zortrax M200 Plus printer station

1.2.2. Ender 3 V2

Ender 3 V2 3D printer (Fig. 2), created by Creality. It is an improved version of the original Ender 3 which introduces several improvements and additional features. Here are the main features of the Ender 3 V2 3D printer:
- stable construction,
- large working surface,
- compatibility with various materials: the Ender 3 V2 printer supports various types of filaments, such as PLA, ABS, PETG and many others (thanks to this, the user has more freedom in choosing materials for his projects [7]),
- working area: 220 × 220 × 250 mm,
- material diameter: 1.75 mm,
- diameter of the print head: 0.4 mm (alternatively 0.2 or 0.3 mm),
- number of printheads: 1 filament: PLA, TPU, carbon filament, ABS,
- device dimensions: 475 × 470 × 620 mm,
- technology: FDM,
- device type: for self-assembly,
- minimum layer height: 100 microns,
- positioning accuracy of the XY axis: no information,
- positioning accuracy of the Z axis: no information,
- maximum table temperature: 110ºC,
- maximum head temperature: 250ºC,
- printing accuracy: 0.1 mm.

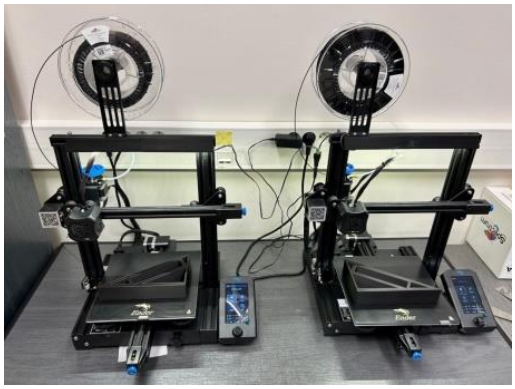


Fig. 2. Ender 3 V2 printer station

1.3. Programs for prepare and slicing a 3D model

In order for the printers to start their work, it is necessary to create a G-code. The G-code language is a programming language used in numerical machine tools. It is based on instructions that control the movement of the tool and the execution of operations on the machine. These functions require input of parameters such as the position on the XYZ axis, turning the coolant on or off, and specifying the spindle speed. Example commands in the G-code language: G91 – incremental positioning, G54-G59 – selection of the workpiece coordinate system. When modeling a 3D object, the resulting file is processed by a slicer that divides the model into layers. The next step is to generate the G-code based on the created layers. Thanks to this, the 3D printer knows, among other things, how it should move or what speed to use. Using specialized software, you can convert the .stl file created in the CAD program to the G-code format [2]. Slicer calculates how much material will be used for the 3D casting process and in what time it will be made. At the end of this process, this command language is produced.

The 3D model was prepared in Autodesk Inventor. A 5×5×5 cm cube was chosen for the settings impact study. Then the model was saved as stl. file. The model was opened in three programs for preparing a 3D print.

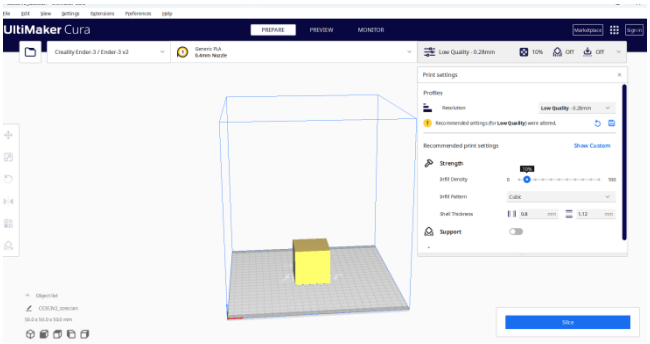


Fig. 3. UltiMaker Cura settings

The first settings were made in the Cura software. Cura is free and open-source software developed by Ultimaker. This slicer offers advanced editing functions, including rotating, cropping and scaling [12]. Cura also allows you to open a preview of the model before printing. All samples were set to 10% infill and 0.28-0.29 mm layer thickness. The settings of the first sample are shown in figure 3.

The second sample was cut in the program dedicated to the Ender 3 V2-Creality printer. Creality (Fig. 4) is the software that comes with the 3D printers of the same brand. The program has specific parameters that can be manipulated, including layer height and thickness, fill density and printing temperature.

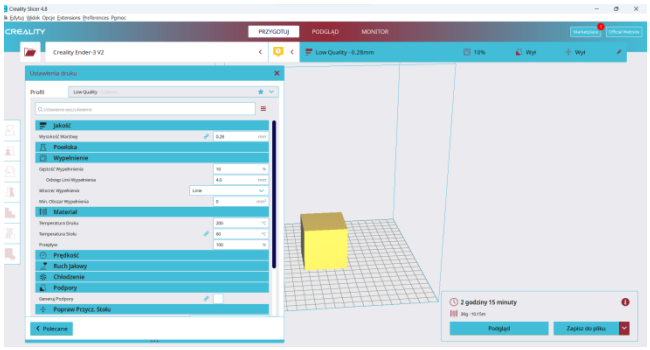


Fig. 4. Creality Slicer settings

Both samples were printed on Ender 3V2 printers. In the case of settings in UltiMaker Cura, the print time was 2h and 15 min. It was the same as in Creality slicer.

The third sample was prepared in the Zortrax – Z-Suite software. The settings are shown in figure 5.

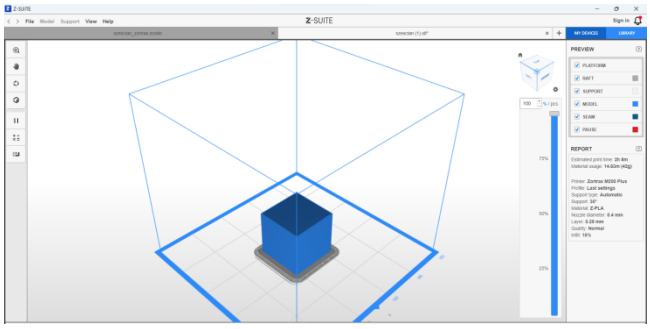


Fig. 5. Creality Slicer settings

2. Results

Table 1 presents the results of the obtained samples set in three different programs and printed on two 3D printers.

Figure 6 an 7 shows the appearance of the samples after printing.

Table 1. Results from 3D printing

3D printer	Ender 3 V2		Zortrax M200 Plus
Slicer	Cura	Creality 3D	Z-Suite
Sample	1	2	3
Material	PLA		
Height of sample	4.95 cm	4.90 cm	4.90 cm
Width of sample	4.90 cm	4.90 cm	4.90 cm
Length of sample	4.90 cm	4.90 cm	4.90 cm
Weight	28 g	28 g	33 g
Color	dark grey	dark grey	light grey
Surface	smooth	smooth	on the first layer, cavities and bulges
Layer visibility	invisible	slightly invisible	invisible

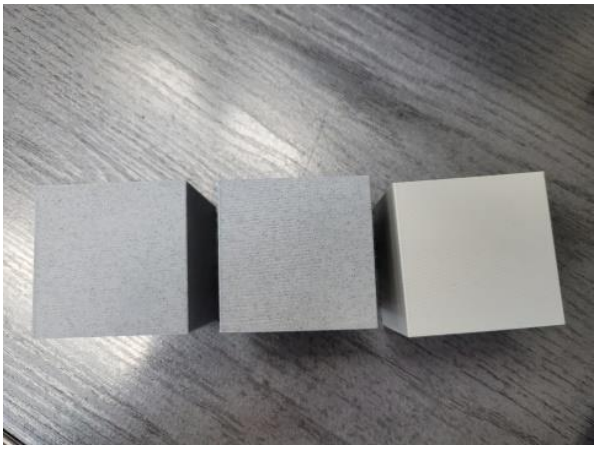


Fig. 6. All samples

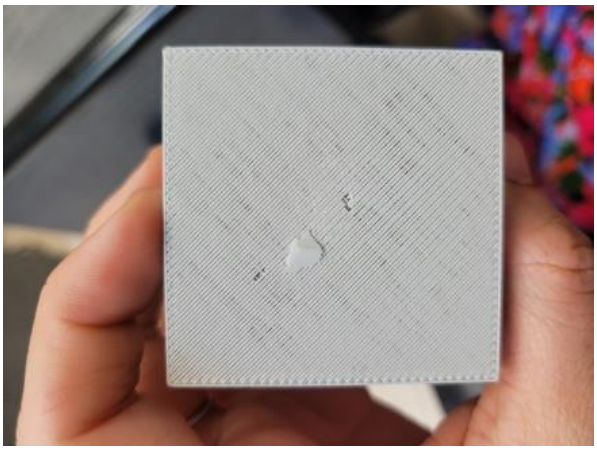


Fig. 7. Sample printing on Zotrax M200 Plus

3. Summary and conclusions

All printed samples are missing about 1mm from the full 5 cm edge of the sidewall. Although the program simulated a printout dedicated to 5×5×5 cm, the final samples have a smaller surface. This is probably due to inaccurate calibration of the printer or shrinkage of the material after cooling down. Samples prepared for printing in UltiMaker Cura and Creality 3D have the same weight. The weight of the sample printed on Zortrax is 5g higher. The samples printed on the Ender 3V2 look practically identical, but the height of the cube prepared in the Cura program is 0.05 cm higher than the other samples. The surface of samples printed on Ender 3 V2 is smooth, while the sample printed on Zortrax has some defects. The last layer has cavities, which was visible in figure 7. The printing time for samples printed on Ender 3 V2 was the same for both the Creality and Cura samples. Although the software showed a weight of 30 g, in the end they weigh 28 g. According to the settings in UltiMaker Cura, the length of material used was 0.07 m longer, but the scale shows that both samples used the same amount of PLA. The sample printed on Zortrax had a shorter print time by 7 minutes compared to samples printed on Ender 3 V2. All the programs used are quite easy to use and have a clear way to change settings. Environmental parameters such as room temperature and humidity also affect the final printout. If the material cools too quickly, the material may crack. To sum up, all the software used fulfill their roles in preparing the model for 3D printing. Economically, less material is used with Cura and Creality 3D sample settings. However, it must be remembered that the Zortrax M200 Plus printer has a different type of data recording in the form of z.code and other components that make up the printer. Selection of the appropriate program depends on availability and user preferences. To sum up, what the final 3D print will look like depends, of course, on the appropriate settings in the software, but also on the correct preparation of the printing device itself.

References

[1] Aslani K., Chaidas D., Kechagias J., Kyratsis P., Salonitis K.: Quality performance evaluation of thinwalled PLA 3D printed parts using the taguchi method and grey relational analysis. Journal of Manufacturing and Materials Processing 4(2), 2020.

[2] Brown A., De Beer D.: Development of a stereolithography (STL) slicing and G-code generation algorithm for an entry level 3-D printer. IEEE AFRICON Conference, 2013.

[3] Cichoń K., Brykalski A.: Zastosowanie drukarek 3D w przemyśle. Przeglad Elektrotechniczny 93(3), 2017, 156–158.

[4] Kiński W., Nalepa K., Miąskowski W.: Analysis of thermal 3D printer head. Mechanik 7, 2016, 726–727.

[5] Paśnikowska-Łukaszuk M. et al.: Time Distribution Analysis of 3D Prints with the Use of a Filament and Masked Stereolithography Resin 3D Printer. Advances in Science and Technology Research Journal 16(5), 2022, 242–249.

[6] Salwierz A., Szymczyk T.: Methods of creating realistic spaces – 3D scanning and 3D modelling. Journal of Computer Sciences Institute 14, 2020, 101–108.

[7] Scutaru B. et al.: Technological restrictions at 3D printing with Zortrax M200 3D printer. Tehnika 74(2), 2019, 181–187.

[8] Szmidt A., Rębosz-Kurdek A.: New approaches of improving FDM/FFF printing technology. Mechanik 90(3), 2017, 258–261.

[9] Szulżyk-Cieplak J. et al.: 3D Printers – New Possibilities in Education. Advances in Science and Technology Research Journal 8, 2014, 96–101.

[10] Tatarczak J. et al.: A review of the newest 3D printing technology for metal objects. Mechanik 90(7), 2017, 612–614.

[11] Zaburko J. et al.: Analysis of thermal operating conditions of 3D printers with printing chamber. AIP Conference Proceedings 2429, 2021.

[12] Zgryza Ł., Raczyńska A., Paśnikowska-Łukaszuk M.: Thermovisual Measurements of 3D Printing of Abs and Pla Filaments. Advances in Science and Technology Research Journal 12(3), 2018, 266–271.

Karolina Tomczyk
e-mail: s101191@pollub.edu.pl

Karolina Tomczyk, 1st year, 2nd degree student in the field of mathematics at the Lublin University of Technology. A graduate of the Maria Curie-Skłodowska University in the field of mathematics. Member of the student science club "AnimGRAF".

http://orcid.org/

Eng. Albert Raczkiewicz
e-mail: albert.raczkiewicz@pollub.edu.pl

Eng. Albert Raczkiewicz is a first year second degree student in the field of mathematics at the Lublin University of Technology. Member of the student science club “AnimGRAF”.
He is passionate about new technologies and would like to pursue his interests in the data analysis department.

http://orcid.org/

M.Sc. Magdalena Paśnikowska-Łukaszuk
e-mail: m.pasnikowska-lukaszuk@pollub.pl

Graduate of Fundamentals of Technology Faculty at Lublin University of Technology. Author of 36 scientific work.
Research interest is environmental engineering, power engineering and IT and telecommunications.

http://orcid.org/0000-0002-3479-6188

http://doi.org/10.35784/iapgos.5352 *received: 26.07.2023 | revised: 10.09.2023 | accepted: 23.09.2023 | available online: 30.06.2023*

SEGMENTATION OF MULTIGRADATION IMAGES BASED ON SPATIAL CONNECTIVITY FEATURES

Leonid Timchenko[1], Natalia Kokriatskaya[1], Volodymyr Tverdomed[1,2], Oleksandr Stetsenko[1], Valentina Kaplun[3], Oleg K. Kolesnytskyj[3], Oleksandr Reshetnik[3], Saule Smailova[4], Ulzhalgas Zhunissova[5]

[1]State University of Infrastructure and Technology, Artificial Intelligence Systems and Telecommunication Technologies Department, Kyiv, Ukraine, [2]Kyiv Institute of Railway Transport, Kyiv, Ukraine, [3]Vinnytsia National Technical University, Vinnytsia, Ukraine, [4]D.Serikbayev East Kazakhstan State Technical University, Ust-Kamenogorsk, Kazakhstan, [5]Astana Medical University, Astana, Kazakhstan

Abstract. *The article aims to study the multi-level segmentation process of images of arbitrary configuration and placement based on features of spatial connectivity. Existing image processing algorithms are analyzed, and their advantages and disadvantages are determined. A method of organizing the process of segmentation of multi-gradation halftone images is developed and an algorithm of actions according to the described method is given.*

Keywords: image segmentation, image processing, halftone images, spatial connectivity

SEGMENTACJA OBRAZÓW WIELOGRADACYJNYCH NA PODSTAWIE CECH ŁĄCZNOŚCI PRZESTRZENNEJ

Streszczenie. *Artykuł ma na celu zbadanie procesu wielopoziomowego segmentacji obrazów o dowolnej konfiguracji i rozmieszczeniu w oparciu o cechy łączności przestrzennej. Przeanalizowano istniejące algorytmy przetwarzania obrazu oraz określono ich zalety i wady. Opracowano metodę organizacji procesu segmentacji wielogradacyjnych obrazów półtonowych i przedstawiono algorytm działań zgodnie z opisaną metodą.*

Słowa kluczowe: segmentacja obrazu, przetwarzanie obrazu, obrazy półtonowe, łączność przestrzenna

Introduction

Image recognition is a relevant and promising direction in the information technologies field; its application scope is expanding every year. The recognition process includes several tasks to obtain the most accurate result, one of which is segmentation. Segmentation is the division of an image into regions based on certain features that characterize these regions or the image in general. Depending on the type of image, brightness segmentation, contour segmentation, shape or texture segmentation is used. Existing algorithms have several disadvantages, leading to the necessity of creating new learning methods [5, 6, 9]. Thus, the region augmentation method provides fairly accurate segmentation of simple scenes with a small number of objects and without texture. However, for more complex scenes, this method does not give good results. The development of methods that allow the segmentation of halftone images of arbitrary configuration and placement, considering their spatial relationships, is an urgent task. This article covers information about the developed method that allows performing effective segmentation in an algorithmically convenient way [1, 13, 15].

Purpose and tasks of the study. The article describes the algorithm of the developed method of multigradation image segmentation based on features of spatial connectivity and gives examples of its operation.

1. Materials and methods

Consider the mathematical model of the proposed segmentation method. We present the initial image in the form of a matrix A (1), which forms a set of image elements $a(m,n)$, where $m = 1..M$, $n = 1..N$.

$$A = \begin{bmatrix} a(1,1) & \cdots & a(1,M) \\ \vdots & \ddots & \vdots \\ a(N,1) & \cdots & a(N,M) \end{bmatrix} \quad (1)$$

This matrix represents the input data for the segmentation problem. The value of the elements $a(m,n)$, limited by condition: $0 < a(m,n) < C$, and belongs to the domain of non-negative integers, where C – maximum brightness value. At the output, it is necessary to obtain several areas that correspond to individual segments of the object, their combination, or the entire object. They can be represented by the corresponding output matrices in halftone

$$V_{1..L} = \begin{bmatrix} v(1,1) & \cdots & v(1,M) \\ \vdots & \ddots & \vdots \\ v(N,1) & \cdots & v(N,M) \end{bmatrix} \quad (2)$$

or binary form (2). They have the same dimension as the original matrix A and are subject to the same conditions. Thus, the domain of feasible solutions is described by an inequality $0 < v(m,n) < C$ and belongs to the domain of non-negative integers [2, 4, 10].

In the first step, the value of all elements of the original image is reduced by a defined discrete value. Find $A^1 = A - D$, where all elements of D have the same values d:

$$A^1 = A - D = \begin{bmatrix} a(1,1) - d & \cdots & a(1,M) - d \\ \vdots & \ddots & \vdots \\ a(N,1) - d & \cdots & a(N,M) - d \end{bmatrix} \quad (3)$$

Based on the obtained matrix A^1, which is a truncated image, a slice is formed (binary matrix):

$$B^1 = \begin{bmatrix} b(1,1) & \cdots & b(1,M) \\ \vdots & \ddots & \vdots \\ b(N,1) & \cdots & b(N,M) \end{bmatrix} \quad (4)$$

by:

$$b^k(m,n) = \begin{cases} 1, & a^k(m,n) = 0 \\ 0, & a^k(m,n) \neq 0 \end{cases} \quad (5)$$

These actions are repeated for the matrix A^1 and for all subsequent matrices A^k until all the elements of their respective slices B^k take zero values. Thus, a set of matrices is formed $A^1 \dots A^k$ and their corresponding sections $B^1 \dots B^k$, and the number of slices K is determined by the formula $k = \frac{C}{d}$.

An example of creating slices for 5×5 image fragment and quantization step 1 is shown in figure 1 [8].

Let's determine the value of the intra-slice connectivity for each binary slice from the obtained set as

$$\Delta^k = \sum_{\substack{m=1,M \\ n=1}} \sum_{\substack{i=-1,1 \\ j=-1}} b^k(m,n) \cap b^k(m+i,n+j) \quad (6)$$

This function can be represented by the spectrum of intra-slice connectivity. The next step is to study the obtained connectivity spectrum, namely [6, 12, 14]:

- determination of the global maximum of the function,
- determination of local maxima of the function.

Determining the global maximum is not difficult and can be given by the function:

$$\Delta'(K) = \max(\Delta^k) \quad (7)$$

where $\Delta'(K)$ – the maximum value of the intra-slice connectivity for the slice with the number K. Determining the mentioned local maxima is often complicated due to the "truncated" and stepped spectrum of the function. In this case, it will be logical to approximate the entire histogram or its section with some analytical function and determine the critical points by calculating the derivatives. The value of the maxima of the function of intra-slice connectivity can be represented [3, 7]

$$\Delta'(K) \geq \max \Delta(l_1) \ldots \geq \max \Delta(L) \quad (8)$$

where $\max \Delta(l_1) \ldots \geq \max \Delta(L)$ – the value of the local maxima of the function of the intra-slice connectivity of the slices$l_1, \ldots L$ respectively [5, 9].

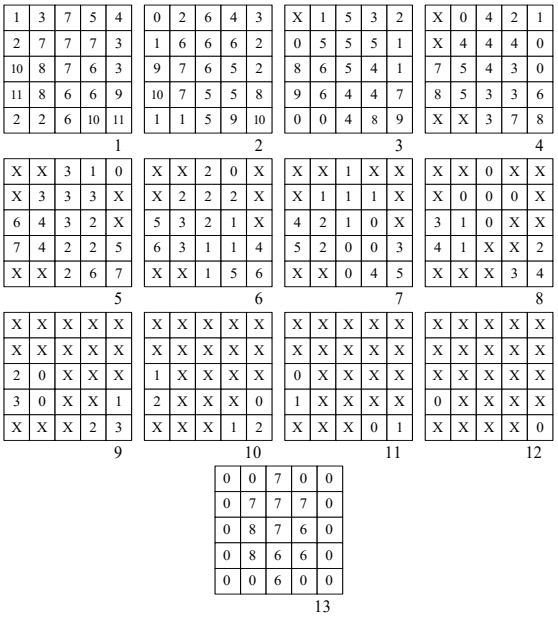


Fig. 1. An example of forming slices

For further analysis, the numbers of sections to which the maximum values belong are of interest. The mathematical description of this procedure looks like this:

$$N(\max \Delta(K)) = \begin{cases} K, \max \Delta(K) = \Delta'(K) \\ 0, \max \Delta(K) \neq \Delta'(K) \end{cases} \quad (9)$$

At the same time, the level of brightness corresponding to the slice with the number K is determined as

$$C^k = d \times K \quad (10)$$

Thus, the element of the binary slice $B(K)$, which has the brightness level C^k, will form the region maximally connected in the given image. These elements belong to the selected area, that is, they belong to one of the output functions [18, 19].

Let us form an auxiliary matrix function R^k, of dimension $M \times N$, corresponding to the binary slice $B(K)$, whose elements have the values

$$r^k(m,n) = \begin{cases} a(m,n), \ b^k(m,n) = 1 \\ 0, \ b(m,n) = 0 \end{cases} \quad (11)$$

Then $R^k(M,N)$ belongs V, if the initial image can be represented in halftone form and $B^k(M,N)$ belongs to V if a binary representation is desired. The formation of the inter-slice connectivity function is carried out for each slice with the next slice adjacent to it and has the form:

$$\Delta(k,k+1) = \sum_{n=1}^{M,N}{}_{,} \sum_{j=-1}^{1,1}{}_{,} b^k(m,n) \cap b^{k+1}(m+i,n+j) \quad (12)$$

However, in contrast to the intra-slice connectivity function, the interest is not the entire spectrum, but its individual sections. The process of selecting these areas is as follows. The slices adjacent to the slice with the maximum intra-slice connectivity, i.e., $K+1$ and $K-1$., are selected as initial slices. The values of the cross-sectional connectivity functions $\Delta(K,K+1)$ and $\Delta(K,K+1)$) are compared with the determined threshold. Based on this, the question of whether the elements of the investigated sections belong to the original function is resolved.

If $\Delta(K,K+1) > P$, then $R(K+1)$ belongs V_l or $B(K+1)$ belongs V_l, respectively, if $\Delta(K,K+1) > P$, then $R(K-1)$ belongs V_l or $B(K+1)$ belongs V_l, where P – is the limit value of the intersection connectivity for selected segments. The process of comparing the values of the intersection connectivity with each adjacent next value takes place until this condition is fulfilled. The process is carried out "to the left" and "to the right" (or "up"-"down") from the slice with the maximum value of intra-slice connectivity. The selected area will form a set of elements whose coordinates are determined by the coordinates of the single elements of the selected binary slices for the binary view (12) or halftone view (13).

$$V_l = \cdots \wedge B(K-1) \wedge B(K) \wedge B(K+1) \wedge \ldots \quad (13)$$

$$R_l = \cdots \wedge R(K-1) \wedge R(K) \wedge R(K+1) \wedge \ldots \quad (14)$$

The same operations (12) ~ (14) can be used for slices with local maxima of intra-slice connectivity. As a result, a set of images is generated, which forms separate segments of the input image A. Within the boundaries of these areas, the brightness changes smoothly.

The obtained segments $V_1, .., V_L$ can be used for further research on the connection of one with the other. This makes sense if the segmentation task includes selection of an object formed from several segments [11, 20]:

$$\Delta(V_l, V_{l+1}) = \sum_{n=1}^{M,N}{}_{,} \sum_{j=-1}^{1,1}{}_{,} v^l(m,n) \cap v^{l+1}(m+i,n+j) \quad (15)$$

where $\Delta(V_l, V_{l+1})$ – the function of cross-sectional connectivity between regions V_l and V_{l+1}.

At the same time, not only the contours of the studied areas are analyzed, but possible "holes" formed in the image from its individual elements are also considered. A software model of the image segmentation method based on features of spatial connectivity was also developed [16, 17].

2. Research results

In summary, we will describe the steps of the general segmentation algorithm, based on the described mathematical model:

1. Image quantization by brightness level.
2. Formation of the intra-slice connectivity function.
3. Determination of the global maximum of the intraslice connectivity function and its slice number.
4. Determination of the slice number with the global maximum connectivity value.
5. Formation of the function of inter-sectional connectivity.
6. Definition of the inter-slice connectivity function for the slice defined in point 4 and adjacent to it.
7. Comparison of the values of cross-sectional connectivity obtained in point 6 with the threshold.
8. Combining the selected sections.

This method can be applied to the segmentation of biomedical images. Below is an example of segmentation based on the method of analyzing connectivity histograms for thermal imaging in figure 2. As can be seen from this example, three main segments were selected (figure 3).

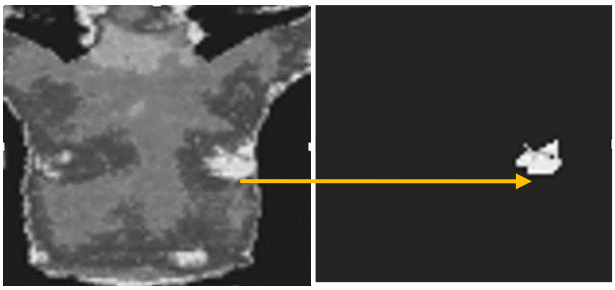


Fig. 2. Input thermal image

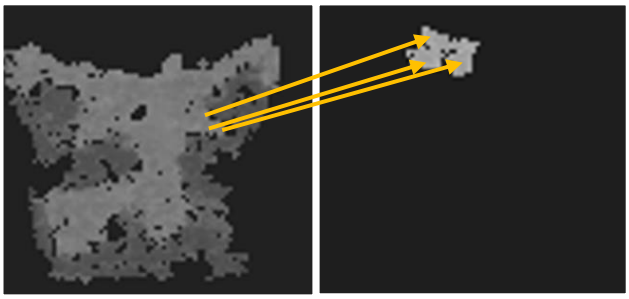


Fig. 3. Selected segments by the described method

Experimental studies included the processing of 500 thermal images. The generalized results of the comparative analysis – correlation comparison of the selected segments based on the proposed segmentation method with known methods, particularly the method of expanding areas and based on the recurrent method of brightness threshold limitation – are presented in table 1.

As a result of the research, a mathematical model is described, which allows the initial image in the form of a matrix A with dimensions $M \times N$ to be presented as a collection of several areas that correspond to individual segments of the object, their combination, or the entire object. They correspond to the original corresponding matrices $V_1, V_2, \ldots, V_L$ in halftone or binary form.

Table 1. Segmentation methods comparative analysis results

Segmentation method	Correlation coefficient of segmented areas
Recursive method of limiting brightness	0.85 – 0.90
Segmentation using region augmentation	0.90 – 0.97
The proposed segmentation method	0.95 – 0.99

On its basis, a step-by-step algorithm for segmentation of multigradation images based on spatial connectivity features is presented. The algorithm's result based on the thermal image sample segmentation is demonstrated. The segmentation method is compared with already existing approaches and its effectiveness is shown.

3. Conclusions

In this article, image segmentation was considered. A method of segmentation of multigradation images based on spatial connectivity features has been developed. The proposed method is characterized by a small number of calculations and ease of implementation.

The division of the image into parts can also be considered as a process of clustering image elements. In this case, the elements of the quantum image are displayed in the vector space of the features of inter- and intra-distinct connections. Clusters consist of various sets of feature space vectors. This division of space into clusters allows simpler methods to classify images. Thus, image division is created on a subset consisting of feature space vectors of inter-distinct and intra-distinct connectivity and sets of elements belonging to various classes.

On the one hand, the proposed approach to the segmentation of multigradation images is considered a method of dividing the image into a set of homogeneous areas. At the same time, the degree of homogeneity determines the change in the level of cross-sectional connectivity, which does not exceed a given threshold.

On the other hand, the proposed approach can be used as a simple method of clustering. At the same time, the formed features, such as intra-slice and inter-slice connectivity, can also be used to assess the degree of similarity and difference of the analyzed images.

References

[1] Avrunin O. G. et al.: Features of image segmentation of the upper respiratory tract for planning of rhinosurgical surgery. 2019 IEEE 39th International Conference on Electronics and Nanotechnology, ELNANO 2019, 485–488.
[2] Avrunin O. G. et al.: Research Active Posterior Rhinomanometry Tomography Method for Nasal Breathing Determining Violations. Sensors 21, 2021, 8508 [http://doi.org/10.3390/s21248508].
[3] Bradski G., Kaehler A.: Learning Open CV, second edition. 2013.
[4] Burgener F. et al.: Differential Diagnosis in Computed Tomography, 2011.
[5] Campbell J.: Human Medical Thermography, 2022.
[6] Comaniciu D., Meer P.: Mean shift analysis and applications. IEEE International Conference on Computer Vision 2, 1999, 1197.
[7] Comaniciu D., Meer P.: Mean Shift: A Robust Approach Toward Feature Space Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 603–619.
[8] Comaniciu D., Ramesh V., Meer P.: Real-Time Tracking of Non-Rigid Objects Using Mean Shift. Conference on CVPR 2, 2000, 1–8.
[9] Gonzalez R., Woods R: Digital Image Processing. Technosphere, 2012.
[10] Haralik R. M.: Statistical and structural approaches to the description of textures. Proceedings of the Institute of Electronics and Radio Engineering, 1979, 98–120.
[11] Kurmi Y., Chaurasia V.: Multifeature-based medical image segmentation. Sensors, 2018.
[12] Linda G. S. Stockman G. C.: Computer Vision, 2001.
[13] Orazayeva A. et al.: Biomedical image segmentation method based on contour preparation. Proc. SPIE 12476, 2022, 1247605 [http://doi.org/10.1117/12.2657929].
[14] Rodriguez-Lozano F. J., León-García F., Ruiz de Adana M., Palomares J. M., Olivares J.: Non-Invasive Forehead Segmentation in Thermographic Imaging. Sensors 19, 2019, 4096 [http://doi.org/10.3390/s19194096].
[15] Romanyuk O. N.: A function-based approach to real-time visualization using graphics processing units. Proc. SPIE 11581, 2020, 115810E [http://doi.org/10.1117/12.2580212].
[16] Rother C., Kolmogorov V., Blake Grabcut A.: Interactive foreground extraction using iterated graph cuts, 2004.
[17] Timchenko L. I. et al.: Q-processors for real-time image processing. Proc. SPIE 11581, 2020, 115810F [http://doi.org/10.1117/12.2580230].
[18] Timchenko L. I., Kutaev Y. F.: Method and organization of image extraction. Patent 2024939C1 RF, MKI G 06 K 9/00, 1992-07-08, 1992.
[19] Vapnik V.N., Chervonenkis A.Y.: Pattern recognition theory (statistical learning problems). Science, 1974.
[20] Wójcik W., Smolarz A.: Information Technology in Medical Diagnostics (1st ed.). CRC Press 2017 [http://doi.org/10.1201/9781315098050].

Prof. Leonid Timchenko
e-mail: tumchenko_li@gsuite.duit.edu.ua

Doctor of Technical Sciences, professor, Head of Artificial Intelligence Systems and Telecommunication Technologies Department, State University of Infrastructure and Technology, Ukraine. 55 articles published in Scopus, 215 citations in 106 articles (h-index = 8).
Research Interests: systems and means of artificial intelligence, parallel-hierarchical image processing systems.

http://orcid.org/0000-0001-5056-5913

Ph.D. Natalia Kokriatskaia
e-mail: kokryatska_ni@gsuite.duit.edu.ua

Associate professor of Artificial Intelligence Systems and Telecommunication Technologies Department, State University of Infrastructure and Technology, Ukraine. 34 articles published in Scopus, 109 citations in 77 articles (h-index = 6). Research Interests: parallel information processing, parallel-hierarchical image processing systems.

http://orcid.org/0000-0003-0090-3886

Ph.D. Volodymyr Tverdomed
e-mail: tverdomed@gsuite.duit.edu.ua

Associate professor, Director of Kyiv Institute of Railway Transport, State University of Infrastructure and Technology, Ukraine.
Research Interests: development of methods for diagnosing the technical condition and forecasting the duration of operational work of railway track elements and track devices.

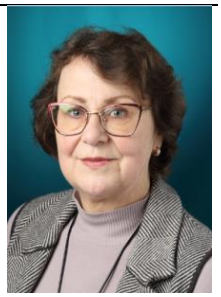
http://orcid.org/0000-0002-0695-1304

M.Sc. Oleksandr Stetsenko
e-mail: stetsenko_oo@gsuite.duit.edu.ua

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.

http://orcid.org/0000-0001-8359-0218

M.Sc. Valentyna Kaplun
e-mail: valentina.kaplun@vntu.edu.ua



Lecturer of the Chair of Safety of Infomation and Communication Systems, Vinnytsia National Technical University, Ukraine.
Research Interests: systems of artificial intelligence, information compression.

http://orcid.org/0000-0003-4353-3694

Ph.D. Eng. Oleg K. Kolesnytskyj
e-mail: kolesnytskiy@vntu.edu.ua

Associate professor, Department of Computer Sciences, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University. Field of scientific interests: information technologies, quality of electricity; improvement of the quality of electricity, management of devices for dynamic compensation of reactive power; increasing the reliability of distribution networks. He has published more than 100 scientific works.

http://orcid.org/0000-0003-0336-4910

Ph.D. Oleksandr O. Reshetnik
e-mail:Degratnik@gmail.com

Assistant at the Department of Software, Vinnytsia National Technical University. AWS Certified Solutions Architect – Professional. Global API and integrations discipline head in Epam Systems.
Research interests include: analog-to-digital conversion and digital-to-analog conversion with redundancy with non-binary notations; big data processing; systems integrations with middleware integration management systems; software architecture; cloud computing. Published more than 50 scientific papers.

http://orcid.org/0009-0006-7320-329X

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 confer ence proceedings. Member of Expert Group in the Computer Science specialization of IQAA.
Her professional interests are teaching, artificial intelligence, software engineering, data processing.

http://orcid.org/0000-0002-8411-3584

Ph.D. Ulzhalgas Zhunissova
e-mail: ulzhalgaszhunisova@gmail.com

Senior lecturer of the Department of Biostatistics, Bioinformatics and Information Technologies, Astana Medical University, Astana, Kazakhstan. She received her doctoral degree from the Lublin University of Technology in 2023,.
Area of scientific interests: data processing, statistics, machine learning methods, neural networks, Bayesian networks, evolutionary algorithms, clustering, information technology.

http://orcid.org/0000-0001-5255-9314

http://doi.org/10.35784/iapgos.5334 *received: 13.08.2023 | revised: 09.09.2023 | accepted: 17.09.2023 | available online: 30.09.2023*

IMPLEMENTATION OF COMPUTER PROCESSING OF RELAXATION PROCESSES INVESTIGATION DATA USING EXTENDED EXPONENTIAL FUNCTION

Andrey Lozovskyi[1], Alexander Lyashkov[2], Igor Gomilko[2], Alexander Tonkoshkur[1]
[1]Oles Honchar Dnipro National University, Department of Electronic Computing Machinery, Dnipro, Ukraine, [2]Oles Honchar Dnipro National University, Department of Applied Radiophysics, Electronics and Nanomaterials, Dnipro, Ukraine

Abstract. *The object of research is the development of a specialized measuring information system for the study and control of relaxation processes in materials and technical systems. The purpose of the work is the use of computer technologies to eliminate routine operations associated with the processing of experimental data, increase the speed, accuracy and information content of the process of studying the control of gas sensors. A variant of using computer data processing to automate the processing and primary analysis of experimental data of scientific research and control of the physicochemical parameters of gas-sensitive materials is proposed. The developed computer data processing system provides a practical opportunity to use the measurements of the kinetic characteristics of the gas sensitivity of gas sensors for their experimental research and control and, thus, to achieve higher accuracy and information content. The testing of the developed information-measuring system confirmed its operability and compliance with the requirements for improving the accuracy and speed of the processing process.*

Keywords: information-measuring system, gas sensor, extended exponential function, hardware, software, data processing

IMPLEMENTACJA KOMPUTEROWEGO PRZETWARZANIA DANYCH BADANIA PROCESÓW RELAKSACYJNYCH Z WYKORZYSTANIEM ROZSZERZONEJ FUNKCJI WYKŁADNICZEJ

Streszczenie. *Przedmiotem badań jest opracowanie specjalistycznego systemu informacji pomiarowej do badania i kontroli procesów relaksacyjnych w materiałach i systemach technicznych. Celem pracy jest wykorzystanie technologii komputerowych do wyeliminowania rutynowych operacji związanych z przetwarzaniem danych eksperymentalnych, zwiększenia szybkości, dokładności i zawartości informacyjnej procesu badania kontroli czujników gazu. Zaproponowano wariant wykorzystania komputerowego przetwarzania danych do automatyzacji przetwarzania i podstawowej analizy danych eksperymentalnych badań naukowych i kontroli parametrów fizykochemicznych materiałów wrażliwych na gaz. Opracowany komputerowy system przetwarzania danych zapewnia praktyczną możliwość wykorzystania pomiarów charakterystyk kinetycznych wrażliwości czujników gazu do ich badań eksperymentalnych i kontroli, a tym samym do osiągnięcia wyższej dokładności i zawartości informacyjnej. Testy opracowanego systemu pomiaru informacji potwierdziły jego funkcjonalność i zgodność z wymaganiami dotyczącymi poprawy dokładności i szybkości procesu przetwarzania.*

Słowa kluczowe: system informacyjno-pomiarowy, czujnik gazu, rozszerzona funkcja wykładnicza, sprzęt, oprogramowanie, przetwarzanie danych

Introduction

Computer technologies for processing experimental data have already become widespread in various fields of science and technology. They are used both in the study of well-known problems of high-energy physics and nuclear physics [3], image processing [4], search and creation of new materials [2], and for solving specific scientific and technical problems research of plasma light sources [8], construction of intelligent information-measuring systems for monitoring heat-insulating materials [9], etc.

Another specific area where the use of the technologies under consideration seems appropriate is the problem of studying relaxation processes in various materials and technical systems. The analysis of such kinetic dependences allows, as a rule, to obtain the most complete information about the properties of an object [7, 13, 15, 18, 19].

Studies of this type are associated with the need to process and analyze significant amounts of experimental data. The previous paper [12] shows the prospects for solving such a problem by creating a specialized information-measuring system.

The object of research is the implementation of computer processing and primary analysis of data from studies of relaxation processes in materials and systems using the example of resistive gas sensors. This process usually takes a lot of time and requires multiple cycles of processing large arrays of experimental data. First of all this is due to the need to ensure sufficient accuracy and information content of the results obtained. Therefore, to exclude routine operations associated with the processing of experimental data, to increase the speed, accuracy and information content of the research process itself, it seems relevant to automate it using computer technology.

The subject of the study is the methods and methods for constructing hardware and software for a specialized automated information-measuring system, which provided digitization and input of measurement data into memory and data analysis within the framework of the well-known model of a stretched exponential function [12, 14]. Known methods of computerization of the data processing process, as a rule, reduced to the use of their digitization and input into computer memory [13].

The purpose of the work is the use of computer technologies to eliminate routine operations associated with the processing of experimental data, increase the speed, accuracy and information content of the process of studying the control of gas sensors.

In this paper, the task is to develop a variant of automation of an information system for processing and analyzing measurement data of the response kinetics of resistive gas sensors for research and control of gas sensor parameters.

The most acceptable option for using computer data processing for the described measuring process is the development of an IMS, which should provide a solution to two problems:

- digitization and input of measurement data into the EOM memory in the form of text files;
- processing and analysis of data within the framework of the previously mentioned model of an extended exponential function, in particular, to ensure the determination of its parameters and the construction of their dependences on various factors (temperature, partial pressure of the detected gas, structural features and chemical composition of the sensor used, etc.).

1. Method and technology for measuring gas sensor parameters

The main characteristics of gas sensitive sensors are [1]:

- response kinetics when a certain amount of adsorbent enters/removes into the atmosphere surrounding the sensor (at a constant sensor temperature);
- temperature dependence of the response at a constant concentration of the adsorbent;
- dependence of the response on the concentration of adsorbent at a constant temperature.

The response defined as R_s/R_o, where R_s is the resistance of the sensor in the medium containing the adsorbent, R_o – is in air.

Based on the need to measure these characteristics, a measuring device was assemble.

The measuring setup consisted of a sealed chamber with a volume of 20 dm^3, to which a small expansion tank was attach, a power supply unit, and an electrical circuit. The expansion tank consisted of elastic rubber and was use to equalize the pressure between the volume of the chamber and the surrounding atmosphere. The electrical circuit ensured the reception of a signal from the sensor and its transmission to a personal computer (PC), and the computer itself.

The gas sensor, together with the system for ensuring its heating, was located in the measuring chamber and electrically connected to the rest of the installation. The chamber also housed a thermometer to control the temperature of the gases and a miniature fan to ensure the isotropy of the gaseous medium.

A dosed amount of active gas was inject at the right time into the measuring chamber.

The most well-known implementations of automation of measurements usually reduced to the process of digitizing the input experimental data and loading it in the form of text files into the computer's memory [13]. Their further processing and analysis were carried out manually or using separate computer programs.

The previous papers [12, 14] used a general phenomenological model based on the extended Kohlrausch-Williams-Wots exponential function $f_{kww}(t) = exp[-(t/\tau)^{\beta}]$, where t is time, β and τ are parameters [10, 17]. It is use to describe the kinetics of the relaxation dependences of the response, in particular, at the recovery stage.

A distinctive feature of such a model is its suitability for describing relaxation phenomena in disordered systems, in which dynamic processes occur simultaneously in many time ranges [6]. As shown in the paper [13] the use of such an approach makes it possible to obtain much more information on the physicochemical phenomena that determine the phenomenon of gas sensitivity of resistive sensor materials. Used in the window application given in the paper [12] it indicates the reality and efficiency of computer processing of a significant amount of data in the process of conducting the physical and chemical measurements under consideration.

2. Software and hardware for digitization and input of measurement data into memory

The hardware for inputting and converting data into digital form is based on the ARM microcontroller STM32F103C8T6 [16], which uses the AMS1117-3.3 chip to provide a stable supply voltage (+3.3 V). The functioning of the microcontroller is carry out using a quartz resonator X1 with a frequency of 8 MHz.

The analog voltage that was digitized came from a voltage divider, to which a constant voltage $U_C = 3.3$ V was applied. The divider was a series connection of a precision current resistor (R_x) and the sensor itself. A part of the voltage taken from the U_{SEN} sensor was impress to the corresponding output of the microcontroller, which is the input of its 12-bit built-in analog-to-digital converter (ADC). The use of this ADC allows digitization with an accuracy of at least 1 mV.

This connection scheme allows you to define the resistance of the sensor as the ratio of the digitized voltage U_{SEN} and the current I passing through the divider $R_{SEN} = U_{SEN}/I$, where $I = (U_C - U_{SEN})/R_X$.

Communication with a personal computer is carry out via the USB interface. For communication, only two common wires D+ and D- are used, which are connect to the micro-USB connector.

To control the electrical circuit of digitization and enter measurement data into the PC memory, a program created in the Python language and retranslated into the machine code language using the script compiler for Windows – Auto-PY-to-EXE is used. The architectural scheme of the developed program is show in Fig. 1.

The user interface of the program is show in Fig. 2. On Fig. 2 area, "1" is used to connect to a PC; in the field "2" – the measured data are displayed. Field "3" is used to set the delay time, the resistance of the current resistor and enable the display of the measured signal in the form of a graph; Field "4" – to set the identifier of the text file, which will be used to write data and comments to the file. Field "5" – chart display area. To start / stop the program, use the "Start / Stop" button.

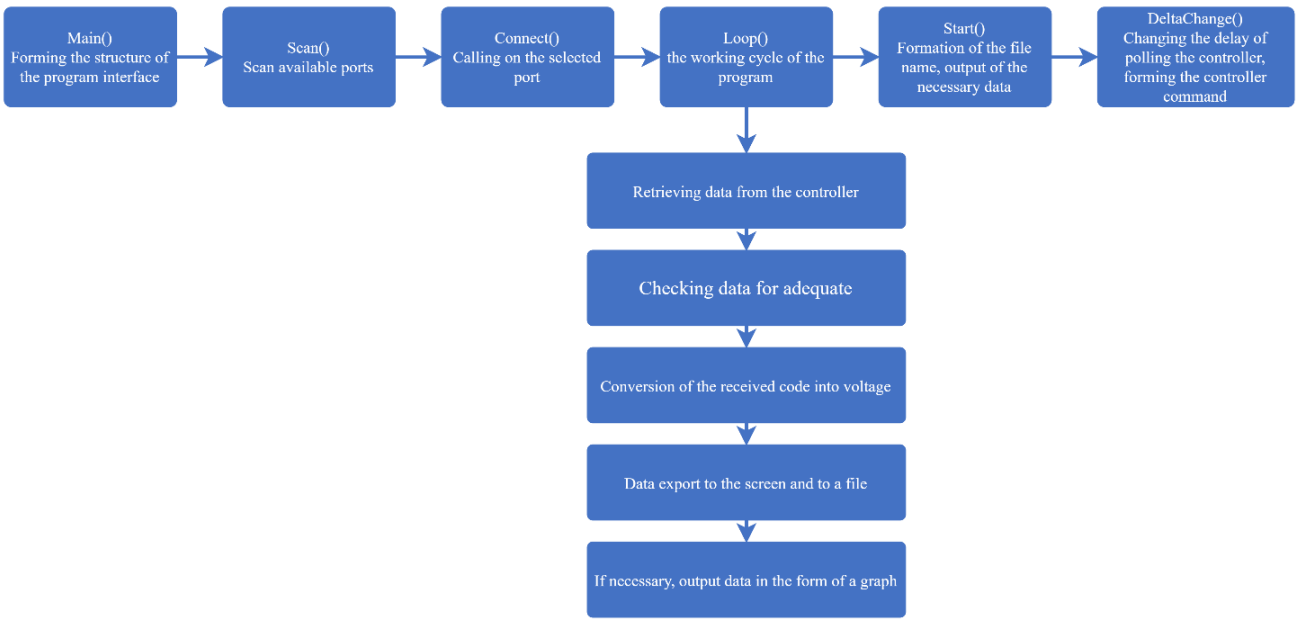


Fig. 1. Architectural scheme of the program for digitization and input of measurement data into memory

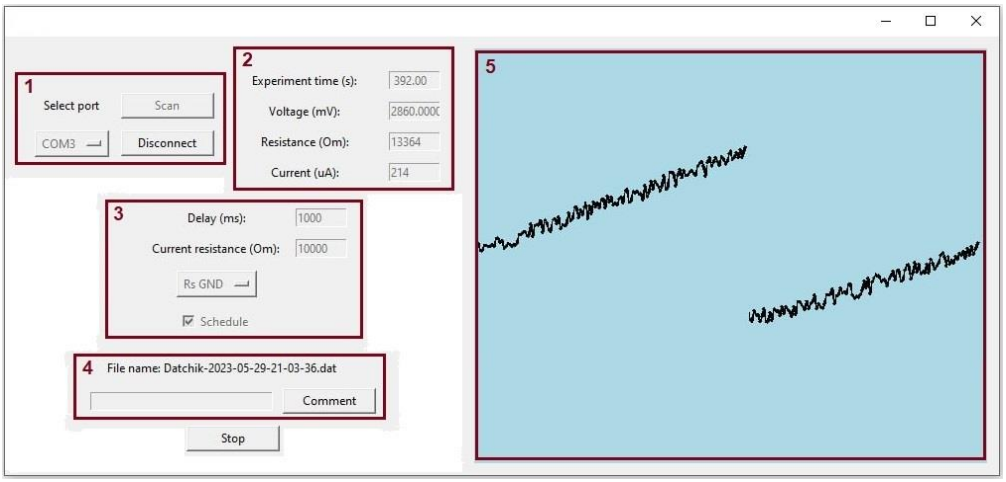


Fig. 2. General view of the interface of the data entry management program

3. Software for data processing and analysis

The presented system of computer technology for processing experimental data provides for the implementation of three modes – test, automated and in the mode with operator's support. The test mode designed to teach you how to work with the software. Automated mode operates as part of an automated information system for measuring and controlling parameters of gas sensors. In the mode with operator's support, the processing and analysis of experimental data is carried out step by step, allowing the user to make changes to its algorithm.

The software base on the integrated use of heterogeneous software products and provides the following service functions for processing and analyzing the specified data:

- data entry using Excel text files;
- calculations in the mathematical package "Mathcad" given parameters;
- visualization of the obtained dependencies in the form of graphs;
- providing the necessary textual information about the calculation algorithms used and test examples of their use;
- internet access if needed for more information.

The environment that ensures the operation of the described software product is the Windows operating system. Navigation between sections occurs with the help of buttons and menus of various levels (main, intermediate and working menus) [20].

To control the processing of source data used the Package Monitor, written using the C# language and the Windows Forms Application project type of the Visual Studio platform. The application software modules of the library of computational algorithms are documents of the well-known mathematical package Mathcad (version 15) [5]. In addition, the application uses universal tools for working with files with extensions like .xlsx, .dat(txt), docx and .pdf.

The main elements of the main window are buttons with labels, the names of which correspond to the purpose of the activated software modules (Fig. 3).

Pressing the "Entering the Experiment Date Argument" button calls up the intermediate menus "Task" (Fig. 4a), where the layout of the input and output information is indicated, and it is possible to select one of the proposed operating modes (Fig. 4b): test ("Test Program"), step-by-step with the participation and control of the operator ("My program") or automatic ("Auto Program"). After selecting and pressing the corresponding field, an Excel spreadsheet document file window will be loaded, where you should enter the values of the variable parameter-argument, the effect on the change in the KWW function of which is supposed to be investigated (Fig. 4c).

Its value is entered in the first column of the electronic table, zeros are entered in the remaining three fields, after which the file must be closed.

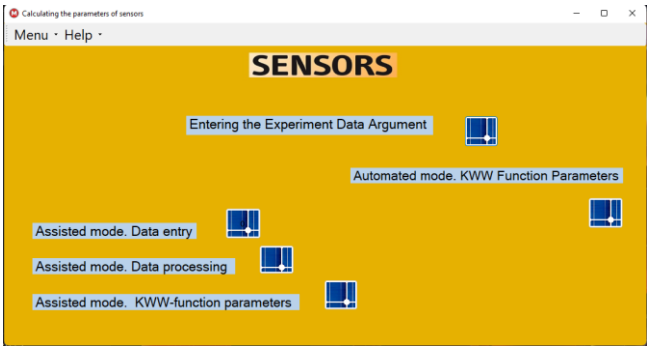


Fig. 3. Main window of the data processing and analysis interface

"**Automated mode. KWW Function Parameters**" run software in automated mode. It is assume that the experimental data of measurements of the kinetic relaxation dependence of the sensor response in the form of a text file (.dat) have already been placed in the corresponding software folder. The program of the mathematical package Mathcad in accordance with the algorithm described in the previous paper [11] will perform their processing. After that, the results will be transferred to a file with an .xlsx type extension, where the values of the variable parameter-argument were entered (instead of zeros, Fig. 4), as well as them – a graphical visualization of the values of the parameters of the KWW-function with the given values of the parameter-argument in the program itself processing.

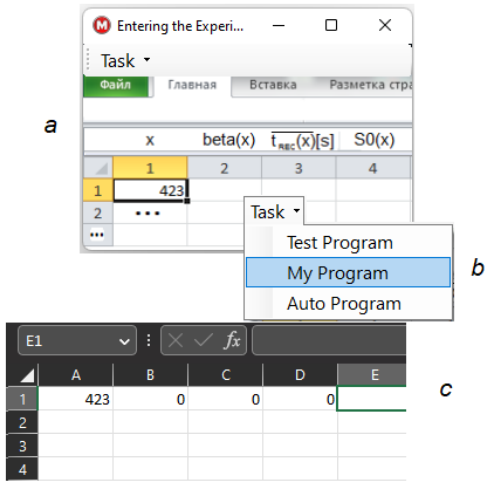


Fig. 4. Intermediate windows for choosing a mode (a and b) and switching to an Excel file (c) to set the value of the argument (the measurement identification factor, in this example, temperature)

Activation of the button "**Assisted mode. Data entry**" provides input of experimental data of measurements of the kinetic relaxation dependence of the sensor response into an Excel spreadsheet type file (or text file) for the test mode and the mode with operator support. In this case, the structure of intermediate windows is used, similar to that shown in Fig. 4.

The remaining buttons "**Assisted mode. Data processing**" and "**Assisted mode. KWW-function parameters**" (Fig. 3) are design to switch to intermediate windows of the same name, contain drop-down menus of the following three items that initialize the following:

"Assisted mode. Data processing" and "Assisted mode. KWW-function parameters" "Test" – call the window of the Mathcad package with a test case (does not allow any changes);

"My program" call of the window of the Mathcad package with a similar file used for processing the experimental data loaded for processing with the participation of the operator and allowing to make changes (modify the algorithm);

"Description" for opening files of the .pdf type, which contained information about the calculation algorithms used in the software modules of the Mathcad package.

The first of these algorithms provides a processing process with the participation of the operator of the data entered at the previous step by the operator. The purpose of the second algorithm is to fix and accumulate an array of values of the parameters of the Kohlrausch-Williams-Watts function, approximating the observed kinetics of changes in the resistance of the sensor under study, with subsequent visualization of their dependences on a given factor.

The main window also contains a toolbar with two buttons "Menu" and "Help". The Menu button contains two items. When you select the first of them, "Open Mathcad", Mathcad will be launch or a prompt will appear on the screen to select the Mathcad startup file yourself. When you select the second item "Exit", the application will be close. The Help button also contains a two-item menu. The "User's Manual" item opens a document with a user's manual. The "Workshop" item launches a browser to access the Internet.

It should be note that before using the described software product on a specific computer, it is necessary to install it together with the standard mathematical package Mathcad, Excel, a browser and .pdf file reader. The configuration of the product presented here is reduce to prescribing in the Visual Studio environment the corresponding file addresses in its monitor, which contain specific software implementations of computational algorithms from its library.

4. Algorithm for computer processing and data analysis

The maintenance algorithm provided by the IS for processing experimental data for measuring and controlling the parameters of gas sensors includes the following operations:

1. Preparation of the measuring setup for the specific conditions of the experimental study (placement of the sensor sample in the chamber, setting the physical parameters of the experiment: temperature, partial pressure, holding a variable interval to achieve a steady state, etc.).

2. Transfer (movement) of the sensor sample into an air atmosphere that does not contain active gas and launch the program for digitization and input of measurement data into the PC memory.

3. After generating a text file containing the experiment data, click the Entering the Experiment Date Argument button on the main menu of the data processing and analysis program. In the new menu, select and run the "**Auto Program**". In the window that opens, in the first column of a new row in the Excel table, load the value of the factor whose influence on the relaxation dependences of the resistance response of the sensor under research is the purpose of the research. The results are the temperature, the partial pressure of the active gas, or the physical parameter of the sensor under research (Fig. 4). In the remaining 3 fields of the columns (values of the parameters of the KWW function), enter "0" (they will be filled as a result of the processing process).

4. Activate and click on the button labeled "**Automated mode. KWW Function Parameters**" (launching the software in automated mode). In the window that opens in the form of a document of the mathematical package Mathcad, the results of processing presented in graphical and textual form (values of the parameters KWW of the function β, S_0 and τ corresponding to the accepted value of the studied influence factor and other experimental conditions).

5. For data processing and data analysis of the next measurement in accordance with paragraph 3, enter the new value of the factor, the influence of which is the purpose of the study, in the next line of the table (Fig. 4c). The file containing the specified.

5. Results of testing

The results of testing the IMS under consideration for computer processing of measurement data of gas sensor parameters in automatic mode are shown in Fig. 5. The effect of temperature on the kinetics of the response of conductivity to methane in ZnO-Ag ceramics was investigate.

The results presented on the presence of the temperature of the maximum sensitivity of the sensor (near 500 K), on the decrease in the characteristic time and the duration of the relaxation process, are in accordance with the available ideas and specific literature data [13].

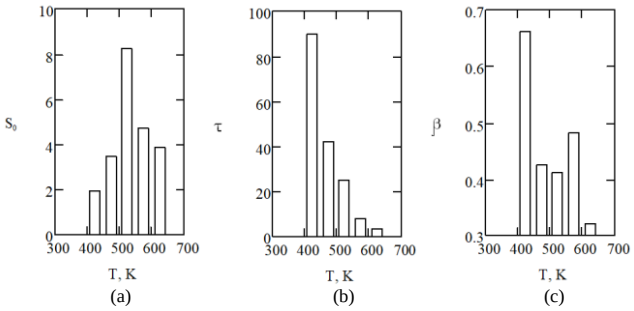


Fig. 5. Diagram of the influence of temperature (T, K) on the stationary (maximum) value of the response S_0 (a). The parameters that determine the relaxation process: the time characteristic τ, c (b) and the index of the extension of the duration of this process β (c) of a sample of zinc oxide ceramics with the addition of silver oxide (1.5 wt.%) after exposure to methane (relative concentration in air 0.005%

6. Conclusions

A variant of the implementation of a specialized information system for computer data processing in measuring and controlling the parameters of gas sensors based on experimental studies of the kinetic dependences of the response of gas-sensitive sensors, and their approximation by the well-known model of the stretched Kohlrausch-Williams-Watts function is present.

The system provides the following service functions for processing and analyzing the specified data:

- automated digitization and input of measurement data into the EOM memory in the form of text files directly from the measurement or control hardware;
- primary processing and analysis of data within the framework of the extended exponential function model, in particular, to determine its parameters and plot their dependences on various factors (temperature, partial pressure of the detected gas, structural features and chemical composition of the sensor used, etc.).

Data processing and analysis software is base on the integrated use of heterogeneous software products and provides the following service functions:

- data entry using Excel spreadsheets;
- calculations in the mathematical package "Mathcad" given parameters;
- visualization of the obtained dependencies in the form of graphs;
- providing the necessary textual information about the calculation algorithms used and test examples of their use;
- Internet access, if necessary, to obtain additional information.

Prospects for further developments and research are in the improvement of software and hardware. The presented results can be extend to studies of a wider class of relaxation processes that can be describe by means of an extended Kohlrausch-Williams-Watts exponential function.

References

[1] Fraden J.: Handbook of Modern Sensors. Springer Verlag, 2004.

[2] Himanen L., Geurts A., Foster A. S., Rinke P.: Data-driven materials science: status, challenges, and perspectives. Advanced Science 6(21), 2019, 1900808 [http://doi.org/10.1002/advs.201900808].

[3] Klimentiev A. A.: Methods for processing very large amounts of data in a distributed heterogeneous computer environment for applications in high energy physics and nuclear physics. Physics of elementary particles and the atomic nucleus 51(6), 2020, 1175–1303.

[4] Labunets V. G., Kokh E. V., Ostheimer E.: Algebraic models and methods of computer image processing. Part 1. Multiplet models of multichannel images. Computer Optics 42(1), 2018, 84–95 [http://doi.org/10.18287/2412-6179-2018-42-1-84-95].

[5] Makarov E. G.: Engineering calculations in Mathcad 15. Peter, Saint Petersburg 2011.

[6] Milovanov A. V., Rasmussen J. J., Rypdal K.: Stretched-exponential decay functions from a self-consistent model of dielectric relaxation. Phys. Lett. A. 372(13), 2008, 2148–2154 [http://doi.org/10.1016/j.physleta.2007.11.025].

[7] Niss K., Dyre J.C., Hecksher T.: Long-time structural relaxation of glass-forming liquids: Simple or stretched exponential? The Journal of Chemical Physics 152(4), 2020, 041103 [http://doi.org/10.1063/1.5142189].

[8] Pochinok A. V., Lazurik V. T., Tseluiko F. F., Borgun E. V.: Computer processing of the measurement results of the characteristics of the plasma ultraviolet source. Bulletin of the Kharkiv National University Physical series "Nuclei, particles, fields" 859, 2008, 59–64.

[9] Selivanova Z. M., Stasenko K. S.: Theoretical foundations for constructing intelligent information-measuring systems for tolerance control of thermal conductivity of heat-insulating materials: monograph. Publishing House of FGBOU VPO "TSTU", Tambov 2015.

[10] Simdyankin S. I., Mousseau N.: Relationship between dynamical heterogeneities and stretched exponential relaxation. Physical Review E. 68(4), 2003, 104–110 [http://doi.org/10.1103/PhysRevE.68.041110].

[11] Tonkoshkur A. S., Lozovskyi A. S.: Algorithm for processing gas sensor's response kinetics data using extended exponential function without numerical differentiation. System technologies 1 (144), 2023, 24–34 [http://doi.org/10.34185/1562-9945-1-144-2023-04].

[12] Tonkoshkur A. S., Lozovskyi A. S.: Application for calculating the parameters of a gas sensor from the experimental kinetic dependence of response. System technologies 2(133), 2021, 26–32 [http://doi.org/10.34185/1562-9945-2-133-2021-04].

[13] Tonkoshkur A. S., Lyashkov A. Y., Povzlo E. L.: Kinetics of Response of ZnO-Ag Ceramics for Resistive Gas Sensor to the Impact of Methane, and its Analysis Using a Stretched Exponential Function. Sensors and Actuators B: Chemical 255, 2018, 1680–1686 [http://doi.org/10.1016/j.snb.2017.08.171].

[14] Tonkoshkur O. S., Povzlo E. L.: Algorithm for data processing of response kinetics of a resistive gas sensor based on the stretched exponential function model. System technologies 1(108), 2017, 129–134.

[15] Tonkoshkur Y. A., Glot A. B.: Isothermal depolarization current spectroscopy of localized states in metal oxide varistors. Journal of Physics D: Applied Physics 45, 2012, 465305 [http://doi.org/10.1088/0022-3727/45/46/465305].

[16] Torgaev S. N., Musorov I. S., Chertikhina D. S., Trygub M. V.: Practical guide to programming STM-microcontrollers. Tomsk Polytechnic University, Tomsk 2015.

[17] Trzmiel J., Weron K., Janczura J., Placzek-Popko E.: Properties of the relaxation time distribution underlying the Kohlrausch-Williams-Watts photoionization of the DX centers in Cdl-xMnxTe mixed crystals. Journal of Physics Condensed Matter 21(34), 2009, 345801 [http://doi.org/10.1088/0953-8984/21/34/345801].

[18] Wang W. H.: Dynamic relaxations and relaxation-property relationships in metallic glasses. Progress in Materials Science 106, 2019, 10056 [http://doi.org/10.1016/j.pmatsci.2019.03.006].

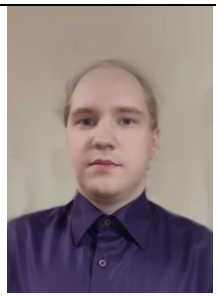
[19] Yuan Q. et al.: Aging Condition Assessment of XLPE Insulated Cables in Various Laying Environments Based on Isothermal Relaxation Current. IEEE 4th International Conference on Electrical Materials and Power Equipment (ICEMPE), 2023, 1–4.

[20] Programming in Windows Forms [https://metanit.com/sharp/windowsforms] (available: 10.06.2023).

M.Sc. Andrii Lozovskyi
e-mail: andrew.lozovsky@gmail.com

He is a postgraduate student of the Department of Electronic Computing Machinery of Oles Honchar Dnipro National University.
Since 2020, he has been working on a computer data processing system for measuring and controlling the parameters of gas sensors.

http://orcid.org/0009-0003-6674-0757



Ph.D. Alexander Lyashkov
e-mail: alexdnu@ ukr.net

Associate Professor of the Department of Applied Radiophysics, Electronics and Nanomaterials, Oles Honchar Dnipro National University. His research interests include semiconducting oxides, oxide ceramics for gas sensing applications and polymer composites.

http://orcid.org/0000-0001-5779-6001

Ph.D. Igor Gomilko
e-mail: gomilko@ffeks.dnu.edu.ua

Dean of the Faculty of Physics, Electronics and Computer Systems Oles Honchar Dnipro National University. His research interests include research of electronic processes in heterogeneous structures; automation of measurements using microprocessor technology; embedded development.

http://orcid.org/0000-0003-3256-9771

D.Sc. Alexander Tonkoshkur
e-mail: tonkoshkuras@ukr.net

Alexander S. Tonkoshkur is a professor of the Department of Electronic Computing Machinery of the Oles Gonchar Dnipro National University, Ukraine.
Education - Dnepropetrovsk State University, Faculty of Physics, specialty "radiophysics and electronics".
He received his DSc:Ukrainian from the Frantsevich Institute for Problems in Materials Science of the National Academy of Sciences of Ukraine in 1993. Scientific interests are in the field of electrical properties of functional electronics materials and the use of computer technologies for data processing and analysis in physical research.

http://orcid.org/0000-0002-1648-675X

http://doi.org/10.35784/iapgos.5350 *received: 11.08.2023 | revised: 09.09.2023 | accepted: 17.09.2023 | available online: 30.09.2023*

URBAN TRAFFIC CRASH ANALYSIS USING DEEP LEARNING TECHNIQUES

Mummaneni Sobhana[1], Nihitha Vemulapalli[1], Gnana Siva Sai Venkatesh Mendu[1], Naga Deepika Ginjupalli[1], Pragathi Dodda[1], Rayanoothala Bala Venkata Subramanyam[2]

[1]Velagapudi Ramakrishna Siddhartha Engineering College, Department of Computer Science and Engineering, Vijayawada, India, [2]National Institute of Technology Warangal, Department of CSE, Warangal, India

Abstract. *Road accidents are concerningly increasing in Andhra Pradesh. In 2021, Andhra Pradesh experienced a 20 percent upsurge in road accidents. The state's unfortunate position of being ranked eighth in terms of fatalities, with 8,946 lives lost in 22,311 traffic accidents, underscores the urgent nature of the problem. The significant financial impact on the victims and their families stresses the necessity for effective actions to reduce road accidents. This study proposes a framework that collects accident data from regions, namely Patamata, Penamaluru, Mylavaram, Krishnalanka, Ibrahimpatnam, and Gandhinagar in Vijayawada (India) from 2019 to 2021. The dataset comprises over 12,000 records of accident data. Deep learning techniques are applied to classify the severity of road accidents into Fatal, Grievous, and Severe Injuries. The classification procedure leverages advanced neural network models, including the Multilayer Perceptron, Long-Short Term Memory, Recurrent Neural Network, and Gated Recurrent Unit. These models are trained on the collected data to accurately predict the severity of road accidents. The project study to make important contributions for suggesting proactive measures and policies to reduce the severity and frequency of road accidents in Andhra Pradesh.*

Keywords: classification, gated recurrent unit, long-short term memory, multilayer perceptron, recurrent neural network, road accidents

ANALIZA KOLIZJI W RUCHU MIEJSKIM Z WYKORZYSTANIEM TECHNIK GŁĘBOKIEGO UCZENIA

Streszczenie. *Liczba wypadków drogowych w Andhra Pradesh niepokojąco rośnie. W 2021 r. stan Andhra Pradesh odnotował 20% wzrost liczby wypadków drogowych. Niefortunna pozycja stanu, który zajmuje ósme miejsce pod względem liczby ofiar śmiertelnych, z 8 946 ofiarami śmiertelnymi w 22 311 wypadkach drogowych, podkreśla pilny charakter problemu. Znaczący wymiar finansowy dla ofiar i ich rodziny podkreśla konieczność podjęcia skutecznych działań w celu ograniczenia liczby wypadków drogowych. W niniejszym badaniu zaproponowano system gromadzenia danych o wypadkach z regionów Patamata, Penamaluru, Mylavaram, Krishnalanka, Ibrahimpatnam i Gandhinagar w Vijayawada (India) w latach 2019–2021. Zbiór danych obejmuje ponad 12 000 rekordów danych o wypadkach. Techniki głębokiego uczenia są stosowane do klasyfikowania wagi wypadków drogowych na śmiertelne, poważne i ciężkie obrażenia. Procedura klasyfikacji wykorzystuje zaawansowane modele sieci neuronowych, w tym wielowarstwowy perceptron, pamięć długoterminową i krótkoterminową, rekurencyjną sieć neuronową i Gated Recurrent Unit. Modele te są trenowane na zebranych danych w celu dokładnego przewidywania wagi wypadków drogowych. Projekt ma wnieść istotny wkład w sugerowanie proaktywnych środków i polityk mających na celu zmniejszenie dotkliwości i częstotliwości wypadków drogowych w Andhra Pradesh.*

Słowa kluczowe: klasyfikacja, gated recurrent unit, pamięć długotrwała i krótkotrwała, perceptron wielowarstwowy, rekurencyjna sieć neuronowa, wypadki drogowe

Introduction

Road accidents on public roads involve vehicles, pedestrians, and cyclists. Factors like driving errors, mechanical malfunctions, weather conditions, and road infrastructure lead to accidents. Negligent behaviors such as speeding, distracted driving, and fatigue worsen the situation. Accidents result in injuries, fatalities, property damage, and emotional distress [5]. Addressing the issue requires strict regulations, speed enforcement, awareness campaigns, driver education, and improved infrastructure.

Prevention, awareness, and technology can reduce accidents, making roads safer for everyone. Various types of road accidents can occur. Rear-end collisions happen when one vehicle hits the back of another. Intersection accidents occur when vehicles collide while crossing paths at intersections. Single-vehicle accidents involve a vehicle colliding with a fixed object like a tree or a pole [29]. Head-on collisions happen when two vehicles collide front to front. When two vehicles crash side by side, the result is a side-impact collision, sometimes known as a T-bone accident. Rollover accidents involve a vehicle flipping over onto its side or roof. Pedestrian and cyclist accidents involve collisions with non-motorized road users.

Road accidents happen for different reasons. When drivers get distracted by using their phones or eating, they lose focus on the road [19]. Driving too fast because of impatience can make it hard to stay in control. If people drive after drinking alcohol or taking drugs, their judgment and coordination are affected. Reckless driving, like taking dangerous actions and ignoring traffic rules, makes accidents more likely [23]. Bad roads, bad weather, tired drivers, problems with vehicles, inexperienced drivers, and not following traffic laws can all cause accidents too. Road accident severity pertains to the extent of seriousness and impact of an accident, which encompasses injuries, fatalities, and property damage [6]. Factors that contribute to this severity include speed, collision type, road conditions, weather, and human behavior. Through the analysis of accident data, patterns and causes can be identified to develop targeted strategies, such as improving infrastructure, enforcing regulations, promoting responsible driving, and enhancing emergency response. The focus is on ensuring road safety and minimizing the occurrence of accidents, with efforts aimed at reducing severity, saving lives, and minimizing harm.

Controlling road accidents demands a comprehensive approach involving various strategies and stakeholders. One of the fundamental pillars is road safety education and awareness, disseminating knowledge about responsible driving practices and the potential consequences of reckless behavior through campaigns and outreach programs. Equally critical is strict enforcement of traffic laws, employing increased police presence and technology like speed cameras to deter dangerous driving habits. Driver education and training programs play a pivotal role in equipping motorists with essential skills and knowledge for safe driving. Continuous infrastructure improvements, including well-designed roads, clear signage, and pedestrian-friendly facilities, contribute to accident prevention. Advanced Driver Assistance Systems (ADAS) empowered by technologies like deep learning can assist drivers in avoiding collisions and maintaining safe driving behavior [18]. Adapting speed limits based on road conditions and implementing weather monitoring and warning systems add additional layers of safety. Involving the community in road safety initiatives fosters shared responsibility for safer roads, while international cooperation enables the exchange of best practices and innovative solutions. Integrating these measures, societies can make significant strides towards minimizing road accidents, protecting lives, and ensuring safer road environments for everyone.

1. Literature review

Mohammad Zarei et al. [27] proposed a non-parametric Empirical Bayes (EB) procedure using CGAN to model crash frequency data. The CGAN-EB method surpasses the conventional NB-EB method concerning hotspot identification and prediction. Compared to the negative binomial (NB) model, it offers superior accuracy for fitting accident data and making predictions. When it comes to network screening and crash modeling, CGAN-EB outperforms the standard NB-EB methodology by a significant margin. The study demonstrates the benefit of the approach through the analysis of datasets obtained from both real-life and hypothetical accident scenarios. It examines various aspects, including the model's fit to the data, the expected accuracy of predictions, and the outputs of network screening. This comprehensive evaluation sheds light on the advantages and effectiveness of the method.

Kaffash Charandabi et al. [14] analyzed a road accident risk mapping approach to reduce road accidents by identifying perilous areas. The proposed hybrid model combines a GRNN and a self-organizing map, considering 22 predictor attributes. An average of 90.74% in the testing on Iran's Tabriz Marand dual carriageway demonstrated a high level of accuracy in estimating accident probability. The study emphasized that areas prone to accidents were notably influenced by factors such as the presence of traffic surveillance cameras, the day of the week, driver age, local climate conditions, altitude variations, and even the specific type of vehicle traversing the area.

Zheng et al. [28] explored a hybrid deep-learning model for short-term traffic flow prediction in smart transportation systems. The framework merges a Bi-Directional Long Short-Term Memory (Bi-LSTM) model, which captures daily and weekly periodicities, with an attention-based Convolutional LSTM (Conv-LSTM) module that encodes spatial and short-term temporal data. Through empirical validation, their approach exhibited remarkable prediction performance surpassing existing methods, and underscoring the efficacy of integrating spatial-temporal features and an attention mechanism. The study not only advances traffic forecasting but also contributes to the foundation of intelligent transportation systems, hinting at the potential of their holistic approach in shaping the future of urban mobility.

Komol et al. [15] examined the effectiveness of advanced warning systems in detecting drivers' intended movements at crossings. Through in-depth analysis, they scrutinized predictors including vehicle speed, acceleration, and yaw rate, leveraging data from the Ipswich Connected Vehicle Pilot project across varying warning distances. Notably, the study demonstrated enhanced prediction accuracy by adopting a strategy of training distinct prediction models for each junction. This innovative approach not only refines prediction precision but also underscores the significance of context-sensitive models in bolstering the functionality of advanced driver assistance systems.

Dr. Rahman et al. [20] studied the significance of considering both the causes and effects of traffic accidents in road safety legislation. They examined purposeful and unintentional driving behaviors and their links to different types and severity levels of accidents using a BBN model. According to the study, which looked at the connections between driving habits and accidents in Al-Ahsa, the chance of occurrence of an accident is significantly increased by speeding alone. It doubles when brake failure is present. The complicated relationships between driving habits and accident causes are effectively captured by the BBN model.

Jie Yan et al. [25] proposed a novel approach for analyzing accident risk associated with highway geometric alignment using satellite maps and clustering algorithms. The method was validated using real-world data from the Nanfu highway in Chongqing, China. The study emphasized the significance of input variables such as section unit length, curve radius, and slope gradient. The MLP model outperformed the negative binomial model in predicting accident risk, revealing complex nonlinear relationships between variables. This research provides insights for enhancing traffic safety in mountainous regions.

Vani Suthamathi Saravanarajan et al. [22] proposed a study, that emphasizes the vital role of accident detection models in Autonomous Vehicles (AVs) safety. It highlights the need for robust deep learning methods to identify various accident scenarios, including isolated vehicle incidents. The study presents an innovative system using three deep learning models, achieving an 86.25% Accident Detection Rate (ADR) and a 33.00% False Alarm Rate (FAR). These results underscore the substantial potential of these models in enhancing AV safety and dependability.

Camilo Gutierrez-Osorio et al. [11] suggested an ensemble Deep Learning Model using GRU and CNN to predict traffic accidents. They leverage social media and open data sources to access detailed information, including unreported incidents. The model undergoes data quality and feature engineering processes. Results show that the ensemble model gave a better output than base models. The model's information can assist traffic control agencies in planning accident prevention activities, especially in high-traffic regions and intersections.

Dragan Gatarić et al. [9] to predict traffic accidents in Serbia and Bosnia & Herzegovina, two ANN models are suggested. These models forecast the number and severity of accidents using objective variables such as the length of the type of terrain, width of the road, volume of traffic, and speed threshold. The ANN models successfully predict outcomes and show good generalization skills. The study emphasizes how ANN might help with accident prediction and transportation planning. Accidents are found to be significantly influenced by section length, indicating that shortening it could lessen accidents and their severity.

Vishal Mandal et al. [17] proposed a novel approach for real-time traffic monitoring using CNNs and a GUI. This cutting-edge system employs advanced deep-learning algorithms and object tracking to identify lines, track stationary vehicles, and perform vehicle counting. Notably adaptable, it excels even in challenging environmental conditions, attesting to its robustness. The presented framework stands as a testament to its potential to revolutionize real-time traffic analysis, offering a reliable and comprehensive solution for varied scenarios.

Carlos M. Ferreira-Vanegas et al. [8] proposed a systematic literature review that investigates the utilization of statistical analysis (SA), machine learning (ML), smart city technologies (TESC), and Geographic Information Systems (GIS) for analyzing road traffic accidents (RTAs). Analyzing a dataset of 3,888 papers published between 2000 and 2021, the review identifies important papers, authors, and journals, as well as highlights emerging regression models and advanced technologies. Addressing the issue of unobserved heterogeneity and the significance of human factors in RTA analysis, the study emphasizes the value of computational algorithms, data visualization, and bibliometric tools for future research.

A. Comi et al. [7] proposed a system that utilizes data mining techniques to analyze road accident data in Rome Municipality. Through clustering and descriptive analysis, significant causes and patterns of accidents are identified, highlighting the influence of vehicle types and road infrastructure conditions on accident severity. The findings emphasize the potential of data mining in planning measures to reduce accidents and predict accident-prone areas. However, limitations are identified, suggesting the need for further research to investigate the impact of focusing on fatal accidents and exploring hybrid prediction approaches combining statistical and machine learning models.

K. Athiappan et al. [4] studied identifying accident-prone areas and the key factors causing accidents in the medium-sized Indian city of Tirunelveli. It employs geospatial analysis to pinpoint these "blackspots" and ranks the most influential factors based on recurrent accidents. The research provides crucial insights for transportation planners to mitigate casualties during

road construction and in existing conditions. The analysis identifies 30 accident-causing factors, with the top 10 including issues like volume-to-capacity ratio, lack of sight distance, uncontrolled intersections, and driver behaviors such as drunk driving and speeding. This comprehensive approach highlights the complexity of addressing accident causes in India, emphasizing the need for proactive measures to significantly reduce road accidents in the future.

Brunna de Sousa Pereira Amorim et al. [3] proposed study employs machine learning, analyzing attributes like spatial data, weather, and accident type. The study focused on Brazilian federal highways, using supervised algorithms. A neural network yielded the best results, with 83% accuracy, 84% precision, 83% recall, and 82% F1-score. Future work involves a smartphone app leveraging this model to alert drivers of high-risk highway sections, incorporating real-time external data for enhanced accuracy.

Piotr Gorzelanczyk et al. [10] studied that increasing vehicular traffic poses a significant and escalating threat of traffic accidents, resulting in loss of lives and economic resources. Limited dataset size challenges effective analysis. This study presents a methodology, utilizing multi-criteria optimization, to identify key factors influencing road accidents in Poland. Weather conditions, province, road type, and even weekdays were found to significantly impact accident rates. This versatile algorithm holds the potential for optimizing accident-reduction strategies, with plans for future research in accident forecasting and synergy-based solutions.

2. Proposed methodology

The proposed system architecture aims to ensure scalability, adaptability, and security. The Process Flow Diagram visually illustrates the sequential steps involved in the system's operation. The model consists of essential modules: Road accident data collection, Pre-processing Dataset, Training the deep learning models, classification of road accidents based on severity, and performance evaluation of the models. This provides valuable insights for analysis and decision-making. Fig. 1 shows the proposed methodology, emphasizing the importance of a well-defined approach in addressing road accidents. Incorporating advanced deep learning techniques and rigorous evaluation, the system aims to improve accuracy and efficiency in predicting the severity of road accidents. This contributes to suggesting effective measures for reducing the severity of road accidents.

2.1. Dataset description

The data utilized in this study has been gathered from the Vijayawada traffic police department and encompasses information from the years 2019, 2020, and 2021. It encompasses various regions within Vijayawada, namely Penamanaluru, Ibrahimpatnam, Gandhinagar, Krishnalanka, Patamata, and Mylavaram. The dataset consists of 13 features that provide valuable insights into road accidents, including Time, Location, P Vehicle, P Vehicle Age, Month, Day, S Vehicle, S Vehicle Age, Victim Gender, Road condition, Victim Age, and Alcohol Consumption. With a total of 12,712 rows, this dataset offers a substantial amount of data for analysis and evaluation. Each row represents a unique observation related to road accidents in the specified regions. These features and data points will serve as crucial inputs for the different modules and tasks involved in the study, facilitating comprehensive exploration and analysis of road accidents in the Vijayawada region. By leveraging this dataset, the study aims to uncover patterns, trends, and factors contributing to road accidents, ultimately assisting in developing effective measures to enhance road safety and reduce accidents in the region.

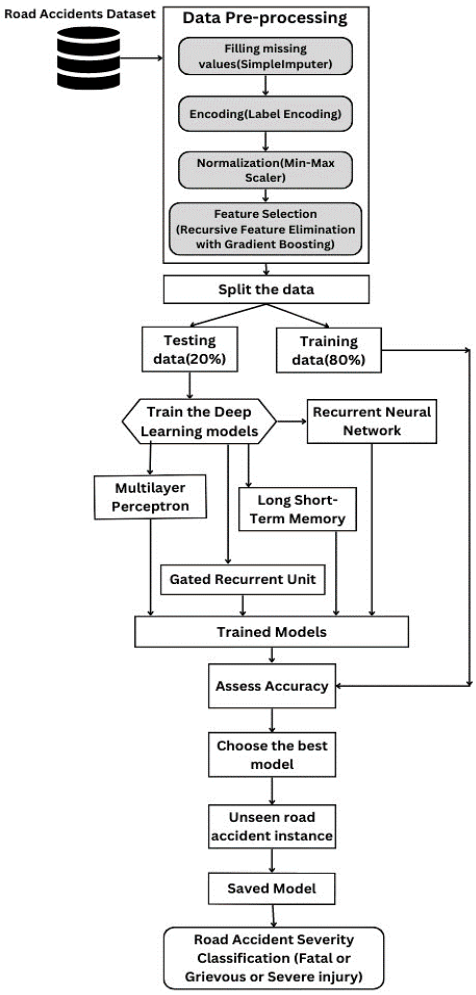


Fig. 1. Process Flow diagram of Road Accident Severity Classification model

2.2. Data pre-processing

Dealing with missing values

Dealing with missing values in a road accident dataset is crucial to maintaining data integrity and completeness for accurate analysis and modeling. A simple imputer algorithm can be used to address this issue. The algorithm identifies columns with missing values, finds the most common category in each column, and fills in the missing values accordingly. By replacing the original missing values with the most common category, the dataset becomes more reliable for further analysis. It is recommended to review the filled-in values to ensure their accuracy and representation of the missing data [2]. Employing a simple imputer algorithm ensures the reliability of the road accident dataset by filling in missing values and enabling accurate analysis and modeling.

Encoding the data

Encoding categorical data is a crucial step in data analysis as it enables numerical representation, algorithm compatibility, relationship identification, and improved model performance. The algorithm for encoding categorical data involves several steps. The algorithm identifies the columns in the dataset that contain categorical values. For each categorical column, a unique numerical label is assigned to each distinct category. A mapping or dictionary is created to keep track of the assigned labels for each category. The categorical values in each column are replaced with their corresponding numerical labels [13].

The analysis can then proceed using the encoded numerical values. The mapping can be retained for future reference or decoding purposes. The process of encoding categorical data facilitates efficient data analysis and enhances the performance of models by transforming categorical values into numerical representations.

Data normalization

Normalization is a crucial step in data analysis as it ensures that all features in a dataset are on a similar scale, preventing the dominance of certain features and enabling fair comparisons. The algorithm for normalization involves several steps. The algorithm identifies the numerical features in the road accident dataset. An instance of the Min-MaxScaler is created [12]. The scaler is fitted to the numerical features, calculating the minimum and maximum values. The numerical features are transformed to scale the values between 0 and 1. The original numerical features are then replaced with the normalized values. This allows for the continuation of the analysis using the normalized data. By normalizing the dataset, the algorithm ensures that all features are treated equally, facilitating unbiased analysis and comparisons.

Feature selection

Feature selection is an essential step in data analysis, and it can be performed using Recursive Feature Elimination (RFE) with Gradient Boosting. This method involves training a gradient-boosting model on the full feature set, ranking the importance of each feature, and repeatedly eliminating the least important features till the desired number of features is obtained. The purpose of this process is to optimize model performance and interpretability. The algorithm takes a normalized dataset as input and outputs a list of selected features. The steps involve loading the dataset, determining the desired number of features, training a gradient boosting model, evaluating feature importance, and repeating the process until the desired number of features is obtained [24]. The selected features are used to build the final model. By applying RFE with Gradient Boosting, the algorithm identifies the most relevant features, enhancing model performance and facilitating better interpretation of the results.

2.3. Classification of road accidents severity

Multi-layer perceptron

The classification of road accident severity is addressed using MLP. The MLP consists of interconnected neurons that process accident-related features to predict severity levels. Training the MLP requires a labeled dataset, divided into training and testing sets. The MLP learns patterns and relationships during training, optimizing its weights and biases. The trained MLP is evaluated using the testing set. Accurate severity classification aids in identifying high-risk situations, enabling targeted interventions for accident prevention [1].

Recurrent neural network

Classification of road accident severity involves utilizing RNNs to process accident data as a sequential time series. RNNs capture temporal relationships and patterns, making predictions based on accident-related features [21]. The parameters of the model are optimized using the training set, and performance is evaluated using the testing set. By leveraging the temporal dynamics and long-term dependencies in the data, RNNs improve the accuracy of severity predictions. This enables timely interventions and targeted safety measures to mitigate the impact of road accidents.

Long-short term memory

The classification of road accident severity can be effectively performed using LSTM networks, a type of RNN architecture. LSTMs capture long-term dependencies in sequential accident data, extract relevant features, and make accurate severity predictions. Proper data preprocessing, architecture design, and hyperparameter tuning are critical for optimal accuracy [16].

Gated recurrent unit

Classification of road accident severity using GRU is an effective approach for modeling temporal dependencies in accident data. GRUs, a variant of RNN, utilize gating mechanisms to selectively remember or forget information from previous time steps [26]. By processing accident data through GRU layers, the model captures sequential patterns and makes accurate predictions. Training involves a labeled dataset, split into training and testing sets. Proper data preprocessing, architecture design and hyperparameter tuning are crucial for optimal classification performance.

2.4. Training the deep learning models

Train deep learning models (MLP, RNN, LSTM, GRU) on labeled road accident data to classify accident severity, leveraging their respective architectures for capturing patterns and temporal dependencies. Algorithm 1 describes the training of deep learning algorithms on the Road accident dataset.

Algorithm 1 ROAD ACCIDENT SEVERITY CLASSIFICATION

Input: Pre-processed Road accident dataset

Output: Classification of road accident severity

1: Split the dataset into training and testing sets using an 80-20 ratio
2: Train an MLP model on the training data
3: Train an RNN model on the training data
4: Train an LSTM model on the training data
5: Train a GRU model on the training data
6: Evaluate the accuracy of each model using the testing data
7: Select the model with the highest accuracy for predicting the severity of road accidents

2.5. Performance evaluation

Evaluate the performance of deep learning models (MLP, RNN, LSTM, GRU) for road accident severity classification using metrics like accuracy, precision, recall, and F1-score to assess their effectiveness in predicting accident severity. Algorithm 2 describes the performance evaluation of deep learning models.

Algorithm 2 PERFORMANCE EVALUATION OF Deep Learning Models

Input: Actual labels y_{true}, Predicted labels y_{pred}

Output: Confusion matrix, Accuracy, Precision, Recall, F-score, ROC-AUC

1: $C \leftarrow confusion\_matrix(y_{true}, y_{pred})$
2: $Accuracy \leftarrow \frac{True\ Positives + True\ Negatives}{Total\ Samples}$
3: $Precision \leftarrow \frac{True\ Positives}{Total\ Samples}$
4: $Recall \leftarrow \frac{True\ Positives}{True\ Positives + False\ Positives}$
5: $F-score \leftarrow 2 \times \frac{Precision \times Recall}{Precision + Recall}$
6: Calculate ROC-AUC Score.

3. Results and discussion

3.1. Confusion matrix and ROC-AUC evaluation for deep learning models in predicting road accident severity

This study rigorously assesses four deep learning models: MLP, RNN, LSTM, and GRU, for their predictive capabilities in road accident severity prediction.

Confusion matrix analysis

For each model, confusion matrices are utilized to visually represent classification accuracy. These matrices provide an exhaustive breakdown of metrics such as True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). The presentation includes clear titles, labeled axes, color-coded cells, and informative annotations, facilitating a precise understanding of each model's effectiveness in accident severity categorization. This approach enables a comprehensive scrutiny of model performance nuances. Fig. 2, Fig. 3, Fig. 4, and Fig. 5 respectively show the confusion matrices for MLP, RNN, LSTM, and GRU.

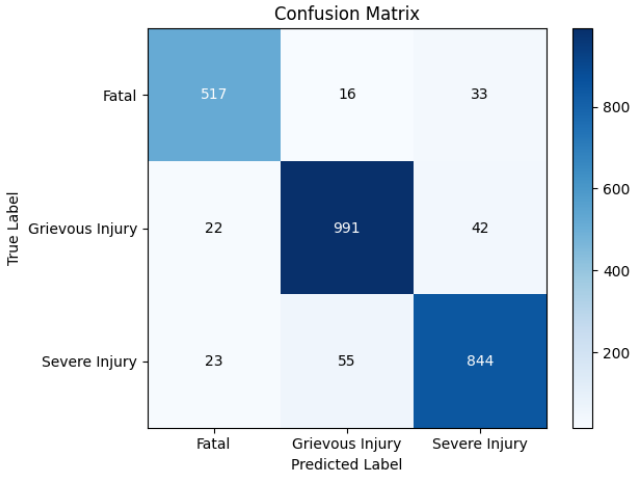


Fig. 2. Confusion Matrix of MLP

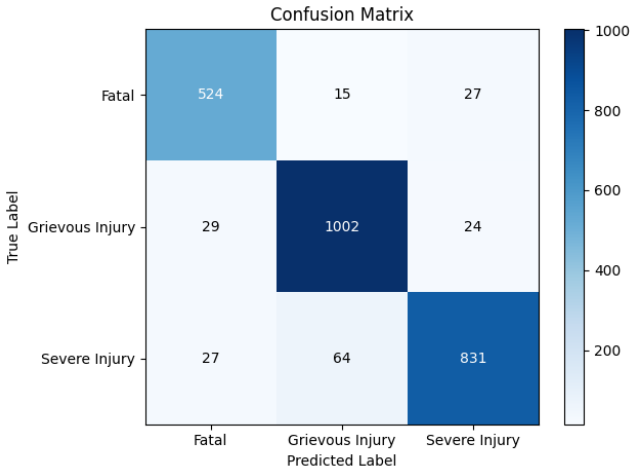


Fig. 3. Confusion Matrix of RNN

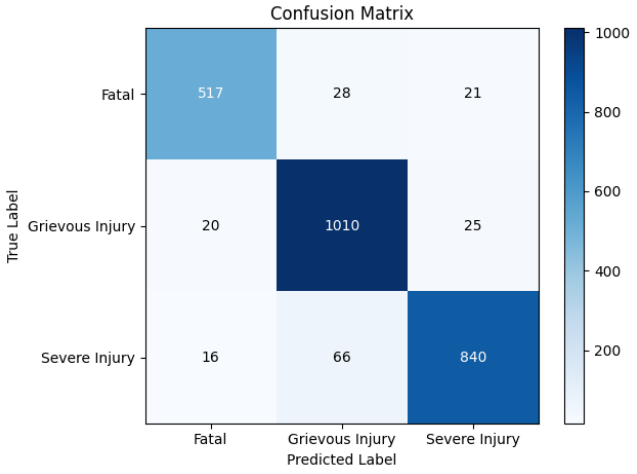


Fig. 4. Confusion Matrix of LSTM

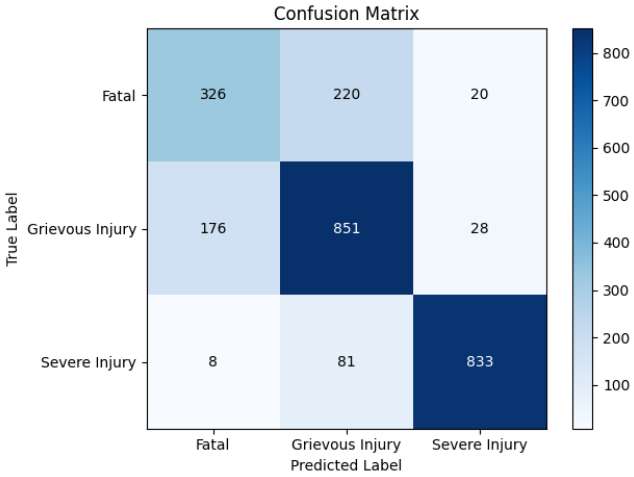


Fig. 5. Confusion Matrix of GRU

ROC-AUC curve assessment

In addition to the confusion matrix analysis, Receiver Operating Characteristic (ROC) curves and corresponding Area Under the Curve (AUC) scores are employed. These metrics offer nuanced insights into the models' discriminatory capabilities. ROC curves plot True Positive Rate (TPR) against False Positive Rate (FPR) at varying classification thresholds, with higher AUC scores indicating superior discriminatory prowess. Fig. 6, 7, 8, and 9 respectively show the ROC-AUC curves for MLP, RNN, LSTM, and GRU.

3.2. Evaluation metrics of MLP, RNN, LSTM, and GRU

This study explores the effectiveness of various deep learning models, including Multilayer Perceptron (MLP), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU), in predicting road accident severity. The models' performance is evaluated using essential metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. The objective is to assess these models' capability to accurately classify the severity of road accidents, providing valuable insights into their suitability for such tasks. Table 1 presents the comprehensive evaluation of these deep learning models.

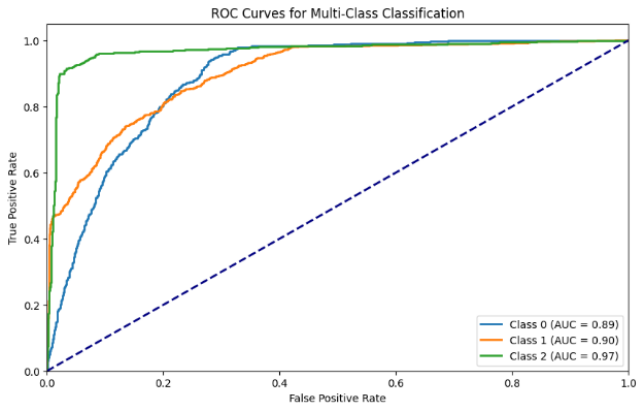


Fig. 6. ROC-AUC of MLP

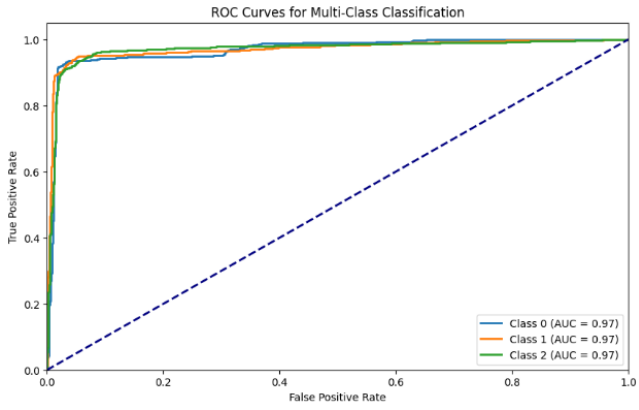


Fig. 7. ROC-AUC of RNN

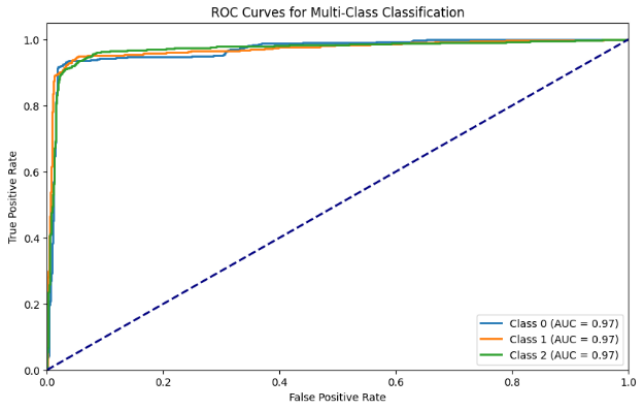


Fig. 8. ROC-AUC of LSTM

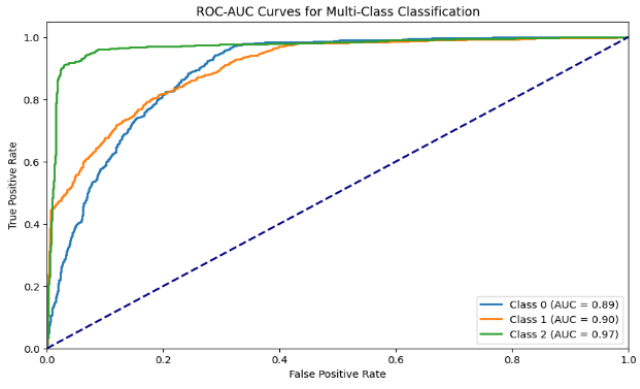


Fig. 9. ROC-AUC of GRU

Table 1. Evaluation metrics of MLP, RNN, LSTM, and GRU

Metrics	MLP	RNN	LSTM	GRU
Precision (%)	92.38	92.41	93.26	77.44
Accuracy (%)	92.48	92.68	93.07	79.04
Recall (%)	92.27	92.56	92.72	76.20
F1-Score (%)	92.32	92.46	92.96	76.70
ROC-AUC (%)	97.67	97.03	97.66	92.14

Among the evaluated deep learning models, LSTM emerges as the most accurate model, achieving an impressive accuracy of 93.07 percent. It is closely followed by the RNN model, which attains an accuracy of 92.68 percent. The MLP model demonstrates an accuracy of 92.48 percent, while the GRU model achieves 79.04 percent accuracy. Fig. 10 visually depicts the accuracy scores obtained by MLP, RNN, LSTM, and GRU, underscoring LSTM's superior performance in accurately classifying road accident severity. These results affirm LSTM as the most promising and effective model for precisely predicting and classifying road accident severity.

3.3. Accuracies of MLP, RNN, LSTM, and GRU for the classification of road accident severity

This study provides valuable insights into the potential of deep learning models, particularly LSTM, for enhancing the prediction and classification of road accident severity. LSTM's remarkable accuracy underscores its suitability for real-world applications in traffic safety and accident management.

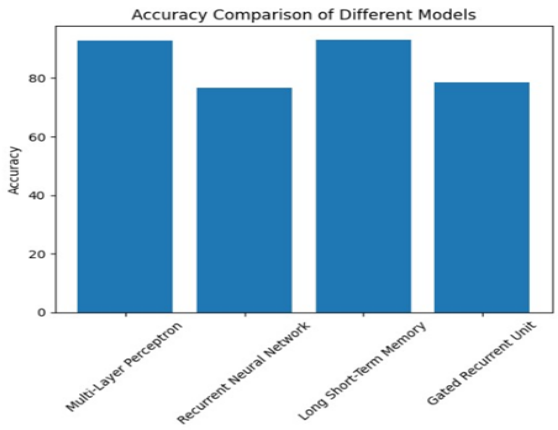


Fig. 10. Accuracies of MLP, RNN, LSTM, and GRU for the Classification of The Severity of Road Accidents

3.4. User interface

The user interface receives various inputs, including weather conditions, details of the primary and secondary vehicles, road conditions, the gender of the victim, the month and day of the incident, alcohol consumption, time, and the ages of the primary and secondary vehicles, as well as the victim's age. It then classifies the severity of the accident into one of three categories: fatal, severe, or grievous. To exemplify the model's performance, we present a specific use case: A scenario where the input parameters correspond to 'hot' weather conditions, a 'three-wheeler' as the primary vehicle, a 'pedestrian' as the secondary vehicle, 'dry' road conditions, a 'female' victim, an incident occurring in the 'month of January' on a 'Saturday' at '8:45 PM,' alcohol consumption, both the primary and secondary vehicles aged '3 years,' and the victim's age noted as '50 years.' The model, when presented with these input values, consistently predicts the accident's severity as 'Fatal' with a high degree of confidence. Consequently, if an actual accident instance mirrors these specific parameter values, our system unequivocally classifies it as 'Fatal.' For enhanced understanding and visualization of this classification process, we incorporate graphical representations into the user interface. Fig. 11 and Fig. 12 visually depict the UI effectively categorizing the road accident instance as 'Fatal.'

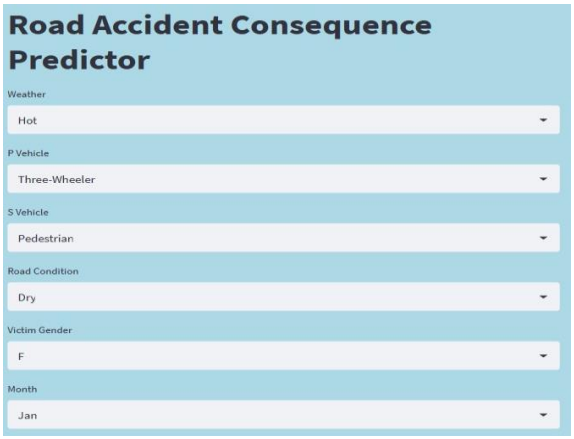


Fig. 11. Prediction of road accident instance as Fatal

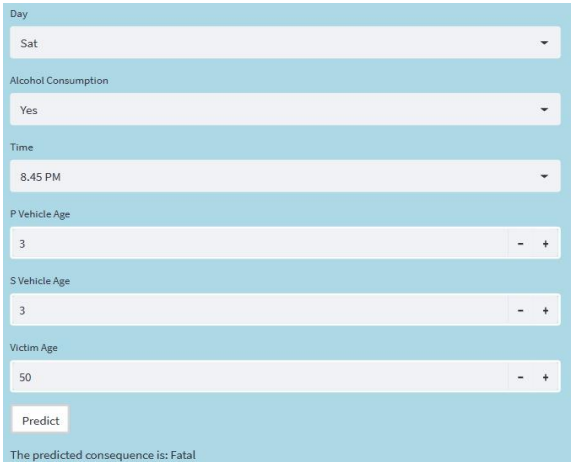


Fig. 12. Prediction of road accident instance as Fatal

4. Conclusion

The scope of this work is confined to the Vijayawada region, a city with its own unique set of road accident challenges. Focusing on this specific geographical area, the study aims to provide targeted insights and solutions to address the severity of the road accident. Analyzing road accident data and utilizing deep learning algorithms, the system aims to categorize the seriousness of road accidents into three distinct classes: Fatal, Grievous, and Severe injuries. This classification model serves as a crucial tool in understanding the magnitude of accidents and enables stakeholders to take proactive measures in terms of policymaking, infrastructure improvements, and awareness campaigns. To accomplish the objectives of the study, future work entails refining the models used for classification. This involves fine-tuning the algorithms, optimizing their performance, and enhancing their accuracy.

References

[1] Al Bataineh A., Kaur D., Jalali S. M. J.: Multi-layer perceptron training optimization using nature-inspired computing. IEEE Access 10, 2022, 36963–36977.
[2] Alghamdi T.A., Javaid N.: A survey of preprocessing methods used for analysis of big data originated from smart grids. IEEE Access 10, 2022, 29149–29171.
[3] Amorim B. d. S.P., et al.: A Machine Learning Approach for Classifying Road Accident Hotspots. ISPRS International Journal of Geo-Information 12(6), 2023, 227.
[4] Athiappan K., et al.: Identifying Influencing Factors of Road Accidents in Emerging Road Accident Blackspots. Advances in Civil Engineering, 2022.
[5] Cai Q.: Cause analysis of traffic accidents on urban roads based on an improved association rule mining algorithm. IEEE Access 8, 2020, 75607–75615.
[6] Chen M.-M., Chen M.-Ch.: Modeling road accident severity with comparisons of logistic regression, decision tree, and random forest. Information 11(5), 2020, 270.
[7] Comi A., Polimeni A., Balsamo Ch.: Road accident analysis with data mining approach: evidence from Rome. Transportation research procedia 62, 2022, 798–805.

[8] Ferreira-Vanegas C. M., Vélez J. I., García-Llinás G. A.: Analytical methods and determinants of frequency and severity of road accidents: a 20-year systematic literature review. Journal of Advanced Transportation, 2022.
[9] Gatarić D., et al.: Predicting Road Traffic Accidents - Artificial Neural Network Approach. Algorithms 16(5), 2023, 257.
[10] Gorzelanczyk P., Tylicki H.: Methodology for Optimizing Factors Affecting Road Accidents in Poland. Forecasting 5(1), 2023, 336–350.
[11] Gutierrez-Osorio C., González F. A., Pedraza C. A.: Deep Learning Ensemble Model for the Prediction of Traffic Accidents Using Social Media Data. Computers 11(9), 2022, 126.
[12] Islam M. J., et al.: Application of min-max normalization on subject-invariant EMG pattern recognition. IEEE Transactions on Instrumentation and Measurement 71, 2022, 1–12.
[13] Jia B.-B., Zhang M.-L.: Multi-dimensional classification via decomposed label encoding. IEEE Transactions on Knowledge and Data Engineering, 2021.
[14] Kaffash Charandabi N., Gholami A., Abdollahzadeh Bina A.: Road accident risk prediction using generalized regression neural network optimized with self-organizing map. Neural Computing and Applications 34(11), 2022, 8511–8524.
[15] Komol, M.M.R., et al.: Deep RNN Based Prediction of Driver's Intended Movements at Intersection Using Cooperative Awareness Messages. IEEE Transactions on Intelligent Transportation Systems 24(7), 2023, 6902–6921.
[16] Le X.-H., et al.: Application of long short-term memory (LSTM) neural network for flood forecasting. Water 11(7), 2019, 1387.
[17] Mandal V., et al.: Artificial intelligence-enabled traffic monitoring system. Sustainability 12(21), 2020, 9177.
[18] Novikov A., Shevtsova A., Vasilieva V.: Development of an approach to reduce the number of accidents caused by drivers. Transportation research procedia 50, 2020, 491–498.
[19] Östh J., et al.: Driver kinematic and muscle responses in braking events with standard and reversible pre-tensioned restraints: validation data for human models. SAE Technical Paper, 2013, 2013-22-0001.
[20] Rahman M.M., et al.: Towards sustainable road safety in Saudi Arabia: Exploring traffic accident causes associated with driving behavior using a Bayesian belief network. Sustainability 14(10), 2022, 6315.
[21] Rezk N. M., et al.: Recurrent neural networks: An embedded computing perspective. IEEE Access 8, 2020, 57967–57996.
[22] Saravanarajan V.S., et al.: Car crash detection using ensemble deep learning. Multimedia Tools and Applications, 2023, 1–19.
[23] Sobhana M., et al.: A Hybrid Machine Learning Approach for Performing Predictive Analytics on Road Accidents. 6th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), 2022.
[24] Upadhyay D., et al.: Intrusion detection in SCADA based power grids: Recursive feature elimination model with majority vote ensemble algorithm. IEEE Transactions on Network Science and Engineering 8(3), 2021, 2559–2574.
[25] Yan J., et al.: Relationship between Highway Geometric Characteristics and Accident Risk: A Multilayer Perceptron Model (MLP) Approach. Sustainability 15(3), 2023, 1893.
[26] Yin Y., et al.: SE-GRU: Structure Embedded Gated Recurrent Unit Neural Networks for Temporal Link Prediction. IEEE Transactions on Network Science and Engineering 9(4), 2022, 2495–2509.
[27] Zarei M., Hellinga B., Izadpanah P.: CGAN-EB: A non-parametric empirical Bayes method for crash frequency modeling using conditional generative adversarial networks as safety performance functions. International Journal of Transportation Science and Technology 12(3), 2023, 753–764.
[28] Zheng H., et al.: A hybrid deep learning model with attention-based conv-LSTM networks for short-term traffic flow prediction. IEEE Transactions on Intelligent Transportation Systems 22(11), 2020, 6910–6920.
[29] Road Accidents in Malaysia: Top 10 Causes & Prevention. Kurnia, 21 Sept. 2022 [http://www.kurnia.com/blog/road-accidents-causes].

Prof. Mummaneni Sobhana
e-mail: sobhana@vrsiddhartha.ac.in

Dr. M. Sobhana 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.

http://orcid.org/0000-0001-5938-5740

Eng. Nihitha Vemulapalli
e-mail: nihithavemulapalli@gmail.com

Nihitha Vemulapalli is a final-year B. Tech student specializing in Computer Science and Engineering at V. R. Siddhartha Engineering College, Vijayawada, India. She is passionate about Deep Learning and has been recognized as a Google Women Engineers Scholar. Additionally, she holds the position of IEEE Chair at the WIE (Women in Engineering) student chapter.

http://orcid.org/0009-0007-8626-0012

Eng. Gnana Siva Sai Venkatesh Mendu
e-mail: sivasaivenkatesh.m@gmail.com

Gnana Siva Sai Venkatesh Mendu is a final-year B. Tech. student, specializing in Computer Science and Engineering at V. R. Siddhartha Engineering College, Vijayawada, India. He is passionate about Artificial Intelligence and Machine Learning. He achieved Bronze in the Cisco NetAcad Riders 2023 competition. He holds the position of IEEE Chair of the GRSS Student Chapter.

http://orcid.org/0009-0004-6406-000X

Eng. Naga Deepika Ginjupalli
e-mail: nagadeepikaginjupalli@gmail.com

Ginjupalli Naga Deepika is currently pursuing her final year of graduation in computer science and engineering from V. R. Siddhartha Engineering College, Vijayawada, India. She is interested in doing projects in the machine learning domain. She holds the position of webmaster of the Computer Society student chapter.

http://orcid.org/0009-0005-0436-0008

Eng. Pragathi Dodda
e-mail: pragathidodda@gmail.com

Pragathi Dodda is currently pursuing B. Tech in computer science and engineering at V. R. Siddhartha Engineering College, Vijayawada, India. Her interests are in Artificial Intelligence and machine learning applied to detecting individuals with Lupus symptoms.

http://orcid.org/0009-0003-4879-6298

Prof. Rayanoothala Bala Venkata Subramanyam
e-mail: rbvs66@nitw.ac.in

R.B.V. Subramanyam received his M. Tech. and Ph.D. from the Indian Institute of Technology Kharagpur, India. Currently, He is working at the National Institute of Technology Warangal, India. He has published many journal and conference papers in the areas of Data Mining. Some of his research interests include Data Mining, Distributed Data Mining, Fuzzy Data Mining, and Big Data Analytics. He is one of the reviewers for IEEE Transactions on Fuzzy Systems and the Journal of Information and Knowledge Management. He is a member in IEEE and The Institution of Engineers (India).

http://orcid.org/0009-0005-8907-1984

http://doi.org/10.35784/iapgos.3740 *received: 13.06.2023 | revised: 20.08.2023 | accepted: 25.09.2023 | available online: 30.09.2023*

UNBALANCED MULTICLASS CLASSIFICATION WITH ADAPTIVE SYNTHETIC MULTINOMIAL NAIVE BAYES APPROACH

Fatkhurokhman Fauzi, Ismatullah, Indah Manfaati Nur
Universitas Muhammadiyah Semarang, Department of Statistics, Semarang, Indonesia

Abstract. *Opinions related to rising fuel prices need to be seen and analysed. Public opinion is closely related to public policy in Indonesia in the future. Twitter is one of the media that people use to convey their opinions. This study uses sentiment analysis to look at this phenomenon. Sentiment is divided into three categories: positive, neutral, and negative. The methods used in this research are Adaptive Synthetic Multinomial Naive Bayes, Adaptive Synthetic k-nearest neighbours, and Adaptive Synthetic Random Forest. The Adaptive Synthetic method is used to handle unbalanced data. The data used in this study are public arguments per province in Indonesia. The results obtained in this study are negative sentiments that dominate all provinces in Indonesia. There is a relationship between negative sentiment and the level of education, internet use, and the human development index. Adaptive Synthetic Multinomial Naive Bayes performed better than other methods, with an accuracy of 0.882. The highest accuracy of the Adaptive Synthetic Multinomial Naive Bayes method is 0.990 in Papua Barat Province.*

Keywords: adaptive synthetic, classification, imbalance data, accuracy

NIEZRÓWNOWAŻONA KLASYFIKACJA WIELOKLASOWA Z ADAPTACYJNYM SYNTETYCZNYM WIELOMIANOWYM NAIWNYM PODEJŚCIEM BAYESA

Streszczenie. *Należy przyjrzeć się i przeanalizować opinie związane z rosnącymi cenami paliw. Opinia publiczna jest ściśle związana z polityką publiczną Indonezji w przyszłości. Twitter jest jednym z mediów, których ludzie używają do przekazywania swoich opinii. Niniejsze badanie wykorzystuje analizę nastrojów, aby przyjrzeć się temu zjawisku. Opinia jest podzielona na trzy kategorie: pozytywną, neutralną i negatywną. Metody wykorzystane w tym badaniu to Adaptive Synthetic Multinomial Naive Bayes, Adaptive Synthetic k-nearest neighbours i Adaptive Synthetic Random Forest. Metoda Adaptive Synthetic służy do obsługi niezrównoważonych danych. Dane wykorzystane w tym badaniu to argumenty publiczne według prowincji w Indonezji. Wyniki uzyskane w tym badaniu to negatywne nastroje, które dominują we wszystkich prowincjach Indonezji. Istnieje związek między negatywnymi nastrojami a poziomem wykształcenia, korzystaniem z Internetu i wskaźnikiem rozwoju społecznego. Adaptive Synthetic Multinomial Naive Bayes działała lepiej niż inne metody, z dokładnością 0,882. Najwyższa dokładność metody Adaptive Synthetic Multinomial Naive Bayes wynosi 0,990 w prowincji Papua Barat.*

Słowa kluczowe: adaptacyjna synteza, klasyfikacja, dane dotyczące nierównowagi, dokładność

Introduction

Humans are created to interact with each other. Besides that, another essential human trait is to respond to a phenomenon around them [8]. Phenomena that are often answered to by society are phenomena that have a direct impact on their lives, for example, primary needs. The community often discusses economic, political, environmental, and humanitarian issues. Moreover, social media is growing very rapidly, so there are more and more platforms for people to express their opinions. Twitter is one of the favourite social media for people to express their views [2].

The expression of public opinion is closely related to the level of education, internet access, and human development index in a country, as well as Indonesia. People's critical thinking skills are trained in education and commenting on a phenomenon [37]. Moreover, education is strongly correlated with the human development index because education is one of the variables in the human development index. Internet access is crucial in the digital age, as most information is online.

The increase in fuel oil (BBM) became a topic of discussion among Indonesians in September 2022 [29]. Fuel is a primary need for most Indonesians. The mobility of Indonesian people is very high, and Fuel is crucial for them. Public comments about the fuel increase were conveyed on social media, and it became a trending topic. Twitter is the social media chosen by the public to express their opinions.

Public sentiment can be classified as positive or neutral, or negative. Sentiment analysis on fuel price increases needs to be done to determine whether people's opinions are classified as positive, neutral, or negative. In addition, the sentiment analysis results can be used as input for the government in taking and evaluating a policy.

In recent decades, Twitter sentiment analysis has used machine learning techniques to classify sentiment into positive, neutral, and negative classes [23]. Sentiment classification is crucial for the government in deciding future policies. A comparison of classification methods between logistic regression, Naive Bayes, Support Vector Machine (SVM), and Stochastic Gradient Descent (SGD) were carried out to obtain the best way to classify Bengali book reviews, Multinomial Naive Bayes (MNB) being the best method with 84% accuracy [16]. The MNB method was applied by Rahman et al. [32] to classify sentiment in Bengali films, and the technique gave an accuracy of 86%. Several other studies on sentiment classification using the MNB method [7, 10, 32, 35, 38].

K-Nearest Neighbors (k-NN) is a machine learning method that uses distance techniques in classifying. Zamzuri et al. [41] use the k-NN method to classify the emotions contained in the text. The results obtained are the k-NN method which is capable of classifying with an accuracy of 79%. Hotel reviews in e-commerce applications using k-NN produce an accuracy of 87% [11].

Besides the MNB and k-NN methods, the Random Forest (RF) method accurately classifies text. The RF method produces better accuracy than the Support Vector Machine (SVM) and logistic regression in classifying feedback and reviewing airline services [31]. An accuracy of 96.42% was obtained by the RF method in classifying public opinion about sexual harassment cases from Brazilian anesthesiologists [5].

Conventional machine learning methods have the limitation of classifying datasets with unbalanced class distribution [40]. In addition, unbalanced data can reduce the accuracy of the classification method because there will be a prediction bias toward the majority class [19]. The phenomenon of fuel price increases in Indonesia has the potential to experience unbalanced data. The frequency of negative sentiment on the fuel increase will be more than neutral and positive.

Furthermore, methods of handling imbalanced datasets are needed; one such method is the Adaptive Synthetic (ADASYN) sampling approach [40]. The ADASYN algorithm works by generating synthetic data from minority classes. In the research of Khan et al. [36], the ADASYN method was able to increase the accuracy of the Extreme gradient boosting multi-classifier from 93% (without ADASYN) to 95% (with ADASYN). This method shows that the classification method will increase accuracy if the data is balanced. Some research on the ADASYN method [6, 20, 42].

Furthermore, the main contribution of this research is to describe the sentiment of fuel price increases per province in Indonesia and opinions that will describe the level of education, internet facilities, and the Human Development Index. In addition, classification using the ADASYN Multinomial Naive Bayes (MNB), ADASYN k-Nearest Neighbors (k-NN), and ADASYN Random Forest (RF) methods will be applied to classify positive, neutral, and negative sentiments in each province in Indonesia. The best method will be evaluated with the accuracy value.

1. Material and methods

This research used multinomial naive bayes for classify sentiments of people in Indonesia about increasing fuel, the sentiments divided into three categories. The three categories are positive, negative, and neutral statements. The positive statement is agreement arguments with increasing fuel, while disagreement is represented by negative statements, and neutral is impartial statement.

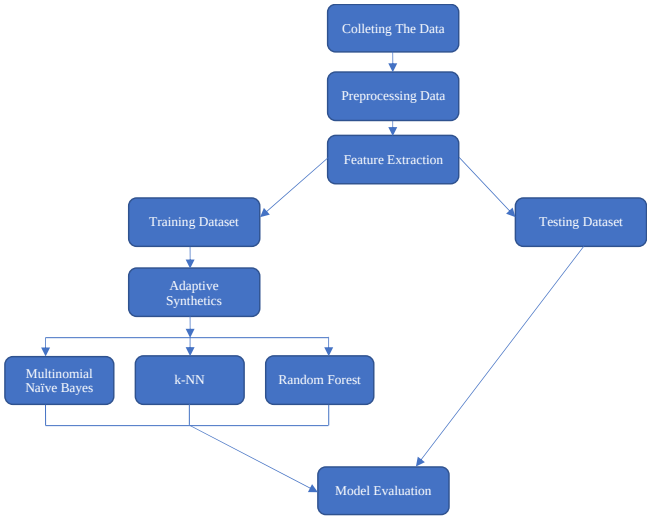


Fig. 1. The Research Step

The first step (figure 1) is collecting the data form twitter with the keywords "#BBMNaik", "BBM Naik", and "BBM", this step repeated for 34 provinces in Indonesia. Data collection in this study using the scraping method. Web scraping is a technique for automatically extracting information from various web documents. Get relevant content based on a query and convert unstructured format to structured representation [26]. The number of data retrieved in each province was as many as 100 tweets with the total tweets used in this study as many as 3400, where the tweets taken in Bahasa. The Sample of collected data is as follows in table 1.

Table 1. Sample of the web scraped data

Username	Tweet	User Location
@SilviaPutrii9	Penyesuaian bbm bertujuan untuk mengurangi beban subsidi besar akibat harga minyak dunia yang terus naik, kebijakan pemerintah sudah benar#BantuanBBMUntukRakyat #BLTBBMTepatSasaran #HematCermatBBM	Pontianak, Kalimantan Barat
@inthesky014	BBM naik itu biar yang sering bawa motor ke masjid pada jalan kaki, supaya pahalanya lebih banyak, subhanallah pemerintah kita ini sangat memperdulikan iman rakyatnya	Cianjur, Indonesia
@Kentringmanikk	@M45Broo_ Saya bisnis laundry dukung 1000% BBM naik. Bila perlu naikan terus, hilangkan subsidi BBM. Larang mobil pake BB fosil secepatnya. Subsidi bisa di pakai utk hal lain. Sudah tdk efisien negeri ini	Tanjung Emas, Indonesia

1.1. Text preprocessing

Results of collecting data was unstructured, so the next step is preprocessing data with Natural Language Processing (NLP) machine learning model [21]. Natural Language Processing have seven procedures, namely case folding, word normalization, cleansing filtering, stemming, and tokenizing.

a) Case Folding

The data is converted into the lowercase so that the uppercase and lowercase words with same meaning are not treated differently [27].

Table 2. Case Folding

Raw Data	Case Folding
Kenaikan harga BBM adalah hal sangat wajar, mengingat harga minyak dunia yang sedang melambung pesat #BantuanBBMUntukRakyat	kenaikan harga bbm adalah hal sangat wajar mengingat harga minyak dunia yang sedang melambung pesat #bantuanbbmuntukrakyat

b) Word Normalization

Word normalization is used to change words that are not standard as an informal word or shortened to a standard word in Bahasa.

Table 3. Word Normalization

Case Folding	Word Normalization
kenaikan harga bbm adalah hal sangat wajar mengingat harga minyak dunia yang sedang melambung pesat #bantuanbbmuntukrakyat	kenaikan harga bbm adalah hal sangat wajar mengingat harga minyak dunia yang sedang melambung pesat #bantuanbbmuntukrakyat

c) Cleansing

Cleansing is used to clean words that are not required such as hashtag (#), website address, username (@username), numbers, emojis and emails.

Table 4. Cleansing

Word Normalization	Cleansing
kenaikan harga bbm adalah hal kenaikan harga bbm adalah hal sangat wajar mengingat harga sangat wajar mengingat harga minyak dunia yang sedang minyak dunia yang sedang melambung pesat melambung pesat #bantuanbbmuntukrakyat	kenaikan harga bbm adalah hal sangat wajar mengingat harga minyak dunia yang sedang melambung pesat

d) Filtering

There are some words in the tokenized text that do not relate to any important concept or result, but may have important implications for the classifier. It is better to delete such words in advance.

Table 5. Filtering/Stopwords Removal

Cleansing	Filtering/Stopword Removal
kenaikan harga bbm adalah hal sangat wajar mengingat harga minyak dunia yang sedang melambung pesat	kenaikan harga wajar harga minyak dunia melambung pesat

e) Stemming

This step is to find the roots of words with deletion suffixes.

Table 6. Stemming

Filtering/Stopword Removal	Filtering/Stopword Removal
kenaikan harga wajar harga minyak dunia melambung pesat	naik harga wajar harga minyak dunia lambung pesat

f) Tokenization

Tokenization refers to the process of converting any text into a series of tokens, each distinct and independent of the other.

Table 7. Tokenization

Token
naik
harga
wajar
harga
minyak
dunia
lambung
pesat

1.2. Feature extraction

The process of converting text data to numbers is called Feature Extraction from text. This is also called text vectorization. The text contained in these tweets is unstructured, so in order to process it, it first needs to be pre-processed, six pre-processing techniques are used, and then features are extracted from the pre-processed data [1]. One of the feature extraction methods is TF-IDF (Term Frequency-Inverse Document Frequency). TF-IDF is an algorithmic method useful for calculating the weight of any commonly used word This method is also known to be efficient, simple and gives accurate results. This approach would calculate the TF and IDF values of each token (word) in each document of the corpus. Generally, the TF-IDF method is used to find out how many times a word occurs in a document [34]. The calculation to find the TF-IDF value is as follows:

$$TF_{t,d}IDF_{t,d} = tf_{t,d} \times idf_t \quad (1)$$

where $TF_{t,d}IDF_{t,d}$ is the weight of the term (t_j) to the documents (d_i). The $tf_{t,d}$ value is the frequency term (t) in document (d). However, if the term is not included in the document, the weight is zero [18]. Calculate the idf_t value:

$$idf_t = log_e \frac{1+n}{1+df_t} + 1 \quad (2)$$

where n is the number of whole documents in the collections while $\mathrm{df_t}$ is that of documents containing term (t).

1.3. Multinomial Naive Bayes (MNB)

Multinomial Naive Bayes (MNB) models the distribution word in the document as a multinomial. A document is treated as a sequence of words and assumed that each word position is generated independently of each other [37]. Multinomial Naive Bayes Classifier can be formulated as follows:

A tweet 'n' being of polarity 'p' is calculated as [30]:

$$P(p|n) \propto P(p) \prod_{1<k<nd} P(t_k|p) \quad (3)$$

where $P(t_k|p)$: represents the conditional probability that whether the term t_k occurs in a tweet of polarity p which is calculated as follows:

$$P(t_k|p) = \frac{count(t_k|p)+1}{count(t_p)+|V|} \quad (4)$$

Here, count ($t_k|p$) means the number of times the term t_k occurs in the tweets which have polarity p and count (t_p) means the total number of tokens present in the tweets of polarity p. Also, 1 and $|V|$ are added as smoothing constants which are added to avoid the mishaps in the calculation when the term does not occur at all in the tweets or the tweets is empty or null. This concept is better known as Laplace Smoothing [3]. $|V|$ is the number of terms in the total vocabulary of tweets.

$P(p)$: represents the prior probability of tweets being of polarity p which is calculated as follows:

$$P(p) = \frac{Number\ of\ tweets\ of\ polarity\ p}{Total\ number\ of\ tweets} \quad (5)$$

1.4. K-Nearest Neighbours (k-NN)

The k-NN classifier is a direct non-parametric classification algorithm [28]. A non-parametric method means that it does not make any assumptions about the distribution of the underlying data. Non-parametric algorithms such as k-NN use a flexible number of parameters, which often increases as more data becomes available [13].

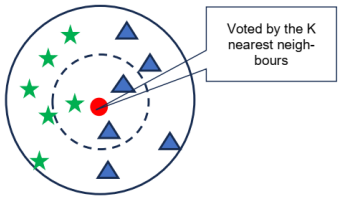


Fig. 2. Illustration of k-NN Algorithm

The k-NN algorithm is a classification algorithm that uses some K nearest data (neighbours) to determine a new data class. This algorithm classifies data based on their similarity or proximity to other data. To calculate the distance between two data in the k-NN algorithm using the Euclidean Distance method. The line length between points p and q is the Euclidean distance between them. If pi and qi are two locations in Euclidean n-space in Cartesian coordinates, then the distance from p to q is given by

$$d_E = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \quad (6)$$

In general, the k-NN logarithm consists of three parts [4, 13]:

1. Determine the number of neighbours (K) used for class determination considerations.
2. Calculate the distance from the new data to each data point in the dataset.
3. Take several K data with the shortest distance, then determine the class of the new data.

1.5. Random Forest (RF)

Random Forest is a machine learning algorithm combining multiple decision trees' output to arrive at a single result [25]. Each tree in the Random Forest will issue class predictions. The class prediction with the most votes becomes the prediction candidate for the model. The greater the number of trees, the higher accuracy will result and prevent overfitting problems.

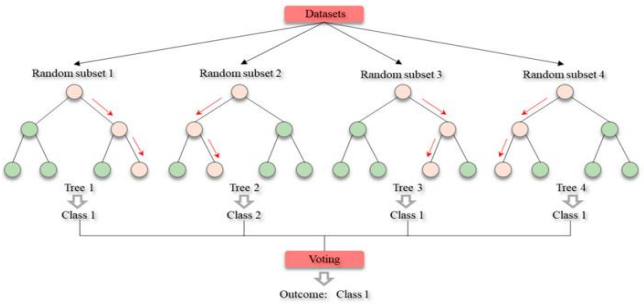


Fig. 2. Illustration of Random Forest Algorithm[39]

The Random Forest algorithm works can be described in the following steps [12].

1. Make a decision tree for each selected sample. Then the prediction results will be obtained from each decision tree that has been made.
2. A voting process is carried out for each prediction result. For classification problems, use the mode (the value that appears most often); for regression problems, use the mean (average value).
3. The algorithm will choose the prediction result with the most votes (most votes) as the final prediction.

1.6. Feature selection

Feature selection is the stage that helps reduce data size and remove features that don't important and improves accuracy [15]. Feature selection serves to reduce the size of the data dimensions as well as aims to select the best features from a feature data set. Feature selected by selecting important and relevant features to the data and reducing irrelevant features. The feature selection method used in this study is Chi Square. The chi square formula is as follows:

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \quad (7)$$

where O_i is the observation value for category i and E_i is expected value for category i.

2. Results and discussion

We begin our analysis by looking at the spread of sentiment in Indonesia. The classification of sentiment in this study is divided into three categories, namely positive sentiment, neutral sentiment, and negative sentiment. The distribution of these sentiments illustrates the perspective of the Indonesian people towards rising BBM prices. In this case, we look at sentiment per province. Distribution can help local governments to make public policies.

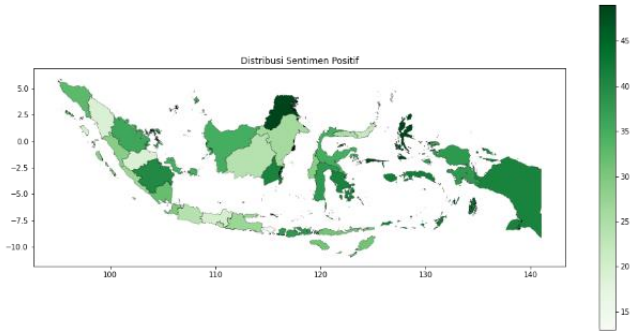


Fig. 4. Distribution of Positive Sentiment

Positive statements are argument that support the increasing BBM price. Based on figure 4, many people in the province disagree with the increase in fuel prices. This is indicated by a light map color. Meanwhile, the provinces of Kalimantan Utara and Maluku Utara were the provinces most supportive of the increase in fuel prices compared to other provinces.

The neutral statement illustrates that public opinion does not care about the increase in fuel prices. The provinces that distributed the most normal reports were Aceh, Jawa Tengah, and the D.I Yogyakarta (figure 5). The highest number of statements is 33. However, most provinces in Indonesia have neutral statements that are low on affiliation.

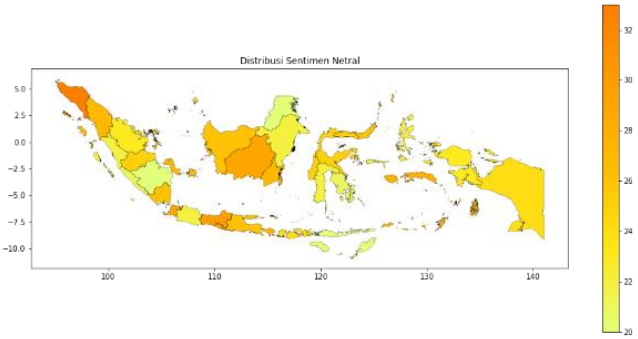


Fig. 5. Distribution of Neutral Sentiment

The majority of the general public reject the increase in fuel prices in Indonesia. This is indicated by the distribution of red in many parts of Indonesia. The province of Bali is the province with the highest negative statement with 58 sentiments. Other regions with a fairly strong red color are DKI Jakarta with 57 negative sentiments, Jambi with 56 negative sentiments and Jawa Barat with 54 negative sentiments. People in these areas often give their opinions in the form of news and suggestions so that these opinions are categorized as negative sentiments (figure 6).

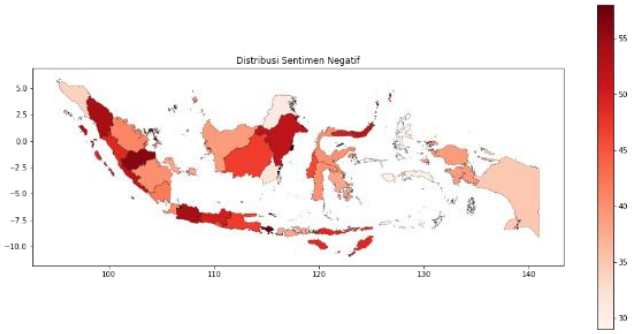


Fig. 6. Distribution of Negative Sentiment

The rejection was carried out because the fuel increase would hit people's purchasing power when salaries did not increase, so the additional allocation to buy fuel would increase and curtail other budgets so that people's purchasing power would decrease. In addition, according to news published by VOA Indonesia related to a national survey of Indonesian political indicators, most people reject the increase in subsidized fuel prices. The wide price gap between subsidized and non-subsidized fuel is considered ineffective in limiting subsidized fuel consumption. The survey results reveal that in terms of income, rejection of the fuel increase comes from those who earn more than Rp 4 million (54.2 percent) and people who make less than Rp1 million per month (69.1 percent). In terms of the type of work, the rejection of fuel increases generally comes from civil servants / private sector (82 percent) and groups of farmers, ranchers, and fishermen (67.3 percent).

Education shapes humans to think critically. Critical thinking is closely related to people expressing opinions on social media. People who express opinions in the media are educated people [9]. Several provinces with high levels of education wrote negative statement comments regarding the increase in fuel prices. It can be seen in figure 7 that DKI Jakarta has the highest level of education compared to other provinces in Indonesia. It is directly proportional to the number of negative comments about the fuel price increase.

The distributions above (figures 4–6) show positive, neutral, and negative reviews and show that the distribution of education levels is uneven. The level of education is directly proportional to the infrastructure (search the literature). The more infrastructure supporting education, the better the level of education in the area.

The word cloud results visualize the words that Twitter users coonly use to express their opinions. The bigger the font size on word cloud then the topic frequently discussed by Twitter users. The following figure shows the word cloud's results.

Based on figure 8 it can be seen positive sentiment, neutral sentiment, and negative sentiment along with the words used. This is useful to know the description of public sentiment in every province in Indonesia regarding the increase in fuel prices set by the government. The word "Harga" dominates in the neutral and negative sentiments. The difference lies in the next term. The negative sentiment contains a sentence that protests. The distribution of positive, neutral and negative sentiment in other provinces has the same pattern.

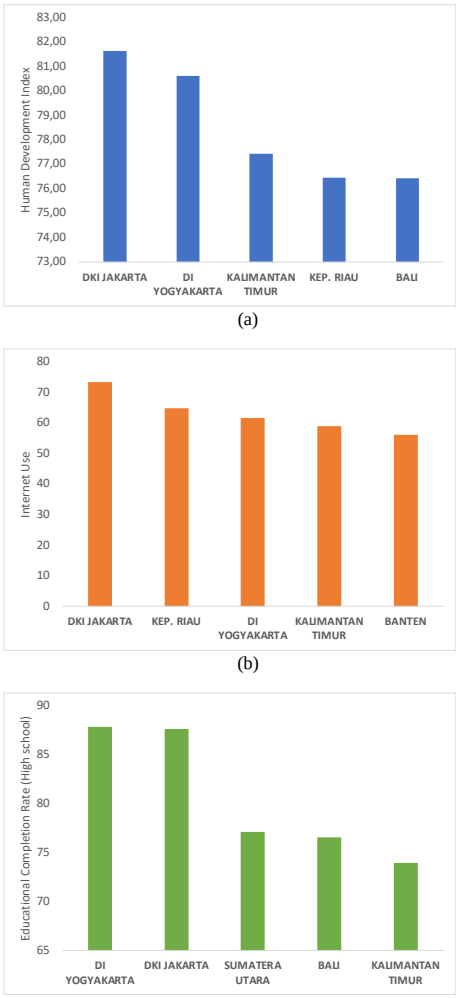


Fig. 7. Variables that describe the level of education (the highest 5 provinces): (a) Human Development Index, (b) Internet Use, (c) Educational Completion Rate (Heigh School)

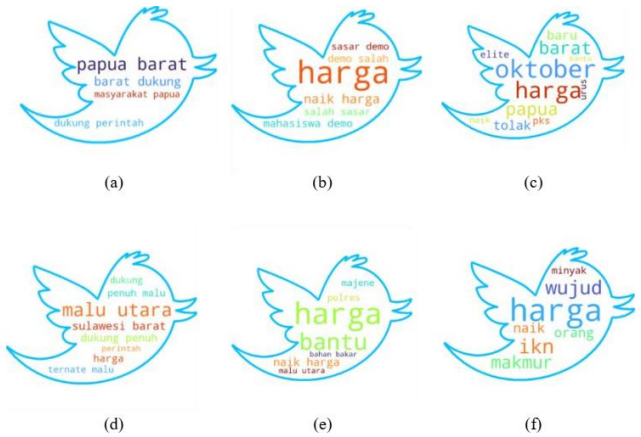


Fig. 8. Sample Wordcloud Sentiment Increase in fuel prices: (a) Papua Barat Positive, (b) Papua Barat Neutral, (c) Papua Barat Negative, (d) Sulawesi Barat Positive, (e) Sulawesi Barat Neutral, (f) Sulawesi Barat Negative

Public response to a particular phenomenon cannot be uniform. If the phenomenon is positive, the frequency of positive comments on social media is more than negative comments, as well as on negative phenomena. See figure 9 (a), the amount of data in each category is disproportionate. The amount of data in the negative category tends to be higher when compared to the neutral category and the positive category. This is imbalance data class.

Imbalanced class data is a common problem in classification machine learning where there is a disproportionate ratio of each class. Most machine learning algorithms don't work very well with imbalanced class data. If this is allowed, then the machine learning model that is built will tend to predict data into the majority data class.

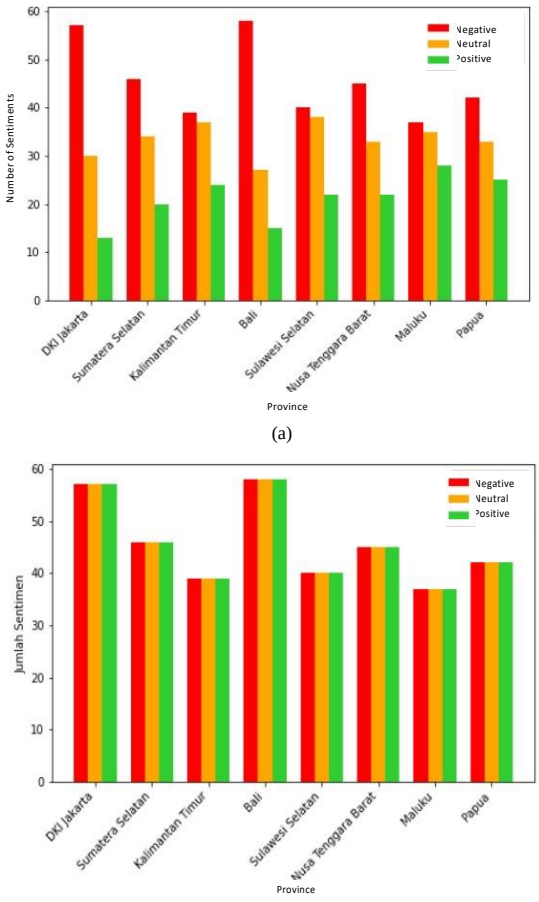


Fig. 9. Before and After Applied ADASYN: (a) before, (b) after

The Adaptive Synthetic (ADASYN) method is used to overcome this problem [14, 17, 22, 24]. After applying ADASYN (figure 9(b)), the data in each category becomes proportional. The data distribution in each category is even so that the data is ready to be used to build machine learning models. Machine learning models are built using Multinomial Naive Bayes (MNB), k-Nearest Neighbours (k-NN), and Random Forest (RF) algorithms.

The classifier model that has been built is used to test data testing. Each data is predicted using the MNB, k-NN, and RF algorithms to be classified into three sentiment categories: positive, neutral, or negative. In this study, researchers tested 680 tweets (data testing) related to rising fuel prices.

Evaluation of results using the confusion matrix to determine the level of accuracy. Based on the data testing process, obtained the level of accuracy in each province in Indonesia is shown in the table below.

Based on table 8, the accuracy of the MNB algorithm is better than the k-NN and RF methods, with an average accuracy of 0.882 (the highest compared to other methods). In general, the MNB accuracy in each province is the highest compared to the k-NN and RF methods. The highest accuracy is in Papua Barat Province, with an accuracy value of 0.990, while the lowest is in Sulawesi Utara Province. The high accuracy value is because the testing data has the vocabulary contained in the training data, so the model can learn it based on the training data and adequately process the testing data.

Table 8. Accuracy

Province	Accuracy		
	MNB	K-NN	RF
Aceh	0.905	0.95	0.80
Bali	0.956	0.90	0.95
Bangka Belitung	0.970	0.80	0.95
Banten	0.960	0.95	0.80
Bengkulu	0.806	0.80	0.75
DI. Yogyakarta	**0.980**	0.95	0.85
DKI Jakarta	0.882	0.90	0.95
Gorontalo	0.942	0.85	0.80
Jawa Barat	0.930	0.95	0.85
Jambi	0.970	0.90	0.80
Jawa Tengah	0.980	0.90	0.90
Jawa Timur	0.966	0.95	0.90
Kalimantan Barat	0.921	0.90	0.90
Kalimantan Selatan	0.980	0.95	0.90
Kalimantan Tengah	0.941	0.85	0.95
Kalimantan Timur	0.900	0.75	0.80
Kalimantan Utara	0.667	0.85	0.95
Kep. Riau	0.933	0.80	0.90
Lampung	0.857	0.80	0.85
Maluku	0.818	0.80	0.70
Maluku Utara	0.957	0.75	0.80
Nusa Tenggara Barat	0.960	0.90	0.90
Nusa Tenggara Timur	0.817	0.80	0.80
Papua	0.814	0.85	0.95
Papua Barat	0.990	0.95	0.95
Riau	0.671	0.85	0.75
Sulawesi Barat	0.986	0.95	0.80
Sulawesi Selatan	0.775	0.90	0.85
Sulawesi Tengah	0.833	0.90	0.85
Sulawesi Tenggara	0.897	0.90	0.80
Sulawesi Utara	0.571	0.65	0.70
Sumatera Barat	0.967	0.90	0.80
Sumatera Selatan	0.750	0.75	0.80
Sumatera Utara	0.750	0.85	0.75
Average Accuracy	**0.882**	0.860	0.850

3. Conclusion

Based on the results of this study it can be seen that the province with a dense population such as DKI Jakarta, Jawa Barat, and Bali tend to resist the increase in fuel prices. The majority of people on each province uses the word “harga” in giving its opinion regarding the increase in fuel prices either negative sentiment, neutral sentiment, or positive sentiment. There is a relationship between negative sentiment and the human development index, internet use, and educational completion rate (high school). The performance of the machine learning model that has been built is measured using accuracy. The ADASYN method is applied to balance the minority and majority classes. The ADASYN Multinomial Naive Bayes (MNB) method has better performance than the ADASYN k-Nearest Neighbors (k-NN) and ADASYN Random Forest (RF) methods. The average accuracy obtained by the MNB method is 0.882 or 88.2%. Near-perfect accuracy values are in the province of Papua Barat. You can use the hybrid method and add testing data to improve performance in further research.

References

[1] Ahuja R. et al.: The Impact of Features Extraction on the Sentiment Analysis. Procedia Computer Science 152, 2019, 341–348 [http://doi.org/10.1016/j.procs.2019.05.008].

[2] Ali H. et al.: Deep Learning-Based Election Results Prediction Using Twitter Activity. Soft Computing 26(16), 2022, 7535–43 [http://doi.org/10.1007/s00500-021-06569-5].

[3] Amity U. et al.: Abstract Proceedings of International Conference on Automation, Computational and Technology Management (ICACTM-2019), 2019.

[4] Andrian R. et al.: K-Nearest Neighbor (k-NN) Classification for Recognition of the Batik Lampung Motifs. Journal of Physics: Conference Series 1338(1), 2019 [http://doi.org/10.1088/1742-6596/1338/1/012061].

[5] Asian J. et al.: Sentiment Analysis for the Brazilian Anesthesiologist Using Multi-Layer Perceptron Classifier and Random Forest Methods. Journal Online Informatika 7(1), 2022, 132 [http://doi.org/10.15575/join.v7i1.900].

[6] Balaram A., Vasundra S.: Prediction of Software Fault-Prone Classes Using Ensemble Random Forest with Adaptive Synthetic Sampling Algorithm. Automated Software Engineering 29(1), 2021, 6 [http://doi.org/10.1007/s10515-021-00311-z].

[7] Budiawan Zulfikar W. et al.: Sentiment Analysis on Social Media Against Public Policy Using Multinomial Naive Bayes. Scientific Journal of Informatics 10(1), 2023 [http://doi.org/10.15294/sji.v10i1.39952].

[8] Bustillos A. et al.: Approaching Dehumanizing Interactions: Joint Consideration of Other-, Meta-, and Self-Dehumanization. Current Opinion in Behavioral Sciences 49, 2023, 101233 [http://doi.org/10.1016/j.cobeha.2022.101233].

[9] Eberwein T.: ‘Trolls’ or ‘Warriors of Faith’?: Differentiating Dysfunctional Forms of Media Criticism in Online Comments. Journal of Information, Communication and Ethics in Society 18(1), 2020, 131–143 [http://doi.org/10.1108/JICES-08-2019-0090].

[10] Farisi A. A. et al.: Sentiment Analysis on Hotel Reviews Using Multinomial Naive Bayes Classifier. Journal of Physics: Conference Series 1192(1), 2019 [http://doi.org/10.1088/1742-6596/1192/1/012024].

[11] Gazali Mahmud F. et al.: Implementation Of K-Nearest Neighbor Algorithm With SMOTE For Hotel Reviews Sentiment Analysis. Sinkron: Jurnal Dan Penelitian Teknik Informatika 8(2), 2023, 595–602 [http://doi.org/10.33395/sinkron.v8i2.12214].

[12] Ghosh D., Cabrera J.: Enriched Random Forest for High Dimensional Genomic Data. IEEE/ACM Transactions on Computational Biology and Bioinformatics 19(5), 2022, 2817–2828 [http://doi.org/10.1109/TCBB.2021.3089417].

[13] Hasdyna N. et al.: Improving the Performance of K-Nearest Neighbor Algorithm by Reducing the Attributes of Dataset Using Gain Ratio. Journal of Physics: Conference Series 1566(1), 2020 [http://doi.org/10.1088/1742-6596/1566/1/012090].

[14] He H. et al.: ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning. IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), 2008, 1322–1328 [http://doi.org/10.1109/IJCNN.2008.4633969].

[15] Herhianto A.: Sentiment Analysis Menggunakan Naive Bayes Classifier (Nbc) Pada Tweet Tentang Zakat. 2020.

[16] Hossain E. et al.: Sentiment Polarity Detection on Bengali Book Reviews Using Multinomial Naive Bayes. Progress in Advanced Computing and Intelligent Engineering (ed.Chhabi Rani Panigrahi et al.), Springer Singapore, 2021, 281–292.

[17] Hu Z. et al.: A Novel Wireless Network Intrusion Detection Method Based on Adaptive Synthetic Sampling and an Improved Convolutional Neural Network. IEEE Access 8, 2020, 195741–195751 [http://doi.org/10.1109/ACCESS.2020.3034015].

[18] Jalilifard A. et al.: Semantic Sensitive TF-IDF to Determine Word Relevance in Documents, 2020 [http://doi.org/10.1007/978-981-33-6977-1].

[19] Jiang C. et al.: Benchmarking State-of-the-Art Imbalanced Data Learning Approaches for Credit Scoring. Expert Systems with Applications 213, 2023, 118878 [http://doi.org/10.1016/j.eswa.2022.118878].

[20] Koh J. E. W. et al: Automated Classification of Attention Deficit Hyperactivity Disorder and Conduct Disorder Using Entropy Features with ECG Signals. Computers in Biology and Medicine 140, 2022, 105120 [http://doi.org/10.1016/j.compbiomed.2021.105120].

[21] Kurniasih A., Lindung P. M.: On the Role of Text Preprocessing in BERT Embedding-Based DNNs for Classifying Informal Texts. International Journal of Advanced Computer Science and Applications 13(6), 2022, 927–934 [http://doi.org/10.14569/IJACSA.2022.01306109].

[22] Kurniawati Y. E. et al.: Adaptive Synthetic-Nominal (ADASYN-N) and Adaptive Synthetic-KNN (ADASYN-KNN) for Multiclass Imbalance Learning on Laboratory Test Data. 2018 4th International Conference on Science and Technology (ICST), 2018, 1–6 [http://doi.org/10.1109/ICSTC.2018.8528679].

[23] Leelawat N. et al.: Twitter Data Sentiment Analysis of Tourism in Thailand during the COVID-19 Pandemic Using Machine Learning. Heliyon 8(10), 2022, e10894 [http://doi.org/10.1016/j.heliyon.2022.e10894].

[24] Liu J. et al.: A Fast Network Intrusion Detection System Using Adaptive Synthetic Oversampling and LightGBM. Computers & Security 106, 2021, 102289 [http://doi.org/10.1016/j.cose.2021.102289].

[25] Liu Y., Wu H.: Prediction of Road Traffic Congestion Based on Random Forest. 2017 10th International Symposium on Computational Intelligence and Design (ISCID) 2, 2017, 361–364 [http://doi.org/10.1109/ISCID.2017.216].

[26] Lytvyn V. et al.: Identifying Textual Content Based on Thematic Analysis of Similar Texts in Big Data. 2019 IEEE 14th International Conference on Computer Sciences and Information Technologies (CSIT) 2, 2019, 84–91 [http://doi.org/10.1109/STC-CSIT.2019.8929808].

[27] Mayo M.: A General Approach to Preprocessing Text Data, 2017.
[28] Moosavian A. et al.: Comparison of Two Classifiers; K-Nearest Neighbor and Artificial Neural Network, for Fault Diagnosis on a Main Engine Journal-Bearing. Shock and Vibration 20(2), 2013, 263–272 [http://doi.org/10.3233/SAV-2012-00742].
[29] Nadhifah D. et al.: Analysis of the Impact of the Increase in Fuel Oil (BBM) on Household Economic Activities. Journal of Contemporary Gender and Child Studies (JCGCS) 1(1), 2022 [https://zia-research.com/index.php/jcgcs].
[30] Nazrul Syed S.: Multinomial Naive Bayes Classifier for Text Analysis (Python). Towards Data Science, 2018.
[31] Patel A. et al.: Sentiment Analysis of Customer Feedback and Reviews for Airline Services Using Language Representation Model. Procedia Computer Science 218, 2023, 2459–2467 [http://doi.org/10.1016/j.procs.2023.01.221].
[32] Rahman R. et al.: Sentiment Analysis on Bengali Movie Reviews Using Multinomial Naive Bayes. 2021 24th International Conference on Computer and Information Technology (ICCIT), 2021, 1–6 [http://doi.org/10.1109/ICCIT54785.2021.9689787].
[33] Rennie J. D. M. et al.: Tackling the Poor Assumptions of Naive Bayes Text Classifiers, 2003.
[34] Ridho Lubis A. et al.: The Effect of the TF-IDF Algorithm in Times Series in Forecasting Word on Social Media. Indonesian Journal of Electrical Engineering and Computer Science 22(2), 2021, 976 [http://doi.org/10.11591/ijeecs.v22.i2.pp976-984].
[35] Sahib N. G. et al.: Sentiment Analysis of Social Media Comments in Mauritius. IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), 2023, 860–865 [http://doi.org/10.1109/CCWC57344.2023.10099291].
[36] Salauddin Khan M. et al.: Comparison of Multiclass Classification Techniques Using Dry Bean Dataset. International Journal of Cognitive Computing in Engineering 4, 2023, 6–20 [http://doi.org/10.1016/j.ijcce.2023.01.002].
[37] Solikah M., Dian N.: The Effectiveness of the Guided Inquiries Learning Model on the Critical Thinking Ability of Students. Jurnal Pijar Mipa 17(2), 2022, 184–191 [http://doi.org/10.29303/jpm.v17i2.3276].
[38] Surya P. P. et al.: Analysis of User Emotions and Opinion Using Multinomial Naive Bayes Classifier. 2019 3rd International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2019, 410–415 [http://doi.org/10.1109/ICECA.2019.8822096].
[39] Yang J. et al.: Delineation of Urban Growth Boundaries Using a Patch-Based Cellular Automata Model under Multiple Spatial and Socio-Economic Scenarios. Sustainability (Switzerland) 11(21), 2019 [http://doi.org/10.3390/su11216159].
[40] Yu B. et al.: Classification Method for Failure Modes of RC Columns Based on Class-Imbalanced Datasets. Structures 48, 2023, 694–705 [http://doi.org/10.1016/j.istruc.2022.12.063].
[41] Zamsuri A. et al.: Classification of Multiple Emotions in Indonesian Text Using The K-Nearest Neighbor Method. Journal of Applied Engineering and Technological Science (JAETS) 4(2), 2023, 1012–1021 [http://doi.org/10.37385/jaets.v4i2.1964].
[42] Zhai J. et al.: Binary Imbalanced Data Classification Based on Diversity Oversampling by Generative Models. Information Sciences 585, 2022, 313–43 [http://doi.org/10.1016/j.ins.2021.11.058].

M.Sc. Fatkhurokhman Fauzi
e-mail: fatkhurokhmanf@unimus.ac.id

He is a lecturer in the statistics department. Research focus: data mining, text mining, machine learning, forecasting, environmental statistics, and climate modeling. He is currently a researcher in the Degrees Initiative research group.

http://orcid.org/0000-0002-8277-8638

B.Sc. Ismatullah
e-mail: ismatullahp17@gmail.com

Research focus: data mining, text mining, and machine learning.

http://orcid.org/0009-0005-7472-1761

M.Sc. Indah Manfaati Nur
e-mail: indahmnur@unimus.ac.id

She is a lecturer in the statistics department. Research focus: text mining and applied statistics.

http://orcid.org/0000-0002-1017-7323

http://doi.org/10.35784/iapgos.3652 *received: 15.05.2023 | revised: 31.08.2023 | accepted: 24.09.2023 | available online: 30.09.2023*

COMPARISON OF THE EFFECTIVENESS OF TIME SERIES ANALYSIS METHODS: SMA, WMA, EMA, EWMA, AND KALMAN FILTER FOR DATA ANALYSIS

Volodymyr Lotysh, Larysa Gumeniuk, Pavlo Humeniuk
Lutsk National Technical University, Faculty of Computer and Information Technologies, Department of Automation and Computer-Integrated Technologies, Lutsk, Ukraine

Abstract. *In time series analysis, signal processing, and financial analysis, simple moving average (SMA), weighted moving average (WMA), exponential moving average (EMA), exponential weighted moving average (EWMA), and Kalman filter are widely used methods. Each method has its own strengths and weaknesses, and the choice of method depends on the specific application and data characteristics. It is important for researchers and practitioners to understand the properties and limitations of these methods in order to make informed decisions when analyzing time series data. This study investigates the effectiveness of time series analysis methods using data modeled with a known exponential function with overlaid random noise. This approach allows for control of the underlying trend in the data while introducing the variability characteristic of real-world data. The relationships were written using scripts for the construction of dependencies, and graphical interpretation of the results is provided.*

Keywords: data analysis, modeling, moving average, Kalman filter

PORÓWNANIE SKUTECZNOŚCI METOD ANALIZY SZEREGÓW CZASOWYCH: SMA, WMA, EMA, EWMA I FILTR KALMANA DO ANALIZY DANYCH

Streszczenie. *W analizie szeregów czasowych, przetwarzaniu sygnałów i analizie finansowej szeroko stosowane są: prosta średnia ruchoma (SMA), ważona średnia ruchoma (WMA), wykładnicza średnia ruchoma (EMA), wykładniczo-ważona średnia ruchoma (EWMA) i filtr Kalmana. Każda z metod ma swoje mocne i słabe strony, a wybór metody zależy od konkretnego zastosowania i charakterystyki danych. Dla badaczy i praktyków ważne jest zrozumienie właściwości i ograniczeń tych metod w celu podejmowania świadomych decyzji podczas analizy danych szeregów czasowych. W niniejszej pracy zbadano skuteczność metod analizy szeregów czasowych z wykorzystaniem danych modelowanych znaną funkcją wykładniczą z nałożonym szumem losowym. Takie podejście pozwala na kontrolowanie głównego trendu w danych przy jednoczesnym wprowadzeniu zmienności typowej dla danych rzeczywistych. Do budowy zależności zostały napisane skrypty. Podana jest graficzna interpretacja wyników.*

Słowa kluczowe: analiza danych, modelowanie, średnia ruchoma, filtr Kalmana

Introduction

Time series analysis is an important tool in many scientific fields, from finance to biology. Accurately modeling and forecasting time series data can be a challenging task, and to solve this problem, numerous statistical methods have been developed. Among the most common methods are Simple Moving Average (SMA), Weighted Moving Average (WMA), Exponential Moving Average (EMA), Exponential Weighted Moving Average (EWMA), and Kalman filter.

SMA is the basic method of smoothing time series data by calculating the average of a fixed number of previous values.

WMA extends the basic idea of SMA by assigning weights to previous values, with the most weight usually given to the most recent values.

EMA is a more complex method that exponentially weighs previous values, with the most weight given to the most recent values.

EWMA is a variant of EMA that allows for different smoothing parameters for different time periods.

The Kalman filter is a powerful tool that can be used to estimate the state of a system based on noisy measurements. It can be applied to time series data by treating the underlying system as a state space model and using the filter to estimate unobserved state variables.

Each of these methods has its strengths and weaknesses, and the choice of method depends on the specific problem facing the researcher. In this article, we will compare these methods by investigating their effectiveness on a given dataset.

The Simple Moving Average (SMA) is discussed in [5]. The authors note that SMA is a simple form of local polynomial regression. SMA is often used to smooth short-term fluctuations in time series data and to identify long-term trends. However, the authors note that SMA has limitations and may not be suitable for all time series data. Additionally, SMA does not take into account data variability, which can lead to erroneous results. Overall, although SMA is a simple and widely used method of smoothing time series data, it is important to consider its limitations and choose the appropriate method based on the characteristics of the data.

The Weighted Moving Average (WMA) is discussed in [3]. The authors present several methods for smoothing time series data, including Simple Moving Average (SMA), Exponential Smoothing, and Weighted Moving Average (WMA). Hyndman and Athanasopoulos provide examples of calculating weighted moving average using the R programming language and discuss the advantages and limitations of this smoothing method. They also compare WMA with other smoothing methods and provide recommendations for choosing an appropriate method based on the characteristics of the time series data and forecasting goals.

Exponential Moving Average (EMA) is also discussed in [3]. The authors describe EMA as a popular method for smoothing time series data and making short-term forecasts. They also describe the formula for calculating EMA, which includes specifying a smoothing parameter, often denoted as alpha, which determines the weight given to the most recent observation. They explain how to choose an appropriate alpha value based on the characteristics of the analyzed data and provide examples of applying the EMA method to different types of time series data. Overall, [3] is a comprehensive introduction to the EMA method and its application for time series forecasting.

The entire article [2] is dedicated to exponential smoothing, which is a method of smoothing time series data. The article discusses the basic idea of exponential smoothing and its various extensions, including the Exponentially Weighted Moving Average (EWMA).

In particular, the article explains that EWMA is a method of exponential weighting of past observations in computing the smoothed value of a time series. The weight assigned to each observation is based on the smoothing parameter, which determines the rate at which the weight decreases with increasing observation time. The article also compares the effectiveness of EWMA with other methods of smoothing time series data and discusses some practical issues related to implementing the method.

The Kalman filter is discussed in several sections of [1]. The authors present the Kalman filter as a method for estimating the state of a linear dynamic system based on noisy measurements. They also discuss the assumptions and limitations of the Kalman filter and provide several examples of its use in time series

analysis and forecasting. It is noted that the Kalman filter is a key component of state space models, as it allows for the estimation of latent variables in the presence of noisy observations. The authors discuss how to set up state space models for time series analysis and provide several examples of models that can be estimated using the Kalman filter. Overall, [1] provides a thorough introduction to the Kalman filter and its use in time series analysis and forecasting.

It is also worth noting classic publications dedicated to economic analysis, which consider time series [4, 6, 7].

Methods such as Simple Moving Average (SMA), Weighted Moving Average (WMA), Exponential Moving Average (EMA), Exponentially Weighted Moving Average (EWMA), and the Kalman filter are widely used in time series analysis, signal processing, and financial analysis.

Each method has its strengths and weaknesses, and the choice of method depends on the specific application and data characteristics.

1. Problem statement

Conduct a study of the effectiveness of time series analysis methods using a known exponential function with added random noise. Mean Squared Error (MSE) will be used as the effectiveness criterion.

2. Main results

Time series analysis is a powerful tool for analyzing data that changes over time. One of the most common applications of time series analysis is in finance, where it is used to detect trends and patterns in financial data and to forecast future values.

One way to study the effectiveness of time series analysis methods is by modeling data using a known exponential function with added random noise. This approach allows the main trend in the data to be controlled while introducing the variability characteristic of real data.

To model data using a known exponential function with random noise, the following formula can be used:

$$y(t) = A \cdot e^{(-k \cdot t)} + \varepsilon(t) \quad (1)$$

where A and k are constants that determine the shape of the exponential curve; t is the time index; $\varepsilon(t)$ is a random variable that represents the random noise.

The exponential function (1) at A = 100 and k = 0.00098 and the random noise imposed on it are shown in Fig. 1.

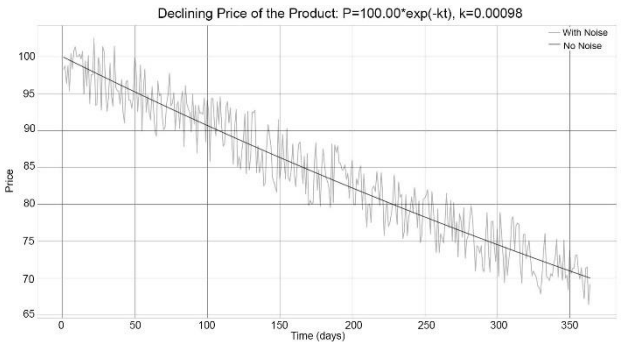


Fig. 1. The exponential function (1) at A = 100 and k = 0.00098 and the random noise imposed on it

Simple Moving Average (SMA) is a commonly used technique for analyzing time series data that helps to identify trends in a dataset over a specific period of time. It is a type of moving average that calculates the average value of a set of data points over a certain time period and is useful for smoothing out short-term fluctuations in the data.

$$SMA = \frac{\sum_{i=1}^{n} D_i}{p} \quad (2)$$

where Di is a data set; p is the length of smoothing or the period of the SMA (the number of values included in the calculation of the moving average).

To calculate the SMA, the sum of a set of data points over a certain period of time is divided by the number of data points in that period. For example, if we are interested in a 10-day SMA, we sum up the data point values for the last 10 days and divide that sum by 10.

SMA can be used to detect trends in a data set over time. When the SMA increases over time, it indicates that the underlying data points are also increasing. Conversely, when the SMA decreases over time, it indicates that the underlying data points are decreasing.

The Simple Moving Average can be configured for different time horizons depending on the specific analysis.

One of the drawbacks of SMA is that it gives equal weight to all data points in the moving average, regardless of how fresh they are. This can lead to a lag in detecting changes in the trend, as SMA may take some time to adapt to new information.

Also, the Simple Moving Average is not ideal for time series data with irregular structure. The smoothing effect of the Moving Average can hide important data features, such as sharp changes or extreme values..

SMA is less effective for long-term trends, as it can overestimate short-term fluctuations and mask the main long-term trend.

The graphical interpretation of the results of the SMA algorithm is shown in Fig. 2.

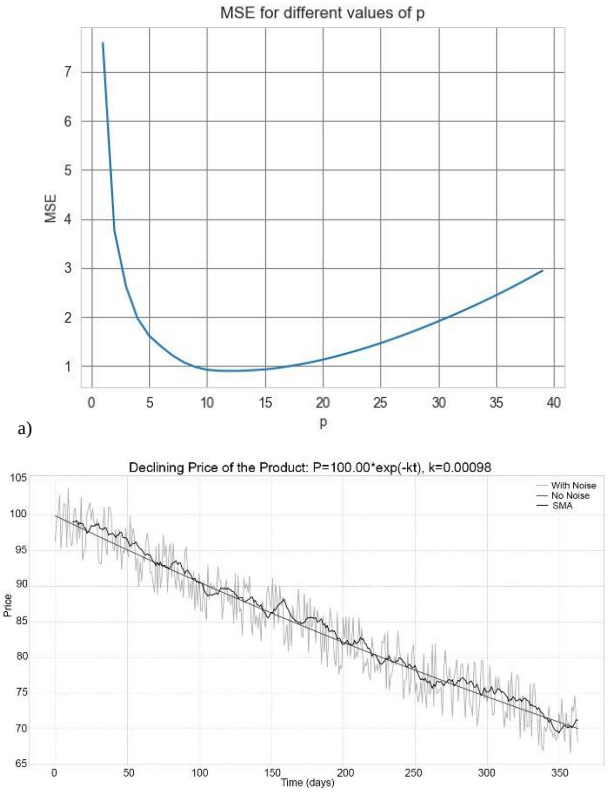


Fig. 2. a) Dependence of the root-mean-square error on the period of SMA; b) SMA at the minimum value of the root-mean-square error (MSE): p = 12; MSE = 0.90

Weighted Moving Average (WMA) is a method of analyzing time series data that is used to analyze and forecast trends in a series of data over time.

It works by computing the average value of a set of data points over a certain period, where each data point is given a weight based on its importance or relevance to the analysis. Weights are usually assigned in such a way that new data is given greater weight than old data, reflecting the fact that newer data is generally more relevant for predicting future trends.

$$WMA = \frac{\sum_{i=1}^{n} D_i \cdot p_i}{\sum_{i=1}^{n} p_i} \quad (3)$$

where Di represents the data for the i-period (with i = *1* being the current period); pi represents the weight values for the i-periods.

Weighted Moving Average (WMA) gives more weight to recent data points, making it more sensitive to changes in data than Simple Moving Average. This allows for better trend detection and forecasting of future values.

WMA can be customized by changing the weight of each data point in the moving average, providing greater flexibility in analysis.

WMA can be useful for smoothing time-series data with irregular structure or volatility, as weights can be adjusted to emphasize certain parts of the data.

However, WMA is prone to lag in detecting changes in trend, as the weights given to recent data points may not be enough to immediately incorporate new information.

The graphical interpretation of the results of the WMA algorithm is shown in Fig. 3.

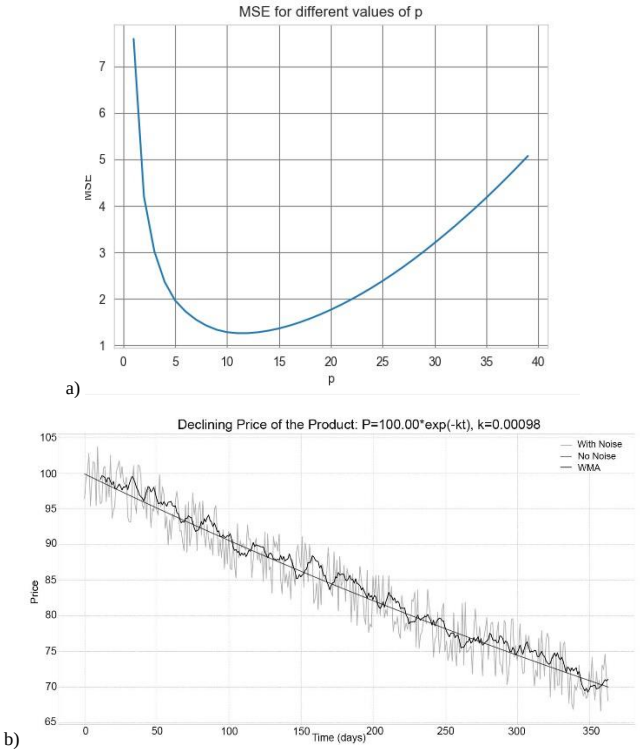


Fig. 3. a) Dependence of the root-mean-square error on weight values; b) WMA at the minimum value of the root-mean-square error (MSE): p = 11; MSE = 1.26

Exponential Moving Average (EMA) is a widely used method for analyzing time series data and predicting trends over time.

Like other Moving Average methods, EMA involves calculating the average value of a set of data points over a certain period. However, EMA gives more weight to the latest data points, using a weighting coefficient that exponentially decreases. This means that EMA responds better to recent changes in the data series and can provide a more accurate forecast of future trends.

$$EMA = \frac{EMA_{i-1} \cdot (p-1) + 2 \cdot D_i}{p+1} \quad (4)$$

where D_i is the value in the i-th period; p is the calculation period; EMA_{i-1} is the value of the previous period's EMA.

Exponential Moving Average (EMA) reacts better to changes in data than Simple Moving Average (SMA) and Weighted Moving Average (WMA). This is because EMA gives more weight to the latest data points and the weight decreases exponentially as the data points become older.

EMA can be customized by adjusting the smoothing factor, providing greater flexibility in analysis.

EMA is less prone to lag than SMA and WMA because it responds more quickly to changes in data.

However, EMA may be more sensitive to outliers than SMA and WMA because outliers have a greater impact on EMA than on other averages.

EMA is still subject to lag when detecting changes in trends because the smoothing factor may not be sufficient to immediately incorporate new information.

The graphic interpretation of the results of the EMA algorithm is shown in Fig. 4.

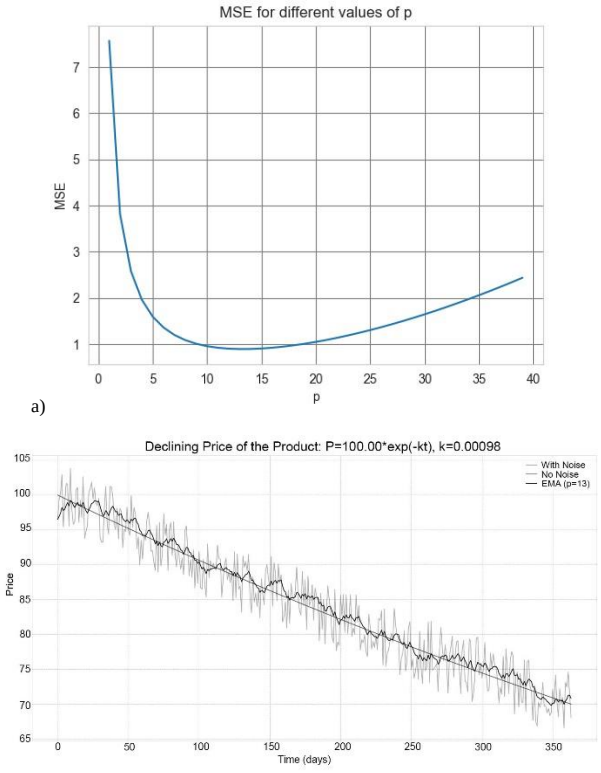


Fig. 4. a) Dependence of the root-mean-square error on the smoothing coefficient value; b) EMA at the minimum value of the root-mean-square error (MSE): p = 13; MSE = 0.89

Exponentially Weighted Moving Average (EWMA) is a time series analysis method used to analyze and forecast trends in data series over time.

EWMA is a variation of the traditional moving average method that gives more weight to recent data points than earlier ones, using a weight coefficient that exponentially decreases. The weight coefficient is usually a smoothing parameter chosen based on the characteristics of the analyzed data series.

$$EWMA = p \cdot r_t + (1-p) \cdot EWMA_{t-1} \quad (5)$$

where p is the weight coefficient; r is the data value in the current period.

EWMA provides greater flexibility in analysis than simple moving average and weighted moving average because it can be adjusted by changing the smoothing coefficient.

However, the exponentially weighted moving average may be more sensitive to outliers than simple moving average and weighted moving average, as outliers have a greater impact on the exponential average than on other averages.

A graphical interpretation of the results of the EWMA algorithm is shown in Fig. 5.

The Kalman filter is a mathematical algorithm used to estimate the state of a system based on a series of measurements. It is widely used in many fields, including finance, engineering, and science, to analyze numerical series and extract significant information from them.

In essence, the Kalman filter works by combining the forecast of the system state based on previous measurements with a new measurement of the system to obtain an updated estimate of the system state. The filter is designed to take into account any uncertainties or errors in measurements, as well as any noise or other sources of variability in the system.

The Kalman filter is particularly useful for analyzing time series data, as it can be used to estimate the underlying trend or pattern in the data, even if the data is noisy or contains outliers.

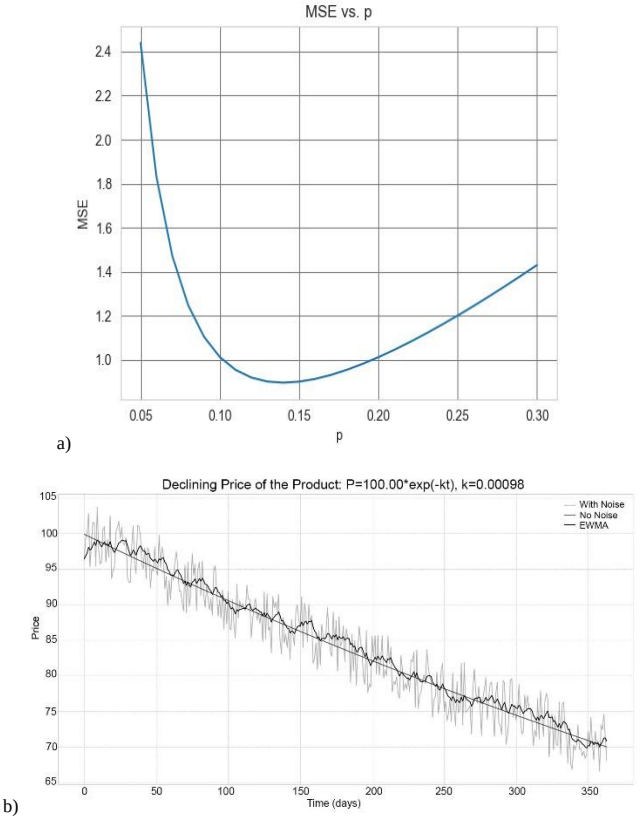


Fig. 5. a) Dependence of the root-mean-square error on the value of the weight coefficient; b) EWMA at the minimum value of the root-mean-square error (MSE): p = 0.14; MSE = 0.89

However, the Kalman filter can require significant computation, especially for large datasets or complex models. The Kalman filter relies on assumptions about the statistical properties of time series data, such as the distribution of noise, which may not always be accurate.

A graphical interpretation of the results of the algorithm using the Kalman filter is shown in Fig. 6.

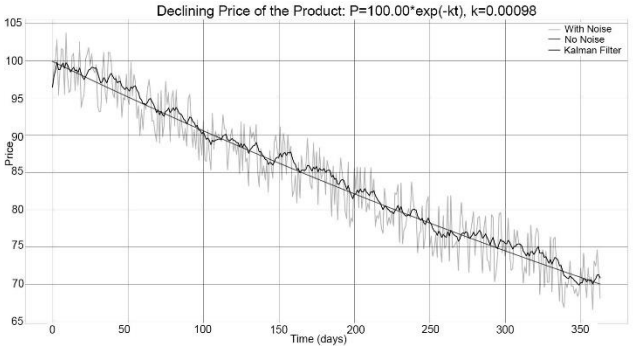


Fig. 6. The results of the Kalman filter at the minimum value of the root-mean-square error (MSE): a = 0.009, b = 0.418, c = 1.000, MSE = 0.8502

To build the dependencies, Python 3.10 scripts were written. For graphical interpretation of the results, the Matplotlib library was used, along with NumPy for numerical operations.

3. Conclusion

Thus, methods such as Simple Moving Average (SMA), Weighted Moving Average (WMA), Exponential Moving Average (EMA), Exponentially Weighted Moving Average (EWMA), and Kalman filter are widely used in time series analysis, signal processing, and financial analysis. Each method has its strengths and weaknesses, and the choice of method depends on the specific application and characteristics of the data. It is important for researchers and practitioners to understand the properties and limitations of these methods to make informed decisions when analyzing time series data.

References

[1] Brockwell P. J., Davis R. A.: Introduction to time series and forecasting. Springer, 2016.
[2] Gardner E. S.: Exponential smoothing: The state of the art. Journal of forecasting 4(1), 1985, 1–28.
[3] Hyndman R. J., Athanasopoulos G.: Forecasting: principles and practice. Otexts, 2018.
[4] Lewis C. D.: Industrial and business forecasting methods. Butterworth, Kent 1982.
[5] Ruppert D., Wand M. P., Carroll R. J.: Semiparametric regression. Cambridge University Press, 2018.
[6] Terry J. W., Parramore K.: Quantitative Methods for Finance. International Thomson Business Press, 1997.
[7] Wynn R. F., Holden K.: An introduction to applied econometric analysis. Macmillan, London 1974.

Ph.D. Volodymyr Lotysh
e-mail: auvp@lntu.edu.ua

Lutsk National Technical University, PhD. (technical). Research interests: open-source software applied for simulations of problems using distributed platforms. Author of nearly 80 publications in this research.

http://orcid.org/0000-0003-0899-8015

Ph.D. Larysa Gumeniuk
e-mail: l.gumeniuk@lntu.edu.ua

Lutsk National Technical University, PhD. (technical), Department of Automation and Computer-Integrated Technologies.
Research interests: modeling of reliability and safety of the automated control systems.
Has more than 60 publications in this area.

http://orcid.org/0000-0002-7678-7060

Ph.D. Pavlo Humeniuk
e-mail: l.gumeniuk@lntu.edu.ua

Lutsk National Technical University, PhD. (technical), Department of Automation and Computer-Integrated Technologies.
Research interests: programming, robotics.

http://orcid.org/0000-0002-6251-8548

http://doi.org/10.35784/iapgos.5320 received: 26.07.2023 | revised: 09.09.2023 | accepted: 28.09.2023 | available online: 30.09.2023

A STANDALONE DC MICROGRID ENERGY MANAGEMENT STRATEGY USING THE BATTERY STATE OF CHARGE

Elvin Yusubov, Lala Bekirova
Azerbaijan State Oil and Industry University, Department of Instrumentation Engineering, Baku, Azerbaijan

Abstract. *This article introduces an enhanced energy management strategy that employs the state of charge (SoC) of batteries in standalone DC microgrids with photovoltaic (PV) modules. Efficient energy management is crucial to ensure uninterrupted power supply to the load units in microgrids. To address the challenges posed by external factors such as temperature fluctuations and variations in solar irradiance, energy storage systems are deployed to compensate for the negative effects of the external factors on the output power of PV modules. The proposed approach takes into account various parameters of the microgrid elements, including the available power from the sources, demand power, and the SoC of batteries, in order to develop an efficient energy control mechanism with load-shedding capability. By considering these parameters, the strategy aims to optimize the utilization of available resources while ensuring a reliable power supply to the connected loads. The SoC of the batteries plays a critical role in determining optimal charging and discharging profiles, enabling effective energy management within the microgrid. To evaluate the effectiveness of the proposed approach, an algorithm is designed and simulations are conducted. The proposed algorithm utilizes a hybrid approach by combining power and SoC-based methods for efficient control. Through analysis of the simulation results, it is found that the presented approach is capable of delivering the intended load power while increasing the life cycle of the batteries with the pre-defined SoC levels.*

Keywords: DC microgrid, energy management strategy, battery state of charge, photovoltaic systems

STRATEGIA ZARZĄDZANIA ENERGIĄ SAMODZIELNEJ MIKROSIECI DC Z WYKORZYSTANIEM STANU NAŁADOWANIA BATERII

Streszczenie. *Niniejszy artykuł wprowadza ulepszoną strategię zarządzania energią, która wykorzystuje stan naładowania akumulatorów (SoC) w autonomicznych mikrosieciach prądu stałego z modułami fotowoltaicznymi (PV). Efektywne zarządzanie energią ma kluczowe znaczenie dla zapewnienia nieprzerwanego zasilania jednostek odbiorczych w mikrosieciach. Aby sprostać wyzwaniom związanym z czynnikami zewnętrznymi, takimi jak wahania temperatury i zmiany natężenia promieniowania słonecznego, systemy magazynowania energii są wdrażane w celu skompensowania negatywnego wpływu czynników zewnętrznych na moc wyjściową modułów fotowoltaicznych. Proponowane podejście uwzględnia różne parametry elementów mikrosieci, w tym dostępną moc ze źródeł, moc zapotrzebowania i SoC akumulatorów, w celu opracowania wydajnego mechanizmu kontroli energii z możliwością zrzucania obciążenia. Biorąc pod uwagę te parametry, strategia ma na celu optymalizację wykorzystania dostępnych zasobów przy jednoczesnym zapewnieniu niezawodnego zasilania podłączonych obciążeń. SoC akumulatorów odgrywa kluczową rolę w określaniu optymalnych profili ładowania i rozładowywania, umożliwiając efektywne zarządzanie energią w mikrosieci. Aby ocenić skuteczność proponowanego podejścia, zaprojektowano algorytm i przeprowadzono symulacje. Proponowany algorytm wykorzystuje podejście hybrydowe, łącząc metody oparte na mocy i SoC w celu zapewnienia wydajnej kontroli. Poprzez analizę wyników symulacji stwierdzono, że prezentowane podejście jest w stanie dostarczyć zamierzoną moc obciążenia, jednocześnie zwiększając cykl życia akumulatorów przy wstępnie zdefiniowanych poziomach SoC.*

Słowa kluczowe: mikrosieć prądu stałego, strategia zarządzania energią, stan baterii akumulatorów, systemy fotowoltaiczne

Introduction

Autonomous DC microgrids have emerged as a promising solution for decentralized and off-grid power systems, operating independently from the central grid [1, 10]. Energy storage systems, particularly batteries, play a crucial role in these microgrids by ensuring reliable power supply, managing renewable energy intermittency, and enhancing system stability. Effective control strategies for battery charging and discharging are essential to optimize energy management within autonomous DC microgrids, enabling efficient resource utilization and uninterrupted power delivery to connected loads [4, 5, 11, 20]. Precise control of battery operations is fundamental to maintaining system performance, prolonging battery life, and meeting load demand requirements.

Different control strategies are employed to manage the power flow in DC microgrids, each with their advantages, and challenges. These include voltage-based DC bus signalling, power-based, state of charge of battery-based and economic dispatch algorithms.

Voltage-based DC bus signalling control is a commonly used method that utilizes voltage thresholds to control battery charging or discharging [6, 12-16]. By monitoring DC bus voltage levels, this strategy ensures grid stability and allows the battery system to support peak load demand or compensate for insufficient renewable power generation. However, these approaches require a change in DC bus voltage which causes instability.

Power-based control strategies focus on adjusting battery charging and discharging rates based on power demand and availability [2, 7, 18, 21]. These techniques consider power flow within the microgrid and dynamically regulate battery operations to balance supply and demand, optimizing energy utilization and minimizing system losses. However, these methods require accurate power measurement and control devices, thus increasing system complexity. Some of them may not fully consider the battery's state of charge in power regulation.

Economic dispatch algorithms [8, 19] considering factors such as electricity prices and demand profiles, aim to minimize energy costs while meeting power requirements within the microgrid. These approaches optimize the scheduling of battery charging and discharging activities, considering the cost-effectiveness of grid power utilization versus stored energy. Economic dispatch algorithms in energy management systems have several disadvantages, which can impact their efficiency and effectiveness. Some of these disadvantages include complexity, computational time, and lack of consideration for environmental factors.

Fig.1. depicts the structure of the simple standalone DC microgrid.

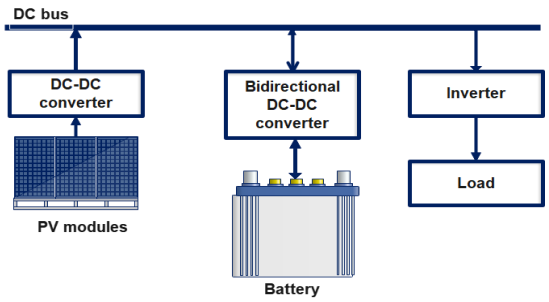


Fig. 1. Simple standalone DC microgrid structure

State-of-charge (SoC) control methods [9, 17] employ the battery's SoC as a critical parameter to determine optimal charging and discharging profiles. By monitoring and managing

SoC, these strategies ensure batteries operate within desired ranges, mitigating the risks of overcharging or over-discharging that could impair battery performance or reduce lifespan. These methods may cause instability in dynamic load and generation scenarios if the real-time power measurement of the units is not conducted. The SoC of the battery using the Coulomb counting method is as follows [3]:

$$SoC(t) = SoC(t_0) - \frac{1}{Q_{rated}} \int_{t_0}^{t_0+\tau} \eta I_b(t)\, dt \tag{1}$$

where: $SoC(t_0)$ is the initial SoC, I_b is battery current, η_b he efficiency of the battery (between 0 and 1).

The proposed algorithm utilizes a hybrid approach by combining power and SoC-based methods for efficient control.

1. DC-DC converter modelling

DC-DC converters play an essential role in interfacing PV modules and batteries with the load units in DC microgrids. ZETA topology is selected to stabilize the voltage received by the PV modules as it can operate in boost and buck modes.

1.1. ZETA converter

The electronic circuit diagram of the ZETA converter which contains two inductors (L1, L2), capacitors (C1, C2), a diode (D), and a transistor (Q), input voltage V_{in} is depicted in Fig. 2.

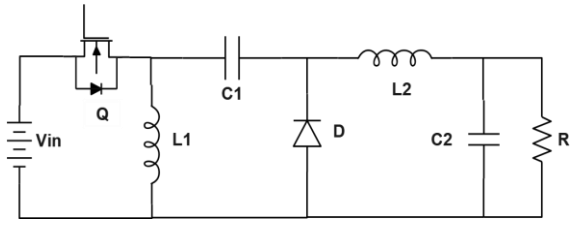


Fig. 2. ZETA circuit diagram

The converter functions in two distinct modes. During the on-mode, the MOSFET is activated, causing the diode, D, to enter its reverse-bias state, thus blocking the flow of electrical current. As a result, the voltage across inductor L1 matches the supply voltage, and there is a gradual rise in the inductor current over time. At the same time, capacitor C1 starts to charge and reaches the output voltage level. The following state-space representation describes the on state of the ZETA converter:

$$\begin{bmatrix} \frac{di_{L1}}{dt} \\ \frac{di_{L2}}{dt} \\ \frac{dv_{C1}}{dt} \\ \frac{dv_{C2}}{dt} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{L_2} & -\frac{1}{L_2} \\ 0 & -\frac{1}{C_1} & 0 & 0 \\ 0 & \frac{-1}{C_2} & 0 & -\frac{1}{C_2 R} \end{bmatrix} \begin{bmatrix} i_{L1} \\ i_{L2} \\ v_{C1} \\ v_{C2} \end{bmatrix} + \begin{bmatrix} \frac{1}{L_1} \\ \frac{1}{L_2} \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} v_{in} \end{bmatrix} \tag{2}$$

$$v_O = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} i_{L1} \\ i_{L2} \\ v_{C1} \\ v_{C2} \end{bmatrix} + \begin{bmatrix} 0 \end{bmatrix} \begin{bmatrix} v_{in} \end{bmatrix} \tag{3}$$

where i_{L1}, i_{L2} are inductor currents through L1 and L2, v_{C1}, v_{C2} are voltages across C1 and C2, respectively, v_{in}, v_o are respective the input and output voltages.

During the second mode, the MOSFET is turned off, causing a polarity change. As a result, the diode, D, switches to the forward-biasing mode, allowing electrical current to flow effortlessly through it. The current passing through the diode leads to the parallel connection of inductor L2 with the output capacitor. Additionally, capacitor C1 discharges its stored energy through inductor L1. The following state-space representation describes the off-state of the ZETA converter:

$$\begin{bmatrix} \frac{di_{L1}}{dt} \\ \frac{di_{L2}}{dt} \\ \frac{dv_{C1}}{dt} \\ \frac{dv_{C2}}{dt} \end{bmatrix} = \begin{bmatrix} 0 & 0 & -\frac{1}{L_1} & 0 \\ 0 & 0 & 0 & -\frac{1}{L_2} \\ \frac{1}{C_1} & 0 & 0 & 0 \\ 0 & \frac{1}{C_2} & 0 & -\frac{1}{C_2 R} \end{bmatrix} \begin{bmatrix} i_{L1} \\ i_{L2} \\ v_{C1} \\ v_{C2} \end{bmatrix} + \begin{bmatrix} \frac{1}{L_1} \\ \frac{1}{L_2} \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} v_{in} \end{bmatrix} \tag{4}$$

The state-space average model is as follows:

$$\begin{bmatrix} \frac{di_{L1}}{dt} \\ \frac{di_{L2}}{dt} \\ \frac{dv_{C1}}{dt} \\ \frac{dv_{C2}}{dt} \end{bmatrix} = \begin{bmatrix} 0 & 0 & -\frac{1-d}{L_1} & 0 \\ 0 & 0 & \frac{d}{L_2} & -\frac{1}{L_2} \\ \frac{1-d}{C_1} & \frac{-d}{C_1} & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{C_2 R} \end{bmatrix} \begin{bmatrix} i_{L1} \\ i_{L2} \\ v_{C1} \\ v_{C2} \end{bmatrix} + \begin{bmatrix} \frac{d}{L_1} \\ \frac{d}{L_2} \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} v_{in} \end{bmatrix} \tag{5}$$

1.2. Bidirectional converter

A power electronics interface is employed to establish a secure and effective link between the battery pack and the DC bus, ensuring safety and efficiency. This interface incorporates bidirectional converters, facilitating energy flow in two directions - from the battery pack to the DC bus and vice versa. In modelling DC microgrids, the bidirectional SEPIC-ZETA converter is frequently utilized due to its ability to charge or discharge the batteries without necessitating a modification in the DC bus voltage.

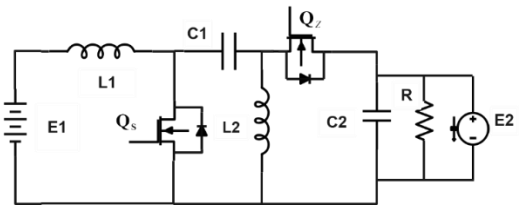


Fig. 3. The circuit diagram of the bidirectional SEPIC-ZETA converter

The bidirectional SEPIC-ZETA converter, illustrated in Fig. 3, comprises the following elements:

- The battery voltage (E1) and DC bus voltage (E2);
- Capacitors (C1 and C2);
- Inductors (L1 and L2);
- Transistors (T1 and T2);
- Load resistance (R).

2. Proposed control algorithm

The proposed algorithm functions through a series of pre-defined steps, taking into account the measured values of PV (Photovoltaic), battery, and demand power, as well as the State of Charge (SoC) of the batteries. These inputs serve as the foundation for the algorithm's decision-making process to efficiently manage the energy system.

To begin, the algorithm compares the PV power output with the load power demand. In cases where the PV power exceeds the load power, the surplus energy is utilized to meet the required load power, eliminating the need for any external source. The excess power is then harnessed to charge the batteries, ensuring optimal utilization of resources. This charging process continues until the SoC of the batteries reaches above 90%, at which point charging is automatically halted to prevent

overcharging. By maintaining the SoC within the safe threshold, the algorithm enhances the batteries' life cycle, promoting their longevity and effectiveness.

On the contrary, when the PV power falls short of the load power demand, a different strategy comes into play. The algorithm then evaluates the SoC of the batteries and compares it with the lower SoC limit, which is set at 25%. If the SoC is above this limit, the batteries are brought into action to assist in delivering the required power to the load. The battery discharging process commences, ensuring that the load's power needs are met until the SoC falls below the 25% threshold. By strategically employing the batteries during such scenarios, the algorithm optimizes energy usage, making the most of available resources.

In certain situations where the total combined power from both PV and batteries is unable to meet the load power demand, load shedding becomes necessary. Load shedding is a controlled process where non-essential loads are intentionally disconnected or reduced to ensure that critical loads receive the required power. This mechanism allows the algorithm to prioritize essential loads, ensuring uninterrupted power supply to vital systems.

To perform load shedding, an objective function is proposed with minimal energy loss. The presented minimization function is:

$$\min \left| \boldsymbol{P}_{gen.} - \boldsymbol{b}_1 \boldsymbol{P}_{LD1} - \boldsymbol{b}_2 \boldsymbol{P}_{LD2} - \boldsymbol{b}_3 \boldsymbol{P}_{LD3} - \boldsymbol{b}_n \boldsymbol{P}_{LDn} \right| \quad (6)$$

when the following condition is met:

$$P_{gen} > \sum_{i=1}^{N} b_i P_{LD.i} \, (for \; b_i = 1) \quad (7)$$

where $\boldsymbol{P}_{gen.}$ is the power received from batteries and photovoltaic, $\boldsymbol{P}_{LD1}$ are the required power for the controllable loads, $\boldsymbol{b}_i$ is the parameter which has binary values. 0 and 1 demonstrate shed and active loads, respectively.

During the shedding process, the random binary numbers are assigned to the parameters until the load with minimum energy loss is shed. Each parameter representing a specific load is assigned a binary value, where "1" signifies the load is retained, and "0" denotes the load will be shed. This binary assignment allows for a systematic evaluation of all loads, ensuring a fair and unbiased decision-making process. The algorithm starts by generating these random binary numbers and associating them with each load parameter. It then proceeds to assess the impact of shedding individual loads on the overall energy loss. By selectively shedding different loads, the algorithm can analyze the resulting changes in the energy consumption pattern and determine which load's removal minimizes the energy loss to the greatest extent.

Pseudo-code for the proposed algorithm

If $P_{pv} > P_{load}$

Supply the load using only PV modules

Charge the batteries until their SoC>90%

else

Check the battery SoC

If SoC>25%

If $P_{pv} + P_{bat.} > P_{load}$

Supply the load with PV and battery power until SoC<20%

else

Perform load shedding

3. Simulation results

When the total PV power exceeds the load power, only PV modules supply the load with energy, eliminating the aid of the battery storage units. This scenario is depicted in Fig. 4, where the cumulative PV power generation amounts to 55 W, whereas the cumulative power consumption by the load is 40 W. In this situation, the battery units commence charging until their State of Charge (SoC) reaches 90% to stop overcharging. If the SoC surpasses the pre-determined threshold, the battery units cease charging. The relationship between the power source and the load is represented using binary digits, wherein "1" denotes a connection, and "0" signifies a disconnection. Concerning the battery conditions, "1" represents the charging state, "0" corresponds to the idle mode, and "-1" signifies the discharging state.

In the event that the total sum of PV power generation is less than the total sum of power consumed by the load during the same period, the PV module output power falls short of meeting the power demand, necessitating the incorporation of additional batteries. The battery's State of Charge (SoC) is checked to determine whether it exceeds 25%. If this condition holds true, the following condition is evaluated. If this condition is also met, the batteries are employed to provide the load with additional energy, and the discharging process commences (state -1) until the SoC falls under 25%. This entire process is illustrated in Fig. 5, where PV amounts to 35W, battery power is 80 W, and the battery's SoC is at 75%, while the cumulative load power is 70 W.

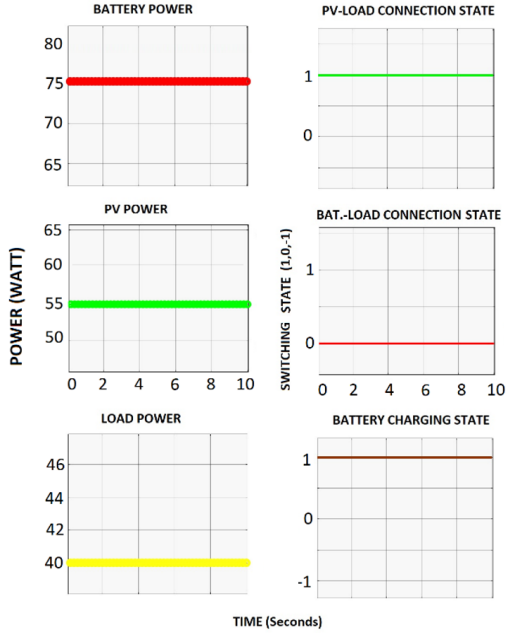


Fig. 4. Load-power source connection and battery charging-discharging states when PV power is higher than the load power

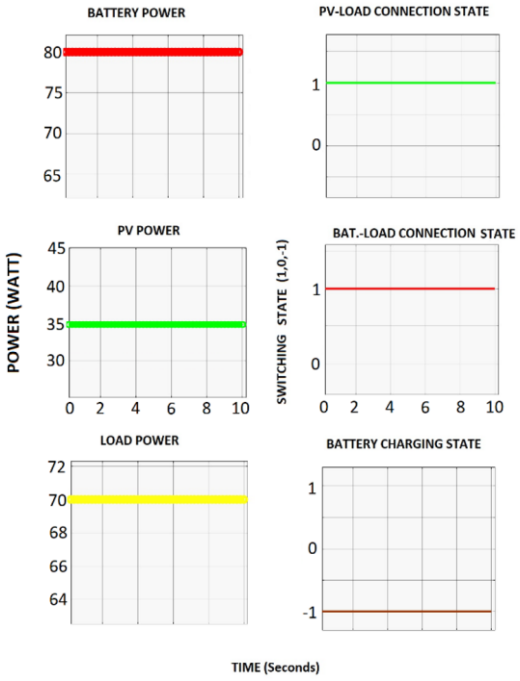


Fig. 5. Load- power source connection and battery charging-discharging states when PV power is lower than the load power

4. Conclusion

In summary, an enhanced energy management strategy for standalone DC microgrids with PV modules, using battery SoC is introduced in this research. A hybrid energy management algorithm with load shedding ability is constructed. Objective functions are built and simulations are carried out. Simulation results confirm its effectiveness, offering a promising solution for sustainable microgrid energy management.

References

[1] Ashok K. A., Amutha N. P.: A Comprehensive Review of DC Microgrid in Market Segments and Control Technique. Heliyon, 2022, e11694 [http://doi.org/10.1016/j.heliyon.2022.e11694].

[2] Caldognetto T. et al.: Power-Based Control of Low-Voltage Microgrids. IEEE Journal of Emerging and Selected Topics in Power Electronics 3(4), 2015, 1056–1066 [http://doi.org/10.1109/jestpe.2015.2413361].

[3] Chang W. Y.: The State of Charge Estimating Methods for Battery: A Review. ISRN Applied Mathematics 2013, 1–7 [http://doi.org/10.1155/2013/953792].

[4] Darwish M. et al.: Review of Battery Management Systems. IEEE International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), 2021 [http://doi.org/10.1109/iceccme52200.2021.9590884].

[5] Fanoro M. et al.: A Review of the Impact of Battery Degradation on Energy Management Systems With a Special Emphasis on Electric Vehicles. Energies 15(16), 2022, 5889 [http://doi.org/10.3390/en15165889].

[6] Garg A. et al.: Implementation of Energy Management Scenarios in a DC Microgrid Using DC Bus Signaling. IEEE International Conference on Power Electronics, Smart Grid and Renewable Energy (PESGRE), 2020 [http://doi.org/10.1109/pesgre45664.2020.9070484].

[7] Giroletti F., Scattolini A.: A Hybrid Frequency/Power Based Method for Industrial Load Shedding. International Journal of Electrical Power & Energy Systems 35(1), 2012, 194–200 [http://doi.org/10.1016/j.ijepes.2011.10.013].

[8] Ibrahim M. S. et al.: Economic Energy Dispatch of Microgrids Using Computational Intelligence. Renewable Energy Microgeneration Systems, Elsevier, 2021, 229–245 [http://doi.org/10.1016/b978-0-12-821726-9.00010-2].

[9] Jayasena K. N. C. et al.: SoC Based Multi-Mode Battery Energy Management System for DC Microgrids. IEEE Moratuwa Engineering Research Conference (MERCon), 2019 [http://doi.org/10.1109/mercon.2019.8818765].

[10] Kandari R. et al.: Chapter 4 – DC microgrid. Microgrids Modeling, Control, and Applications. Elsevier, 2022, 91–139 [http://doi.org/10.1016/b978-0-323-85463-4.00007-1].

[11] Lawder M. T. et al.: Battery Energy Storage System (BESS) and Battery Management System (BMS) for Grid-Scale Applications. Proceedings of the IEEE 102(6), 2014, 1014–1030 [http://doi.org/10.1109/jproc.2014.2317451].

[12] Li F. et al.: Active DC bus signalling control method for coordinating multiple energy storage devices in DC microgrid. IEEE Second International Conference on DC Microgrids (ICDCM), 2017 [http://doi.org/10.1109/icdcm.2017.8001048].

[13] Li X. et al.: DC Bus Signaling-Based Coordination Through Masked Sensing for DC Microgrid. 9th International Power Electronics and Motion Control Conference (IPEMC2020-ECCE Asia), IEEE, 2020 [http://doi.org/10.1109/ipemc-ecceasia48364.2020.9368065].

[14] Li X. et al.: A novel control strategy based on DC bus signalling for DC micro-grid with photovoltaic and battery energy storage. China International Conference on Electricity Distribution (CICED), IEEE, 2016 [http://doi.org/10.1109/ciced.2016.7576132].

[15] Liu S. et al.: A Multivariable Hysteresis-Based DC Bus Signaling Control for DC Microgrid With Enhanced Reliability. Third International Conference on DC Microgrids (ICDCM), IEEE, 2019 [http://doi.org/10.1109/icdcm45535.2019.9232907].

[16] Ma T. et al.: A Control Strategy of DC Building Microgrid Connected to the Neighborhood and AC Power Network. Buildings 7(4), 2017, 42 [http://doi.org/10.3390/buildings7020042].

[17] Mitra S. K., Srinivas B. K.: A SOC Based Adaptive Energy Management System for Hybrid Energy Storage System Integration to Grid. 12th Energy Conversion Congress & Exposition – Asia (ECCE-Asia), IEEE, 2021 [http://doi.org/10.1109/ecce-asia49820.2021.9479107].

[18] Tenti P., Caldognetto T.: Chapter 4 – Master/Slave Power-Based Control of Low-Voltage Microgrids. Microgrid Advanced Control Methods and Renewable Energy System Integration, Elsevier, 2017, 101–135 [http://doi.org/10.1016/b978-0-08-101753-1.00004-8].

[19] Wu H. et al.: Dynamic Economic Dispatch of a Microgrid: Mathematical Models and Solution Algorithm. International Journal of Electrical Power & Energy Systems 63, 2014, 336–346 [http://doi.org/10.1016/j.ijepes.2014.06.002].

[20] Yang Y. et al.: Modelling and Optimal Energy Management for Battery Energy Storage Systems in Renewable Energy Systems: A Review. Renewable and Sustainable Energy Reviews 167, 2022, 112671 [http://doi.org/10.1016/j.rser.2022.112671].

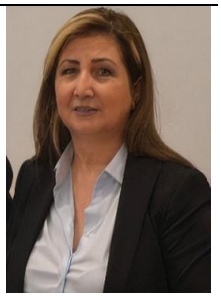
[21] Yusubov E., Bekirova L.: Adaptive Metaheuristic Moth-Flame Optimized Droop Control Method for DC Microgrids. International Conference Automatics and Informatics (ICAI), IEEE, 2022 [http://doi.org/10.1109/icai55857.2022.9960119].

M.Sc. Elvin Yusubov
e-mail: elvinyusifov05@gmail.com

Elvin Yusubov, born in Azerbaijan on December 29, 1993, obtained his BSc degree in Electronics from the National Aviation Academy of Azerbaijan in 2015. He further earned his M.Sc. degree in Computerized Information Measurement and Control from Azerbaijan State Oil and Industry University in 2019. Elvin's research interests encompass power electronics and metaheuristic controllers. His academic contributions consist of more than 10 scientific articles.

http://orcid.org/0000-0001-6199-9266

D.Sc. Lala Bekirova
e-mail: lala.bakirova@asoiu.edu.az



Professor Bekirova Lala Rustam is the Head of the Department of Instrumentation Engineering at Azerbaijan State Oil and Industrial University. Bekirova has authored over 115 articles, patents, and monographs. Her research focuses on various areas, including the physical and technical foundations of information and measurement systems, remote measurements and aerospace research, and optical engineering.

http://orcid.org/0000-0003-0584-7916

http://doi.org/10.35784/iapgos.4457 *received: 30.08.2023 | revised: 09.09.2023 | accepted: 26.09.2023 | available online: 30.09.2023*

MACROMODELING OF LOCAL POWER SUPPLY SYSTEM BALANCE FORECASTING USING FRACTAL PROPERTIES OF LOAD AND GENERATION SCHEDULES

Daniyar Jarykbassov[1], Petr Lezhniuk[2], Iryna Hunko[3], Vladyslav Lysyi[2], Lubov Dobrovolska[4]

[1]Academy of Logistics and Transport, Almaty, Kazakhstan, [2]Institute of Renewable Energy of the National Academy of Sciences of Ukraine, Kyiv, Ukraine, [3]Vinnytsia National Technical University, Vinnytsia, Ukraine, [4]Lutsk National Technical University, Lutsk, Ukraine

Abstract*. A method of forecasting the balance of electricity consumption of urban development objects, civil purposes using discrete macromodels is proposed. We consider the power supply system (PSS) of the district, which is characterised by power supply from general-purpose power grids, as well as having its own generation of electricity from renewable energy sources (RES). Such a local electric power system (LES) under certain conditions can be operated as an independent balanced electrical facility. For optimal operation of the LES under these conditions, it is necessary to predict its power consumption schedules. The proposed macromodelling method allows to develop deterministic models of power consumption with the required accuracy on the basis of retrospective information without the use of data preprocessing procedures. The solution to the problem of forecasting electricity consumption schedules is simplified by using only basic or deterministic characteristics in the construction of the model. These include fractal properties of PSS load schedules*

Keywords: power supply system, urban development, power consumption forecasting, macromodelling, fractal properties

MAKROMODELOWANIE PROGNOZOWANIA BILANSU LOKALNEGO SYSTEMU ELEKTROENERGETYCZNEGO Z WYKORZYSTANIEM FRAKTALNYCH WŁAŚCIWOŚCI PLANÓW OBCIĄŻENIA I GENERACJI

Streszczenie*. Zaproponowano metodę prognozowania bilansu zużycia energii elektrycznej przez obiekty miejskie o przeznaczeniu cywilnym z wykorzystaniem makromodeli dyskretnych. Rozważany jest system zasilania (SZ) dzielnicy, który charakteryzuje się zasilaniem z sieci ogólnego przeznaczenia, a także posiada własną produkcję energii elektrycznej z odnawialnych źródeł energii (OZE). Taki lokalny system elektroenergetyczny (LES) w pewnych warunkach może być eksploatowany jako niezależny, zrównoważony obiekt elektryczny. Aby zapewnić optymalne działanie LES w takich warunkach, konieczne jest przewidywanie jego harmonogramów poboru mocy. Proponowana metoda makromodelowania pozwala na opracowanie deterministycznych modeli poboru mocy z wymaganą dokładnością na podstawie informacji retrospektywnych bez stosowania procedur wstępnego przetwarzania danych. Rozwiązanie problemu prognozowania planów zużycia energii elektrycznej jest uproszczone poprzez wykorzystanie w konstrukcji modelu jedynie podstawowych lub deterministycznych charakterystyk. Obejmują one fraktalne właściwości harmonogramów obciążenia SZ.*

Słowa kluczowe: system zasilania, rozwój miast, prognozowanie zużycia energii, makromodelowanie, właściwości fraktalne

Introduction

Macromodelling [1, 4, 8] is used in the evaluation of complex systems, which include power supply systems (PSS). Such SES models are distinguished by the fact that they describe only the external characteristics of the modelled system and are intended for its general evaluation. Problem solving is simplified without losing adequacy due to the use of only the main or defining characteristics in the construction of the model [9]. Such characteristics that characterise the SES load schedules include their fractal properties [2, 5, 7].

That is, self-affine structures – fractals are used, each part of which repeats in its development the development of the whole model as a whole. Such models are useful for obtaining preliminary estimates of the system, for example, in pre-design decisions. However, the problem is that the forecasting of powerful power systems, industrial enterprises does not allow to take into account the composition, specificity and technological features of the mode of operation of electrical consumers and power consumption of urban buildings.

Therefore, the aim of the article is to develop an effective, sufficiently accurate for practical application forecasting method using fractal properties for SES of urban buildings.

1. Highlighting essential properties of SES in the macromodelling method

Residential electrical loads are random, depending on the agenda of the occupants and the availability of a set of electrical appliances (EA). They vary significantly during the day and depending on the season. In addition, localised SESs have to take into account the instability of RES generation that is part of their electricity supply. All this creates difficulties in forecasting the power consumption schedules of SES. Thus, the determination of forecasted electrical loads is the basis for the design of internal and urban grids [1, 6].

The main group of consumers in residential areas are residential buildings. The electrical load of houses is determined by the lighting of flats and various household appliances used by the population. In practice, the value of electricity consumption is influenced by the most powerful electrical appliances of everyday use. These include air conditioners, washing machines, dishwashers, electric heaters such as boilers, electric cookers, heating systems, etc.

Fig. 1 shows the structural diagram of the models designed to predict the electricity consumption of civilian objects of urban development.

1. Autoregressive models are time series models in which the values of a time series at a given moment depend linearly on previous values of the same series. An autoregressive process of order p is defined as

$$X_t = c + \sum_{i=1}^{p} a_i X_{t-i} + \varepsilon_t \tag{1}$$

where a_i − model parameters (autoregressive coefficients); c – constant (often equal to zero for simplification); ε_t – white noise.

Autoregressive models can be used to model seasonality, in which case the number of model coefficients will correspond to the number of cyclically changing factors that are taken into account.

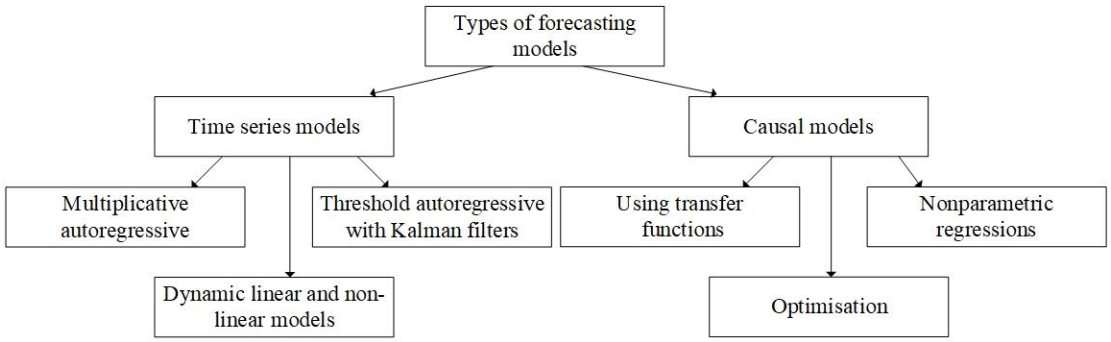


Fig. 1. Types of electricity consumption forecasting models

In order to forecast electricity consumption, the following form of autoregressive model can be applied:

$$L(t,d) = \sum_{k=1}^{4} a_k L_k \left(t,d\right) \tag{2}$$

where a_k – linear weights that provide the optimal combination of the four individual predictions; L_1 (t, d) – forecast $L(t,d)$ based on a first-order autoregressive model with a delay of one hour; $L_2(t,d)$, $L_3(t,d)$, $L_4(t,d)$ – the same with a delay of 24 hours, a week and a year respectively.

2. Generalised exponential smoothing that can be applied to predict total hourly electricity consumption:

$$L(t) = a^T f(t) + \varepsilon(t) \tag{3}$$

where a^{T} – the transposed vector of exponentially smoothed weights; $f(t)$ – vector of smoothing functions.

3. Neural networks and fuzzy logic is one of the new approaches used to solve the problem of forecasting based on fuzzy logic and neural networks. The method involves the use of a priori information, allows the use of new information in the construction process and takes into account the properties of the modelled process. It can also use previously known information, subjected to training and sufficiently visual for the observer. Neural networks are able to identify complex dependencies between input and output data and perform generalisation of existing but hidden properties and relationships. This is where the ability of a trained neural network to predict follows, to anticipate the future value of a particular sequence based on several previous values or currently existing factors.

This way of applying neural networks for forecasting will not lose performance under the condition of incomplete input information, but requires considerable time for training. Application of systems with fuzzy logic, i.e. sets with a set of elements of arbitrary nature, which do not clearly define belonging to a certain set, allows to eliminate the disadvantages of artificial neural networks [3].

Modelling with application of fuzzy sets is expedient in case of research of too complex technical system or process.

Modern mathematical statistics, long claimed to be the main tool for data analysis, is not always suitable for solving problems from real diverse life. This happens because averaged sample characteristics are used, which often turn out to be fictitious values. Therefore, the methods of mathematical statistics are useful for testing pre-formulated hypotheses.

2. Building a macromodel of forecasting

The essence of alternative forecasting based on macromodelling is the process of building a model of electricity consumption at such stages shown in Fig. 2.

The first stage of the analysis procedure involves the collection and processing of data on the electrical consumption of the real object under study. The system of organisation of collection and registration of information on electric load of the object was carried out with the help of the automated system of commercial electricity metering (ASCEM). Electricity metering devices of residential buildings, public buildings and their complexes are recommended to be united in the ASCEM, which promotes its use as a tool for energy audit and control of energy resources saving issues through the introduction of appropriate measures [10]. The use of ASCEM promotes the economical use of energy resources through careful control of electricity consumption and development of appropriate measures.

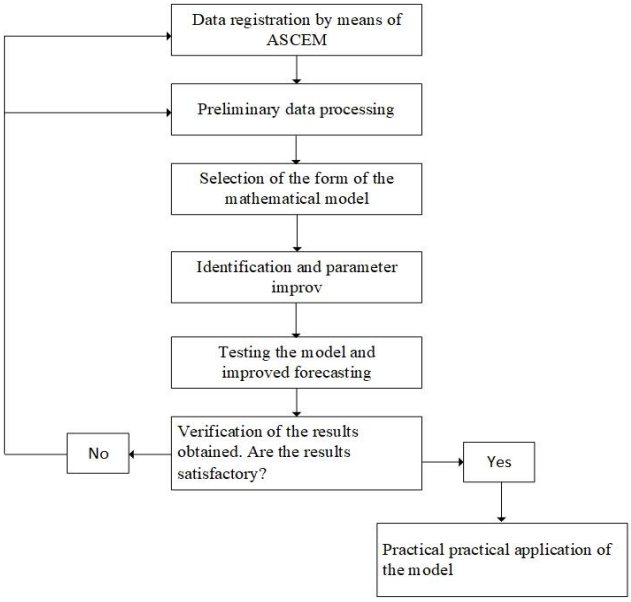


Fig. 2. Stages of building a macro model of forecasting

The collected information is displayed as continuous graphs (Fig. 3) or digitally. Based on the information obtained, the maximum electrical load is determined (on the graphs of Fig. 3 it is 95 kW). Graphs of electric load have a probabilistic character and change during the day, have maximums in the morning from 7 to 9 and evening from 19 to 23 hours, which is due to the modes of operation of a variety of electrical consumers.

Preliminary data processing involves selection of the time period for which forecasting is carried out and the period of further data verification. Filtering is carried out when there is a significant amount of information on electricity consumption in order to select the form of the model and simplify the forecasting procedure.

The choice of model form may involve mathematical macromodelling using discrete autonomous macromodels in the form of a „black box". The process is derived from the recorded characteristics of electricity consumption by using homogeneous differential or difference equations of state in form (4).

$$\begin{cases} \dfrac{d\vec{t}}{dt} = \vec{f}(\vec{x}) \\ \vec{y} = \vec{g}(\vec{x}) \end{cases} \tag{4}$$

$$\begin{cases} \vec{x}^{(k+1)} = \vec{f}(\vec{x}^{(k)}) \\ \vec{y}^{(k+1)} = \vec{g}(\vec{x}^{(k)}) \end{cases} \tag{5}$$

where $\vec{x}$ is a vector of state variables; $\vec{y}$ is a vector of output variables; $\vec{f}(\cdot), \vec{g}(\cdot)$ are vectors functions selected using optimisation algorithms.

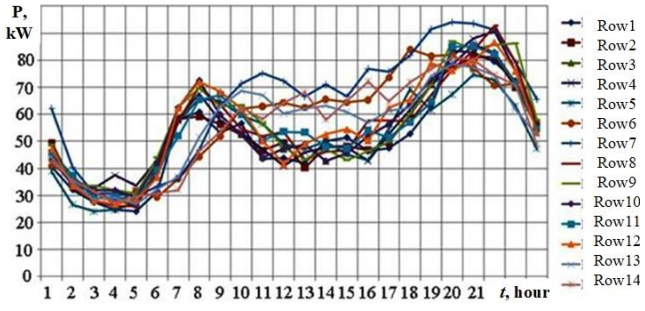


Fig. 3. Example of daily electric load graphs of a residential building (day of the month coincides with the row number)

Since there is no explicit vector of input variables when building the electricity consumption model, we consider the case when the initial value of the state variables of the modelled object is zero. Let's choose the form of macromodel description in the following form:

$$\begin{cases} \vec{x}(k+1) = F\,\vec{x}(k) + \Phi\left(\vec{x}(k), \vec{v}(k)\right) \\ \vec{y}(k+1) = C\left(\vec{x}(k+1)\right) \end{cases} \tag{6}$$

The initial state of the modelled object is described by the zero discrete vector of state variables $\vec{x}^{(0)}$. Therefore, the components of this vector must be added to the set of unknown model coefficients. However, $\vec{x}^{(0)}$ one cannot simply enter in the set of model parameters, as each dynamic process will have its own independent value of $\vec{x}^{(0)}$. To take this fact into account, we need to divide the vector of unknown coefficients $\vec{\lambda}$ into two parts: the first, which includes coefficients that are the same for all processes, and the second with an independent set of elements of the vector $\vec{x}^{(0)}$ for each process, which increases the unknown number of coefficients and complicates the optimisation problem.

When using the proposed macromodel, a problem arises in determining the zero discretisation of the vector $\vec{x}$, since the components of this vector, as a rule, are not directly measured experimentally, but are determined through certain values of the initial elements $\vec{y}$. In general, this means that it is necessary to find additionally a linear or non-linear dependence $\vec{x}^{(0)}$ on the experimentally measured values of $\vec{y}$.

In particular, in forecasting problems this dependence is constructed as a function of several first discretes of initial values:

$$\vec{x}^{(0)} = \vec{f}\left(\vec{y}^{(1)}, \vec{y}^{(2)}, \ldots, \vec{y}^{(l)}\right) \tag{7}$$

where l – number of samples used to find the zero sample of the vector $\vec{x}$.

The optimisation approach due to its universality with respect to the form of representation of macromodels can be used to find additional dependencies. This means that the elements of the vector $\vec{x}^{(0)}$, added to the unknown coefficients, $\vec{\lambda}$ should be replaced by the coefficients of expression (3), i.e., actually introduce this expression into the model itself.

If the model looks like (7), we get:

$$\begin{cases} \vec{x}^{(k+1)} = F\,\vec{x}^{(k+1)} + \Phi(\vec{x}^{(k)}) \\ \vec{y}^{(k+1)} = C\,\vec{x}^{(k+1)} \\ \vec{x}^{(0)} = \vec{f}(\vec{y}^{(1)}, \vec{y}^{(2)}, \ldots, \vec{y}^{(l)}) \end{cases} \tag{8}$$

which will contribute to the validation of the macro-model of electricity consumption.

In order to verify the performance of the proposed approach, a macro model of the daily electricity consumption of a 9-storey 216-apartment residential building is constructed, using input data on average daily electricity consumption and weeks (Fig. 4 and 5).

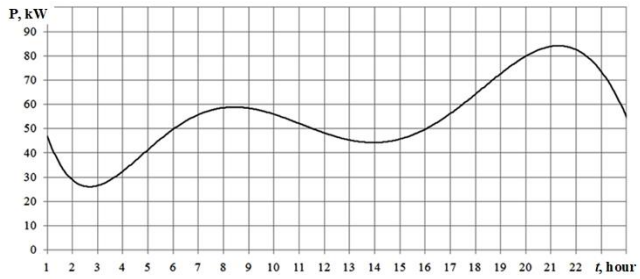


Fig. 4. Example of daily electric load graphs of a residential building (day of the month coincides with the row number)

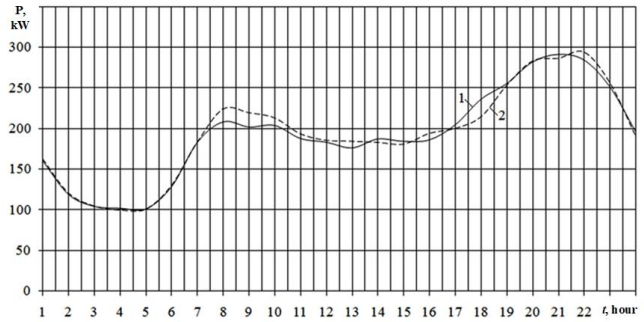


Fig. 5. Example of daily electric load graphs of a residential building (day of the month coincides with the row number)

A discrete macromodel of the electrical consumption of a residential building is created in the form of such a ratio:

$$\begin{cases} \vec{x}^{(i+1)} = F\,\vec{x}^{(i)} \\ \vec{y}^{(i+1)} = C\,\vec{x}^{(i+1)} \end{cases} \tag{9}$$

The input data of daily electricity consumption during the first and second weeks were used to build the macromodel. The initial value of the state variables was determined based on a linear dependence on the level of electricity consumption at randomly selected times, namely at 9, 20 and 22 hours:

$$\vec{x}^{(0)} = S\begin{pmatrix} y_9 \\ y_{20} \\ y_{22} \end{pmatrix} \tag{10}$$

Verification of the stand-alone macromodel was performed on an independent data set. For this purpose, the input data of electricity consumption of the residential building for the third week and the created macromodel of electricity consumption for the third week were used (Fig. 6).

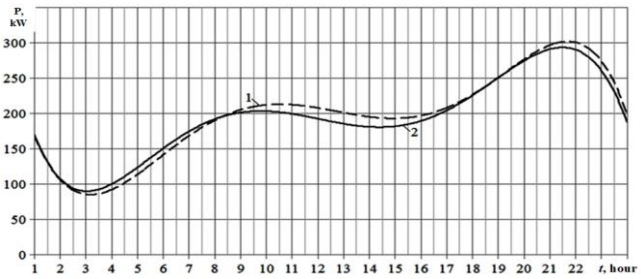


Fig. 6. Input data (curve 1) and macromodel response for week 3 (curve 2)

Comparison of the mean square error of forecasting by the regression model for forecasting electricity consumption of a residential building and by using a neural network showed the following. According to the retrospective information of residential building electricity consumption, the regression model for forecasting is as follows:

$$w = 55t + 4562 \tag{11}$$

The mean-square error of electricity consumption forecasting by the regression model was 6.0%.

Comparing the actual value of electricity consumption of the residential building for the 3rd week of October with its macro-modelling value, the RMS error of forecasting is about 3.1%. Thus, forecasting by macro-modelling better approximates the adequacy of the process than by regression model.

3. Conclusions

The proposed forecasting method allows to develop with sufficient accuracy deterministic models of power consumption on the basis of retrospective data without using data preprocessing procedures, which is typical for other methods. Forecasting of LES power consumption schedules is simplified by using only basic and deterministic characteristics in the construction of the macromodel. These include fractal properties of PSS load schedules.

Using the example of residential building power consumption forecasting, it is shown that the autonomous macromodelling approach, effectively using a priori information, is able to learn and, at the output, provide clear and adequate information about the LES power consumption process. Using the ASCEM information, it is possible to obtain a number of illustrative daily electric load schedules and generalise them for the future. Due to the use of fractal properties of LES load graphs in macromodelling, the accuracy of electricity consumption forecasting is improved and the term of satisfactory forecast increases.

References

[1] Ahmad N. et al.: Load Forecasting Techniques for Power System: Research Challenges and Survey. IEEE Access 10, 2022, 71054–71090 [http://doi.org/10.1109/ACCESS.2022.3187839].
[2] Feder E.: Fractals. Mir, Moscow 1991.
[3] Kaur D. et al.: Energy forecasting in smart grid systems: recent advancements in probabilistic deep learning. IET Gener. Transm. Distrib. 16, 2022, 4461–4479 [http://doi.org/10.1049/gtd2.12603].
[4] Lezhniuk P. et al.: Natural-simulation Model of Photovoltaic Station Generation in Process of Electricity Balancing in Electrical Power System. Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska 12(3), 2022, 40–45 [http://doi.org/10.35784/iapgos.3030].
[5] Lezhniuk P. D., Bondarchuk A. S., Shullie Iu. A.: Fractal analysis of dynamics of electric loading of civil objects. Sciences of Europe 27-1(27), 2018, 46–54.
[6] Liu H. et al.: A Power Load Forecasting Method Based on Intelligent Data Analysis. Electronics 12(16), 2023, 3441.
[7] Peters E.: Fractal Analysis of Financial Markets: Application of Chaos Theory in Investments and Economics. Internet-trading, Moscow 2004.
[8] Stakhiv P. G., Kozak Y. Y., Gogolyuk O. P.: Discrete macromodelling in electrical engineering and related fields. Publishing House of Lviv Polytechnic, Lviv 2014.
[9] Venikov V. A.: Theory of similarity and modeling. Higher School, Moscow 1976.
[10] Zhelezko Y. S.: Power losses. Reactive power. Electricity quality: Guide for practical calculations. ENPAS, Moscow 2009.

M.Sc. Daniyar Jarykbassov
e-mail: daniyarjarykbassov@yandex.kz

M.Sc., graduated from NJSC "Almaty University of Energy and Communications named after Gumarbek Daukeev", Faculty of Power Engineering Works at “Alatau Zharyk Kompaniyasy JSC”. Engaged in the transmission and distribution of electricity in Almaty and Almaty region.
He has published 6 scientific papers. The main direction of his scientific activity is automation of optimal control of electric power systems modes.

http://orcid.org/0000-0002-0051-5560

D.Sc. Eng. Petr Lezhniuk
e-mail: lezhpd@gmail.com

Professor of the Department of Electric Stations and Systems, Vinnytsia National Technical University. Research interests include application of similarity theory and modeling for optimum control of electro power systems modes. He has published 600 scientific papers, developed a number of basically new devices and software complexes for computers. The main directions of his scientific activity are mathematical modelling and optimization of electric power systems modes.

http://orcid.org/10000-0002-9366-3553

Ph.D. Iryna Hunko
e-mail: iryna_hunko@ukr.net

Ph.D., associate professor, Department of Power Plants and Systems, Vinnytsia National Technical University. She has published 75 scientific papers. The main direction of his scientific activity is optimal control of electric power systems with renewable energy resources.

http://orcid.org/0000-0003-2868-4056

Vladyslav Lysyi
e-mail: vladyslavlysyi@gmail.com

Postgraduate student at the Institute of Renewable Energy, dissertation topic: "Improving the energy efficiency of renewable energy sources in the electricity balance of power systems".

http://orcid.org/0009-0007-0211-9100

Ph.D. Lyubov Dobrovolska
e-mail: lsobchuk@gmail.com

Ph.D., professor, Department of Electrical Engineering Lutsk National Technical University. Total 180 publications including 90 publications in scientific journals and in the conference scripts.
The main directions of her scientific activity is optimal control of electric power systems with renewable energy resources.

http://orcid.org/0000-0001-8175-7635

http://doi.org/10.35784/iapgos.3756 received: 18.06.2023 | revised: 09.09.2023 | accepted: 27.09.2023 | available online: 30.06.2023

PV PANEL COOLING USING STACK EFFECT

Kudithi Nageswara Rao, Ganesamoorthy Rajkumar
Hindustan Institute of Technology and Science, Department of Electrical and Electronics Engineering, Chennai, India

Abstract. *Unsatisfactory feat of the photo-voltaic cell is solitary in every of the foremost issues among the promotion of PV technology. A vital strand changing cell performance is operative temperature of the cell. The voltage of the cell, declines near directly with a rise in the operative temperature of the photo-voltaic cell. The temperature condition on the in-operative surface of a photo-voltaic panel remains usually 22–30°C beyond the normal temperature conditions. Upon the premise concerning those issues, the potency of the photo-voltaic cell will be refined by sustaining their operative temperatures as low as possible. This paper portrays a technique of PV/solar panel cooling using convection generated by the chimney effect. This paper considers the diminution of warmth from the Photo-voltaic panel for both active and inactive conditions.*

Keywords: air-cooling, convection, solar-cell cooling, solar photo-voltaic

CHŁODZENIE PANELI FOTOWOLTAICZNYCH Z WYKORZYSTANIEM EFEKTU STOSU

Streszczenie. *Niezadowalająca wydajność ogniwa fotowoltaicznego jest jednym z najważniejszych problemów związanych z promocją technologii PV. Istotnym czynnikiem wpływającym na wydajność ogniwa jest jego temperatura robocza. Napięcie ogniwa spada niemal bezpośrednio wraz ze wzrostem temperatury roboczej ogniwa fotowoltaicznego. Temperatura na powierzchni działającego panelu fotowoltaicznego pozostaje zwykle o 22–30°C wyższa od normalnych warunków temperaturowych. Zgodnie z założeniami dotyczącymi tych kwestii, moc ogniwa fotowoltaicznego zostanie poprawiona poprzez utrzymanie ich temperatury roboczej na jak najniższym poziomie. Niniejszy artykuł przedstawia technikę chłodzenia paneli fotowoltaicznych/solarnych z wykorzystaniem konwekcji generowanej przez efekt kominowy. W artykule uwzględniono zmniejszenie ciepła z panelu fotowoltaicznego zarówno w warunkach aktywnych, jak i nieaktywnych.*

Słowa kluczowe: air-cooling, convection, solar-cell cooling, solar photo-voltaic

Introduction

Today, the world is dealing with some serious issues of energy deficiency, warming and degeneration of environment sources and energy power resources, renewable-resources have gotten more attention. Energy is one amongst comparable candidates. Energy source is widely obtainable with no cost. Energy is transformed into electric power by PV effect. PV system is consistent, quiet and freed from moving parts which leads to reduction of operation cost and servicing cost of the system. Being pollution free and pure source of energy, PV method has earned far more importance. Power output of PV method depends mainly on solar irradiance, temperature of cell and operating voltage. Anyway they actually present an infinite region of rivalry contrasting to straightforward energy assets by reason of their hefty expense and low potency during energy change.

The PV cells are ready to produce energy source from the abundant resource of sunlight. Since the PV components are exposed directly to the sunlight, it produces heat still as electricity. A PV module transforms solely 12–15% of the input solar power to electricity and also the undefined power is actually released as heat energy. The good and comfortable atmosphere affects this density/voltage (J/V) features of the photo-voltaic components wherever their electrical capabilities are adversely stricken with the numerous rise of cell operative temperature throughout absorption of radiation [10].

By applying a cooling methodology to a PV component, the worth of different energy is reduced in 3 ways. First, cooling improves the electrical output of PV components. Second, cooling method makes doable the utilization of concentrating solar cell method by shielding the photo-voltaic cells from outreaching the temperatures and so permanent loss occurs, even below the irradiation of several suns. As these drives it doable to alter solar cells with probably more cost-effective concentrators. Finally, the heat is unfastened by the photo-voltaic cell cooling method is employed for constructing the heating system or cooling system, or in commercial applications. to the current outcome, hybrid PV/T solar systems are explored as a way of reducing the heat of PV components and boost their electrical capabilities. This category of a PV scheme is termed Solar PV thermal collector (PV/T collector), which create thermal and current concurrently.

The hybrid PV thermal systems are still under development. J. K. Tonui, Y. Tripanagnostopoulos [16] discussed about an enhanced solar PV-thermal collector an addition to warmth extortion of air passage by forcibly and naturally, while R. Mazon-Hernandez et al [7] discussed about the heat reduction process out of the photo-voltaic arrays located on top of the hothouse by using air because the thermal energy carriers. The steady-state effect concerning the fins that's been vertical in structure, working on PV/T air warmer was been studied by Marc A. Rosen, Rakesh Kumar [6] while experimental research of air formed in Photo-voltaic Thermal gatherer in addition to various styles of Heat exchangers was being carried by H. J. Mohd. Yusof [9].

This paper illustrates a photo-voltaic array cooling method. The uninvolved cooling method is within the model of an oblong channel, which envelops the rearmost of the console and expands up-above towards the zenith of console. The cooling effect is induced here by stack effect. Both dynamic and uninvolved cooling is being studied using the identical system with an addition of fan connected at the very best section that's being powered using the device. During this technique, air is used for the reason that of the warmth carrier. Designing and Theoretical investigation of the technique is explained below.

1. Proposed system

The design comprises of an absorber section additionally to the planning proposed by Tonui and Tripanagnostopoulos [15]. the pace of air ascending within the tube (chimney) is commensurate to the energy utilized by the air, so addition of an absorber section will assist to stiff the natural draft of air which might be employed for reducing the heat of photo-voltaic arrays. The uninvolved cooling method will accommodate mainly two segments, top portion and center segment as in Fig. 1. Top-head section contains of the upright addendum to the center segment. It will also act sort of a chimney during this structure arrangement to uphold the natural draft fabricated by the nice and cozy uprising breeze. The beam of radiated energy falling over the absorbent section that's product of glass. These incident solar rays are soaked up by means of the air that's under the absorber section that in due course gets thaw out furthermore its density declines. The nice and cozy air will attempt to arise, thanks towards the buoyant force and flee all the way through the vent. The elevated vent will deliver a pressure variation in the middle of the underside and also the zenith which can further assist within the breeze of the surrounding air.

The middle portion of cooling chimney includes section of PV array as well as absorber section. The absorber section is meant such its size is greater than the size of PV array. As in Fig. 2 a glass sheet, which is transparent, is planted over the duct that's

formed of low- carbon steel. Photo-voltaic array is implemented on the lower 1/2 middle section whereas the glass sheet at the upper 1/2 the centre section acts because the absorber section. Centre segment of it, is fixed at an slant inclination capable local range to create sure that the photo-voltaic array as well as the absorber part obtains maximal energy over the year.

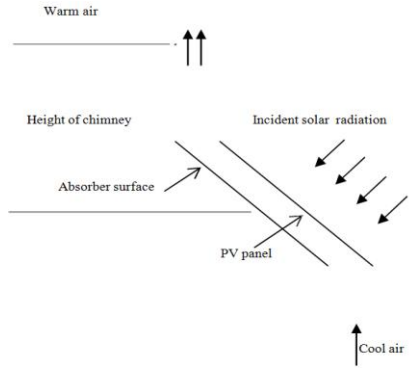


Fig. 1. Layout of cooling system

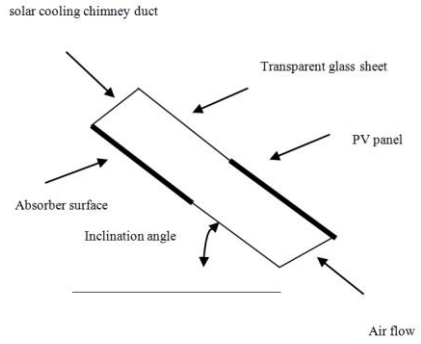


Fig. 2. Center portion of cooling system

The most of all solar beams which incident on it, will undergo the sheet, transparent in nature, reckoning on its reflectiveness and emissivity's and drop on surface of absorber, which is coated dense-black to maximise the portion of radiation energy absorbed. Consequently, a huge increment in absorbing the heat on the surface will occur. The heat transfer happens through the warmth absorbing surface to the atmosphere surrounding, impending from section of the photo-voltaic board panel. Updraft current with minor density will soar aloft in regard to the best section. Elevation about the superior part will aid the great and comfortable air current to soar towards the best owing to the static-head difference. Proportion of air gush will shoot-up if we increase the altitude of tube (chimney). The natural air draft within the channel duct instigated attributable to the variation in density between the current of air within the centre portion and nominal whiff of air up-above the best part will generate the suction force, in which it empowers the normal air draft to infiltrate from base section. Cool breeze entering from the underside tends to engulf the surface of heated photo-voltaic array and lessens the temperature of it.

2. Experimental setup

The cooling scheme was fabricated utilizing mild steel for demonstrating the concept. The particular experimental model embraces two key parts as illustrated in the "Design" segment. The head section was optimized to refine the natural draft in a better way. A photo-voltaic array of 10 W, 12 V Polycrystalline was utilized in the test setup. The panel is introduced on the inclined section, below the absorber and the same panel is attributed on a standalone frame to have a comparison study. The relative view of the system is made known in Fig. 4 Plain glass of 5mm thickness was accustomed to act as the transparent sheet that allow to pass solar radiation hooked on the absorber section. The channel wall or duct made of mild steel is insulated with glass wool so that to avoid heat loss. Fig. 3 portrays the isometric view of the air channel for the air conditioning setup with dimensions in mm.

The system is installed on a stand that is fabricated using mild steel L angle. During experiments temperature of the photo-voltaic array Tpv, inlet Tin and outlet Tout air temperatures for different solar insolation's are estimated using a Infrared thermometer and the readings are noted for further calculations. Here, Submissive (passive) temperature reduction of the photo-voltaic panel is being studied using this experimental setup.

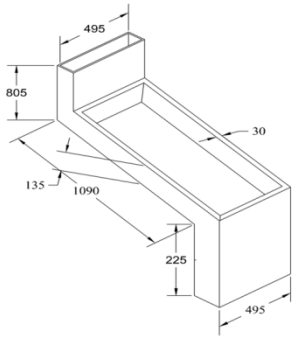


Fig. 3. Isometric view of cooling system

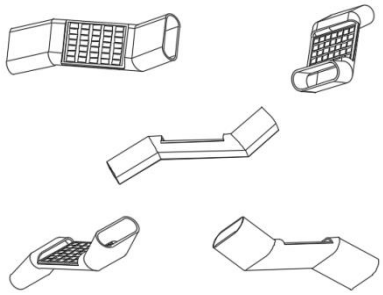


Fig. 4. Relative view of cooling system

The active cooling of photo-voltaic panel by means of the same setup can be studied by making small changes in the design. In dynamic cooling scheme exhaust fan driven with help of photo-voltaic panel that is employed to provide the desired cooling effect as in Fig.4. Here cooling effect is expected to be more than that could be achieved by any passive heat reduction system. 12 V exhaust fan is connected at the top part of the particular system. 12 V, 7 Ah battery is also used here.

3. Model analysis

A logical framework is defined to replicate the level of energy balance. Rays from the sun, falls on the chamber of absorber and the photo-voltaic array in the replica. The altitude of head segment is enhanced to maximize the natural air draft and also to reduce the shading effect. Profundity of duct is regarded as 0.1m. As considering the surface of the absorber, its size is greater than the size of the photo-voltaic array. Warmth quantity that penetrates the respective system, which should be equivalent up to the volume of energy that abandons the same system. Energy of heat that infiltrate to the structure, can be stated as (Randall,Mitchell [12]; Brinkworth, Frank P. Incropera [3]; Johnson et al, Akbarzadeh [1]):

$$Q_{in} = I \times A_{absorber} \times (1 - r) \times \tau \quad (1)$$

Fig. 5 illustrates the photo-voltaic array section where the photo-voltaic array is seated onto the inclined channel (duct) of the cooling chimney and Fig. 6 illustrates the three dimensional view of the proposed contraption.

Contemplating the model, it consists of walls which are adiabatic in nature, excluding the glassy sheet. Summation of the heat amount which escapes the structure can possibly state as

$$Q_{out} = Q_{air} + Q_{conv-cond} \quad (2)$$

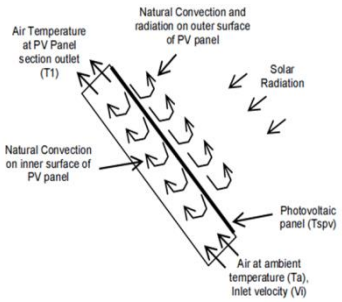


Fig. 5. Transfer of heat over the surface of the panel

Fig. 6. Three-dimensional view of the proposed contraption

The amount of heat absorbed by the air inside the channel or duct possibly provided as

$$Q_{air} = V_i \times A_{duct} \times \rho_i \times C_{p\,air} \times (T_o - T_i) \quad (3)$$

The amount of heat that lost on account of conduction and convection via the glassy sheet possibly indicated as

$$Q_{conv-cond} = \frac{\frac{T_i+T_o}{2} - T_i}{R_{conv-inside}+R_{cond}+R_{conv-outside}} \quad (4)$$

Convective thermal resistance inside the chimney duct

$$R_{conv-inside} = \frac{1}{h_i \times A_{absorber}} \quad (5)$$

Conductive thermal resistance for transparent sheet

$$R_{cond} = \frac{Thickness}{K \times A_{absorber}} \quad (6)$$

Convective thermal resistance provided to outside of the chimney is stated as

$$R_{conv-outside} = \frac{1}{h_o \times A_{absorber}} \quad (7)$$

Inside the chimney duct, if the heat is forced into the setup, that will increase the air warmth in it. Therefore, the total draft pressure generated over the chamber of absorption with the zenith of the vent can be developed as (Brinkworth 2000 [3]; Akbarzadeh, Johnson et al. 2009 [1])

$$\Delta P = (\rho_i - \rho_o) \times g \times H \quad (8)$$

Pressure fall athwart the system possibly stated by (Leenders et al., Bazilian [2]; Cengel,Turner et al. [4]; Ong [11])

$$P_{drop} = f \times \frac{1}{D_h} \times \frac{1}{2} \times \rho_i \times V_i^2 \quad (9)$$

Comparing these equations to invade the heat in the system and to escape the heat from the system, an equation has been developed with reference to the exit velocity. Likewise the equations for pressure and total draft pressure fall apart the setup, can be reorganized with reference of velocity at outlet. Resolving these equations altogether, the velocity at outlet and calefaction of the exhaust air can be predicted.

4. Conclusion

By using the effects of buoyancy in the chimney, the natural air draft can be attained and it can be also utilized as a apathetic cooling medium for photo-voltaic arrays. In the plan of the setup, simple modification was made and adding the section of absorber had helped to develop the induced natural draft of air as well as also as a consequence, aids to develop the competence concerning the photo-voltaic array. Elementary and overture investigation illustrates that in the process of coupling, an air passage channel and an absorber section with a photo-voltaic array and also possibly attain the appreciable conditioning system of air, of the photo-voltaic array as well as the potency of photo-voltaic array can possibly enhance.

References

[1] Akbarzadeh A. et al.: Examining potential benefits of combining a chimney with a salinity gradient solar pond for production of power in salt affected areas. Solar Energy 2009, 1345–1359.
[2] Bazilian M. et al.: Photo-voltaic cogeneration in the built environment. Solar Energy, 2001, 57–69.
[3] Brinkworth B. J.: Estimation of flow and heat transfer for the design of PV cooling ducts. Solar Energy 2000, 413–420.
[4] Cengel Y. et al.: Fundamentals of Thermal- Fluid Sciences. Applied Mechanics Reviews, 2003.
[5] Ibrahim A. et al.: Hybrid Photo-voltaic Thermal (PV/T) Air and Water Based Solar Collectors Suitable for Building Integrated Applications. American Journal of Environmental Sciences 5, 2009, 618–624.
[6] Marc A. R., Rakesh K.: Performance of a photo-voltaic/thermal solar air heater:Effect of vertical fins on a double pass system. International Journal of Energy and Environmental Engineering 2(4) 2011, 1–2.
[7] Mazon-Hernandez R. et al.: Development of an installation to reduce the temperature of photo-voltaic modules and improve their efficiency. International Conference on Renewable Energies and Power Quality (ICREPQ'10), 2010.
[8] Mittelman G. et al.: A model and heat transfer correlation for rooftop integrated photo-voltaics with a passive air cooling channel. Solar Energy 83, 2009, 1150– 1160.
[9] Mohd Yusof H. J. et al.: Performance Study of Air-based Photo-voltaic-thermal (PV/T) Collector with Different Designs of Heat Exchanger. Sains Malaysiana 42(9), 2013, 1319–1325.
[10] Nishiokaa K. et al.: Field-test analysis of PV system output characteristics focusing on module temperature. Solar Energy Materials & Solar Cells 75, 2003, 665–671.
[11] Ong K. S.: A mathematical model of a solar chimney. Renewable Energy 2003, 1047–1060.
[12] Randall K. R. et al.: Natural Convection Heat Transfer Characteristics of Flat Plate Enclosures. Journal of Heat Transfer 1979, 120–125.
[13] Sandberg M., Moshfegh B.: Buoyancy-induced air flow in photo-voltaic facades: Effect of geometry of the air gap and location of solar cell modules. Building and Environment 2002, 211–218.
[14] Sandberg M., Moshfegh B.: Investigation of fluid flow and heat transfer in a vertical channel heated from one side by PV elements, part II – Experimental study. Renewable Energy, 254–258.
[15] Tonui J. K., Tripanagnostopoulos Y.: Performance improvement of PV/T solar collectors with natural air flow operation. Solar Energy 82, 2008, 1–12.
[16] Tonui J. K., Tripanagnostopoulos Y.: Improved PV/T solar collectors with heat extraction by forced or natural air circulation. Renewable Energy 32, 2006, 623–637.

Ph.D. Kudithi Nageswara Rao
e-mail: nageswarak@hindustanuniv.ac.in

He is working as Assistant Professor in Department of Electrical & Electronics Engineering, Hindustan Institute Technology and Science, Padur, Chennai. He possesses 12+ years of experience from includes embedded technology industry, and rich teaching experience. He published more than 15 research articles including reputed conferences and he filed 1 patent in Intellectual property India, also he published a book. His area of interest is Renewable energy systems, Electric Vehicles and controller designs.

http://orcid.org/0000-0003-4490-0900

M.Sc. Rajkumar Ganesamoorthy
e-mail: raj06@ymail.com

He is a Research Scholar in Department of Electrical & Electronics Engineering, Hindustan Institute Technology and Science, Padur, Chennai. He possesses a decade of experience in the academic field, and in addition to one year of experience working as a Project Engineer in the EV industry. His professional background includes the successful publication of over 40 academic journals and completion of seven funded projects. He has received multiple national-level awards for my work.

http://orcid.org/0000-0002-6505-892X

http://doi.org/10.35784/iapgos.3755 *received: 18.06.2023 | revised: 08.09.2023 | accepted: 27.09.2023 | available online: 30.09.2023*

A NEW AUTOMATIC INTELLIGENCE-BASED SOLAR LOAD CONTROL SYSTEM

Kudith Nageswara Rao, Ganesamoorthy Rajkuma
Hindustan Institute of Technology and Science, Department of Electrical and Electronics Engineering, Chennai, India

Abstract. *In modern times, solar panels have become a common sight in many households as they provide electricity for various purposes. Typically, the solar panel's charges a battery, and any excess energy generated is usually wasted once the battery is fully charged. However, by utilizing this extra energy, heavy loads can be powered as well. This is where a solar power controller comes into play, which measures the parameters of the solar cell through multiple sensor and adjusts the load accordingly. When the power output of the PV cell is high, the load runs on solar power, and if the power is not sufficient, the load switches to the main supply. The load switches back to solar power when it becomes high again. Monitoring the solar cell parameters allows for real-time identification of the power produced by the solar panel.*

Keywords: load control system, solar PV

NOWY, AUTOMATYCZNY SYSTEM KONTROLI OBCIĄŻENIA SŁONECZNEGO OPARTY NA SZTUCZNEJ INTELIGENCJI

Streszczenie. *W dzisiejszych czasach panele słoneczne stały się powszechnym widokiem w wielu gospodarstwach domowych, ponieważ zapewniają energię elektryczną do różnych celów. Zazwyczaj panele słoneczne ładują akumulator, a nadmiar wytworzonej energii jest zwykle marnowany po pełnym naładowaniu akumulatora. Jednakże, wykorzystując tę dodatkową energię, można również zasilać duże obciążenia. W tym miejscu do gry wkracza kontroler energii słonecznej, który mierzy parametry ogniwa słonecznego za pomocą wielu czujników i odpowiednio dostosowuje obciążenie. Gdy moc wyjściowa ogniwa fotowoltaicznego jest wysoka, obciążenie jest zasilane energią słoneczną, a jeśli moc nie jest wystarczająca, obciążenie przełącza się na główne źródło zasilania. Obciążenie przełącza się z powrotem na zasilanie energią słoneczną, gdy jej moc ponownie staje się wysoka. Monitorowanie parametrów ogniwa fotowoltaicznego pozwala na identyfikację mocy wytwarzanej przez panel fotowoltaiczny w czasie rzeczywistym.*

Słowa kluczowe: system kontroli obciążenia, panel słoneczny

Introduction

Solar radiation is a readily available and clean source of energy. Being situated in tropical regions, India has a high availability of solar energy, which can be harnessed to meet a significant portion of household energy needs. As a result, the solar power market has emerged as one of the fastest-growing renewable energy markets globally. Solar panels are widely used for both industrial and household purposes, with power stations available to monitor the solar panel circuits and parameters. In households, it's essential to monitor solar panel parameters such as voltage, current, and power. Solar power is stored in batteries, and when the battery is fully charged and the energy production is high, shifting loads to solar power becomes necessary. However, this shifting process is currently manual and not particularly user-friendly.

1. Objectives

In order to effectively utilize solar energy by selecting loads to operate on solar power and to monitor solar panel parameters and power output, a system is required. The solar power controller is an essential component that enables the optimal use of solar energy by measuring the solar parameters in real-time. Based on the measured values, the system adjusts the loads and allows for seamless switching of loads between the solar inverter and main supply, ensuring efficient utilization of solar energy.

2. Proposed system design

The block diagram and circuit diagram of the solar power controller with an auto intelligence load control system are shown in figures 1 and 2 respectively.

When a solar panel is exposed to sunlight, it generates electricity. To measure the output voltage of the solar panel, a voltage divider circuit is connected across it, which reduces the voltage to a measurable range [2, 3, 6, 7, 8, 9, 14, 15]. The output voltage is then directed to pin A0 of the Arduino microcontroller. In addition, a current sensor is also incorporated into the circuit to sense the current produced by the solar panel. The current sensor output is connected to pin A3 of the Arduino [17]. To regulate the battery charging voltage and avoid voltage fluctuations, a buck converter is used as a charge controller in the circuit [1, 5, 10, 11, 16].

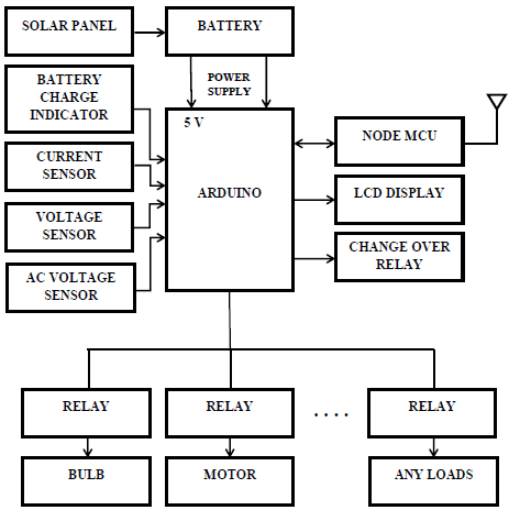


Fig. 1. Block diagram of Solar power controller

In order to monitor the battery level of the solar panel system, a voltage divider circuit is connected across the battery, and the output is displayed on an LCD screen [4, 12, 13, 18]. The voltage divider output is then directed to pin A1 of the Arduino. To measure the KSEB (Kerala State Electricity Board) current, a small step-down transformer is used, which converts 230 V to 9 V, and its output is connected to pin 2 of the Arduino. A voltage regulator reduces the 9 V to 5 V. If KSEB power is available, pin 2 of the Arduino becomes high, and low otherwise. The relay circuit is connected to pin 3 of the Arduino. When the solar panel output is high, the relay circuit closes, and the common pin switches from normally closed to normally open, connecting KSEB power to the normally closed pin and solar power to the normally open pin. An alarm is connected to pin A2 of the Arduino. An LCD converter is used to connect the LCD screen to the Arduino, with pins A4 and A5 of the Arduino connected to the SDA and SCL pins of the LCD converter. The LCD displays the panel voltage, panel current, battery voltage, and source of supply. Four loads are connected to pins 10, 11, 12, and 13 of the Arduino. Two light loads

are connected to pins 10 and 11, and two heavy loads are connected to pins 12 and 13. A NodeMCU module is connected to pins 0 and 1 of the Arduino, which enables communication with the user.

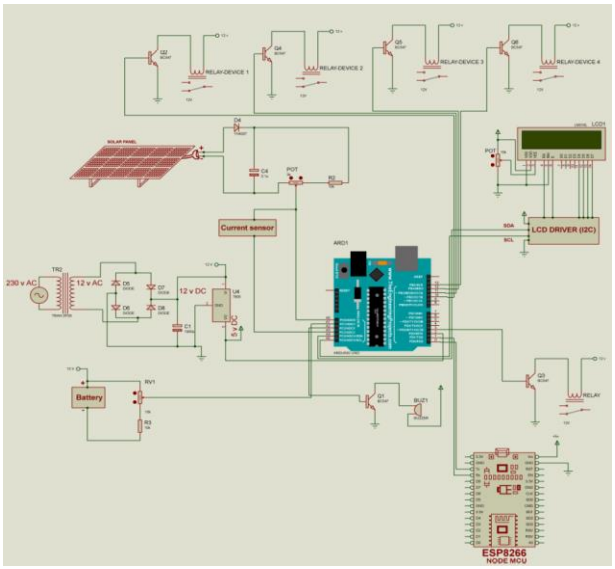


Fig. 2. Circuit diagram of Solar power controller

2.1. Experimental analysis

The system uses a current sensor called the ACS712 to sense current, and if the output voltage of the solar panel exceeds 12 V, the system shifts the supply from the KSEB to the battery. This is done by triggering a transistor, which energizes the changeover relay and switches the load connection from the AC mains to the battery. During this shift, light loads like fans and lights can continue to operate using solar power without interruption, while heavy loads like washing machines and induction cookers can be manually switched on as needed.

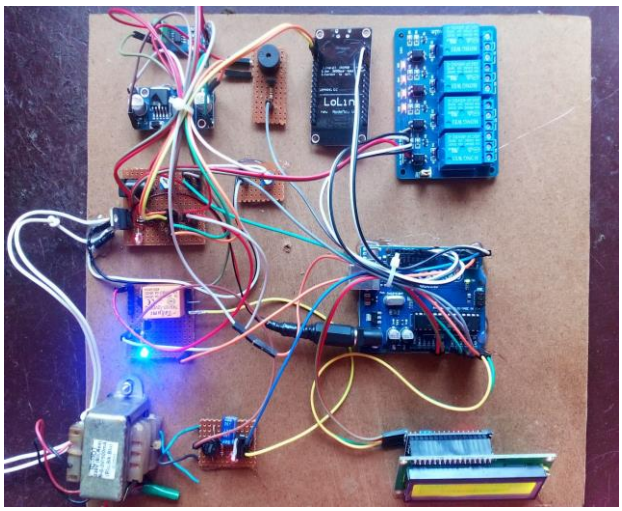


Fig. 3. Hardware implementation

2.2. Result analysis

2.2.1. Battery charging estimation

We have taken a 10 W solar panel and 12 V, 7 A battery.
So current produced,

$$I = P/V$$

$$I = 10/12 = 0.833 \text{ A}$$

So charging time,

$$T = 7/0.833 = 8.5 \text{ hours}$$

2.2.2. Working Conditions

There are three conditions are considered.

- Condition 1 – Solar panel output power is higher than 12 V.
- Condition 2 – Solar panel output power is less than 12 V.
- Condition 3 – Solar power and KSEB are not available.

Condition 1: When solar panel output becomes higher than 12 V then the system will shift from KSEB to solar power and displayed values in LCD are shown in figure 4(a). Here panel voltage is 19 V and panel current is 0.53 A. The Source of supply is displayed as solar.

Condition 2: When solar panel voltage becomes less than 12 V then the system switched to KSEB or main supply and the displayed value in LCD is shown in figure 4(b). Here the source of supply is displayed as KSEB.

Condition 3: If solar power and KSEB are not available, then the system will be shut down and the status displayed in LCD is shown in figure 4(c). Here the source of supply is displayed as no power.

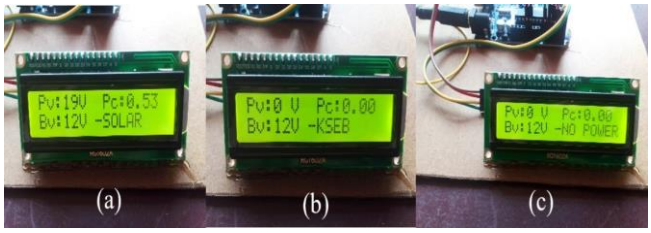


Fig. 4. Displayed values in: (a) condition 1, (b) condition 2, (c) condition 3

Measured parameters can be checked on a webpage. It shows solar panel voltage, solar panel current, and battery voltage. The webpage contains device status, and it is possible to on/off devices through it.

3. Design and implementation

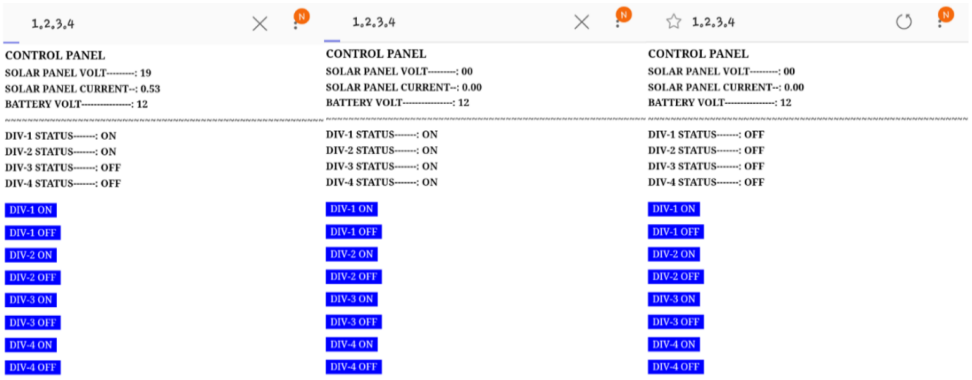
The process of system sizing involves determining the minimum sizes of PV panels, inverters, batteries, and charge controllers necessary to meet the required electrical energy demand based on the specific solar conditions at the site. This process ensures a balance between the system's output and solar input, while also accounting for any losses within the system. The design process encompasses the following steps:

1. Identify the watt-hour/day load that needs to be served.
2. Determine the average monthly solar energy available.
3. Calculate the solar panel size required to meet the load during the least favorable month.
4. Determine the appropriate size and type of battery to ensure reliable power supply.
5. Select the suitable type of charge controller.
6. Determine the capacity of the inverter needed for the system.

To perform system sizing accurately, the following information is crucial:

- The solar energy available in kWh/m2/day at the site during the month with the lowest solar energy.
- The average daily energy requirement in watt-hours (Wh) for operating the desired appliances, as well as any specific power needs that exceed the average.
- The losses that occur within the PV system, which diminish the energy available to the user.

System sizing involves determining the minimum sizes of PV panels, inverters, batteries, and charge controllers necessary to deliver the required electrical energy under the specific solar conditions at the site. It aims to achieve a balance between the system's output and solar input while taking into account losses within the system.



a) b) c)

Fig. 5. Webpage for: (a) condition 1, (b) condition 2, (c) condition 3

The design process for system sizing consists of the following steps:

1. Identify the load to be served, measured in watt-hours per day.
2. Determine the average solar energy available, considering a month-by-month basis.
3. Calculate the required size of the solar panel to meet the load demand during the worst month conditions.
4. Determine the appropriate size and type of battery necessary to ensure the desired power reliability.
5. Select the suitable type of charge controller for the system.
6. Determine the capacity of the inverter required to meet the power needs.

In order to do the system sizing, we need to know the following in formations:

- The solar energy available in kWh/m^2/day at the site for the lowest solar energy month of the year.
- The average Wh/day required by the user to operate the desired appliances and any special need for power that go much beyond the average.

The losses that occur in the PV system that reduces the energy available to the user.

3.1. Load estimation

The process of estimating the total watt-hours to be supplied involves the following steps:

1. Determine the power rating, measured in watts, for each appliance that will be used in the household.
2. Estimate the number of hours per day that each appliance is expected to be used.
3. Multiply the power rating of each appliance by the estimated hours of use to calculate the watt-hours per day for each individual appliance.
4. Calculate the total watt-hours per day by summing up the watt-hour values obtained for all appliances combined.

The different power ratings of household appliances in watts are listed in Table 1.

Table 1. Household appliances usage

Load	Watt(W)	Number of Load	Working Hours per Day
LED Bulbs	12	3	5
	9	5	5
CF Lamps	20	4	5
	15	3	5
	8	4	5
LED TV	45	1	4
Washing Machine	2150	1	1
Pump	500	1	½
Refrigerator	135	1	5
Mixer Grinder	600	1	1/6
Fan	40	10	4
Computer	90	1	1/3

Average watt-hours used by,
LED Bulbs = (9×5×5) + (12×3×5) = 225 + 180 = 405 Wh
CF Lamps = (20×4×5) + (15×3×5) + (8×4×5) = 400 + 225 + 160 = 785 Wh
LED TV = 45×1×4 = 180 Wh
Washing Machine = 2150×1×1 = 2150 Wh
Pump = 500×1×(1/2) = 250 Wh
Refrigerator = 135×1×5 = 675 Wh
Mixer Grinder = 600×1×(1/6) = 100 Wh
Fan = 40×10×4 = 1600 Wh
Computer = 90×1×(1/3) = 30 Wh
Total watt-hour required per day = 405 + 785 + 180 + 2150 + 250 + 1600 + 675 + 30 + 100 = 6175 Wh

3.1.1. Solar panels sizing

If the solar panel of 250W is chosen, with Open circuit voltage of 37.8 V, with physical dimensions of 250 watt panel and the dimension is 1650 mm×992 mm×40 mm.
Surface area of 250 W/ panel = 1.65 × 0.992 = 1.6368 m^2
Therefore, the roof area required for installing the solar PV system will be:
Roof area required = Number of panels × Surface area of one module = 10 × 1.6368 = 16.368 m^2

3.1.2. Battery sizing

The battery specification is 2500 cycles to 50% depth of discharge means that a solar system is discharged on an average to a depth of 50% per day then it could be expected to have a life of around 2500 days or 6.8 years.

The following assumptions are made for selecting the capacity of the battery bank:

1. Battery efficiency is assumed to be 85%
2. Depth of discharge is assumed to be 50%
3. Cloudy days assumed to be 2 (battery supports 2 days of energy requirements)
4. Battery voltage levels are chosen based on the wattage capacity of the plant

Battery sizing calculation:
Wh/day required by the load = 6175 W
Battery must supply = 6175 ÷ 0.85 = 7264.7 W
≈ 7265Wh/day (based on asssump-1)
Battery bank must supply = 7265 ÷ 0.5 = 14530 Wh (based on assump-2)
Battery bank must be sized = 14530 × 2 = 29060 Wh (based on assump-3)
With 48 V battery, Ah capacity = 29060 ÷ 48 = 605 Ah
Therefore, it is better to choose 16 Numbers of 200 Ah, 12 V batteries connected as 4 in series and 4 in parallel (4×200 = 800 Ah at 48 volts) and this will be able to supply the load for 2 day without 36 interruption.

3.1.3. Load shifting

Solar power can be utilized to power various household loads, but it is essential to install the appropriate panels and batteries based on the load requirements. In the earlier mentioned case, the total watt-hours required per day amounts to 6175 Wh. All loads can operate using the main supply, and when the shift is made from the main supply to solar power, the light loads can continue functioning seamlessly. The light loads include LEB bulbs, CF lamps, fans, LED TVs, refrigerators, and computers. On the other hand, other loads such as pumps, washing machines, and mixer grinders can be switched on directly or controlled through a webpage. Figure 6 illustrates the loads that can operate without any interruptions during the transition.

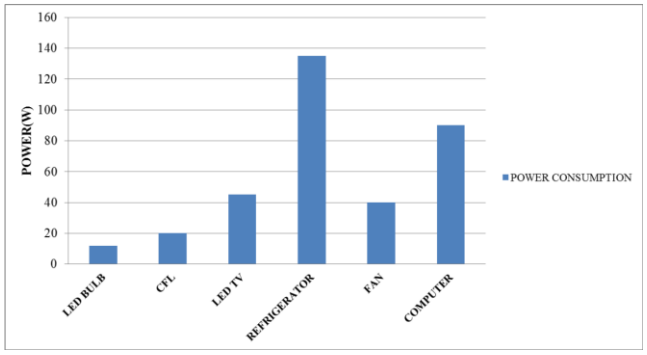


Fig. 6. Load operation with solar power

4. Conclusion

The typical means of measuring solar panel performance involve either costly and specialized systems utilized by the industry or specialized testing laboratories. These systems are both rare and expensive, and they require skilled personnel to maintain them. In contrast, the system developed for this project is designed for household use, providing a means to measure solar panel parameters and calculate solar power output in real-time. By using this system, it is possible to optimize the use of solar energy and reduce electricity costs. The IoT-based platform also enables communication with the user.

Conflict of interest

The authors have declared no conflict of interests.

Funding

This research work is not funded by any external agencies.

References

[1] Dabou R. et al.: Development of autonomous monitoring and performance evaluation system of grid-tied photovoltaic station. International Journal of Hydrogen Energy 46(59), 2021, 30267–30287.
[2] Dini H. S., Rizki P. P.: Automation and mobile phone based-monitoring of hydroponic farming style using solar energy. Acta Electron. Malays. (AEM) 5(5), 2021.
[3] Ergashevich K. K. et al.: Wsn-based Monitoring Systems for the Solar Power Stations of Telecommunication Devices. IIUM Engineering Journal 22(2), 2021, 98–118.
[4] Gallego A. J. et al.: Model Predictive Control of the Mojave solar trough plants. Control Engineering Practice 123, 2022, 105140.
[5] Ganesamoorthy R., Sekar S.: An improved intermittent power supply technique for electrostatic precipitators. Energy Harvesting and Systems, 2022.
[6] Hu S. et al.: Disguised Tailing and Video Surveillance With Solar-Powered Fixed-Wing Unmanned Aerial Vehicle. IEEE Transactions on Vehicular Technology 71(5), 2022, 5507–5518.
[7] Imtiaz A. et al.: Grid-Assisted Rooftop Solar PV System: A Step toward Green Medina, KSA. Smart Science 7(2), 2019, 130–138, [http://doi.org/10.1080/23080477.2019.1565053].
[8] Jino R. et al.: A self-powered, real-time, LoRaWAN IoT-based soil health monitoring system. IEEE Internet of Things Journal 8(11), 2021, 9278–9293.
[9] Kandeal A. W. et al.: Infrared thermography-based condition monitoring of solar photovoltaic systems: A mini review of recent advances. Solar Energy 223, 2021, 33–43.
[10] Karthik B., Priyanka Gandhi M.: Solar powered based DC drives controlled by using IoT. Materials Today: Proceedings 37, 2021, 2513–2516.
[11] Kerem A., Abdusselam Y.: Design and prototyping of GSM-bluetooth based solar energy remote monitoring system. COMPEL-The international journal for computation and mathematics in electrical and electronic engineering, 2021.
[12] Kotb M. et al.: Enriching the stability of solar/wind DC microgrids using battery and superconducting magnetic energy storage based fuzzy logic control. Journal of Energy Storage 45, 2022, 103751.
[13] Liu J. et al.: Domain size control in all-polymer solar cells. Iscience 25(4), 2022, 104090.
[14] Rajkumar G., Ajna K. I.: Design and Performance Evaluation of Efficiency Enhanced Solar Food Processor. Smart Science 10(4), 2022, 283–293, [http://doi.org/10.1080/23080477.2021.2012009].
[15] Shakya S.: A self monitoring and analyzing system for solar power station using IoT and data mining algorithms. Journal of Soft Computing Paradigm 3(2), 2021, 96–109.
[16] Trevathan J. et al.: An IoT general-purpose sensor board for enabling remote aquatic environmental monitoring. Internet of Things 16, 2021, 100429.
[17] Wahyu S. et al.: Efficiency Testing Of Solar-Powered Smart Green House Systems For Plant Culture. Jurnal Teknik Elektro dan Komputer 11(1), 2022, 9–14.
[18] Zhou K. et al.: Morphology control in high-efficiency all-polymer solar cells. InfoMat 4(4), 2022, e12270.

Ph.D. Kudithi Nageswara Rao
e-mail: nageswarak@hindustanuniv.ac.in

He is working as Assistant Professor in Department of Electrical & Electronics Engineering, Hindustan Institute Technology and Science, Padur, Chennai. He possesses 12+ years of experience from includes embedded technology industry, and rich teaching experience. He published more than 15 research articles including reputed conferences and he filed 1 patent in Intellectual property India, also he published a book. His area of interest is Renewable energy systems, Electric Vehicles and controller designs.

http://orcid.org/0000-0003-4490-0900

M.Sc. Rajkumar Ganesamoorthy
e-mail: raj06@ymail.com

He is a Research Scholar in Department of Electrical & Electronics Engineering, Hindustan Institute Technology and Science, Padur, Chennai. He possesses a decade of experience in the academic field, and in addition to one year of experience working as a Project Engineer in the EV industry. His professional background includes the successful publication of over 40 academic journals and completion of seven funded projects. He has received multiple national-level awards for my work.

http://orcid.org/0000-0002-6505-892X

http://doi.org/10.35784/iapgos.3650 *received: 13.05.2023 | revised: 19.07.2023 | accepted: 23.09.2023 | available online: 30.09.2023*

OPTIMIZATION OF PARTS CUTTING PROCESS PARAMETERS WORKING IN CONDITIONS OF CYCLIC LOADS

Kateryna Barandych[1], Sergii Vysloukh[1], Grygoriy Tymchyk[1], Oleksandr Murashchenko[2], Saule Smailova[3], Saule Kumargazhanova[3]

[1]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine, [2]Vinnytsia National Technical University, Vinnytsia, Ukraine, [3]D.Serikbayev East Kazakhstan State Technical University, Ust-Kamenogorsk, Kazakhstan

Abstract*. The paper is devoted to questions of technological fatigue life assurance of parts working in conditions of cyclic loads by optimization their cutting conditions for finish turning process. In order to solve the task of optimizing the parts cutting conditions, the corresponding software, based on the previously created mathematical model of the finishing turning process, was developed in the C# programming language. With the purpose of technological providing the necessary fatigue life of the part, taking into account the real conditions of its operation for the maximum productivity of the finishing turning process, the methodical recommendations for determining the optimal parts cutting conditions at the phase of production technological preparation are given. An example application of the proposed solution is presented.*

Keywords: technological support, fatigue life, finishing turning process, optimization

OPTYMALIZACJA PARAMETRÓW PROCESU CIĘCIA CZĘŚCI PRACUJĄCYCH W WARUNKACH OBCIĄŻEŃ CYKLICZNYCH

Streszczenie*. Artykuł poświęcony jest zagadnieniom zapewnienia technologicznej trwałości zmęczeniowej części pracujących w warunkach obciążeń cyklicznych, poprzez optymalizację ich warunków skrawania w procesie toczenia wykańczającego. W celu rozwiązania zadania optymalizacji warunków skrawania części, zostało opracowane odpowiednie oprogramowanie w języku C#, oparte na wcześniej stworzonym modelu matematycznym procesu toczenia wykańczającego. W celu technologicznego zapewnienia niezbędnej trwałości zmęczeniowej części, biorąc pod uwagę rzeczywiste warunki jej działania dla maksymalnej wydajności procesu toczenia wykańczającego, podano metodyczne zalecenia dotyczące określania optymalnych warunków skrawania części na etapie technologicznego przygotowania produkcji. Zaprezentowano przykładowe zastosowanie proponowanego rozwiązania.*

Słowa kluczowe: wsparcie technologiczne, trwałość zmęczeniowa, proces toczenia wykańczającego, optymalizacja

Introduction

Among these are shafts, axles, crank rods, pivot pins, gears [4, 6], rotors and their fastening elements [8], bearings, disks [5, 13, 14], etc., which have to work under loads that vary in magnitude and direction. As a result in the material of such parts stress (in size smaller than the limits of strength) vary in time arise. Shafts and axles, when running under constant external loading, experience symmetrical cycle alternating bending stresses that can cause fatigue failure of these parts. Fatigue life is one of the main fatigue strength properties of critical parts, which allows to determine the period of their exploitation.

More than 70% of all technological breakdowns are associated with fatigue failure. Local processes of origin and initial development of the crack do not have a visible effect on the deformation of the part as a whole and the accelerated development of the crack, as a rule, is not long lasting. As a result, the skin often occurs suddenly and becomes the cause of emergency situations. Thus, at insignificant stresses corresponding to multi-cyclic fatigue, the possibility of the origin of a crack after a given number of cycles is determined by fatigue life of the part's material and the state of it's surface layer.

On the part's fatigue strength three main groups of factors influence [7, 10]:

- constructive – parts geometry, stress concentrators, fit, safety margin, method of applying loads;
- operational – operating conditions (temperature, humidity and other physical and chemical properties of the environment), quality and frequency of lubrication and repair, compliance with the rules of operation of the product;
- technological – the process of obtaining the necessary material, the method of manufacturing the workpiece, the further processing methods of parts and assembly of the product.

In works [5, 13] the fracture distribution of the compressor blades according to the indicated factors is presented: 29% – constructive, 17% – technological, 11% – operational, 43% – from a combination of all factors. The destruction of turbine blades due to technological factors is 25.1%, constructive – 23.2%, the remaining 51.7% is due to an unfavorable combination of these factors.

Thus, one of the ways to ensure the trouble-free operation of the parts working under the influence of vary loads is to determine the optimal technological parameters of the processing and their use.

1. Formulation of the task

The purpose of the work is technological providing necessary part's fatigue life at their finishing turning process with maximum production efficiency by determining the optimal cutting conditions. To achieve this goal, it is necessary to develop a methodology for solving the optimization problem based on the mathematical model of the finishing turning process [8].

2. Task realization

The finishing turning process mathematical model of parts made from materials of the structural alloyed chromium steels group (1) [1, 11], is created in solving the optimization problem. The maximum productivity of the finishing turning process in the area specified by a set of constraints is chosen as the criterion of optimality.

The method of sliding admission, which relates to methods of nonlinear programming, is used to solve the optimization problem [12, 15]. Acceptable or almost acceptable solution for maximizing the optimality criterion found using the algorithm Neldera and Mead (deformed polyhedron).

$$\max f(S,V) = \begin{cases} S_{min} \le S \le S_{max} \\ \frac{\pi D_0 n_{min}}{1000} \le V \le \frac{\pi D_0 n_{max}}{1000} \\ P_{\max a.f.} \ge P_x = 10 C_{P_x} h^{X_{Px}} S^{Y_{Px}} V^{n_{Px}} K_{Px} \\ N_{eng}\eta \ge \frac{10 C_{P_z} h^{X_{Pz}} S^{Y_{Pz}} V^{n_{Pz}} K_{Pz}}{1000 \cdot 60} \\ 0.5 \cdot TD \ge \Delta_{\Sigma} \\ T \le \left(\frac{C_v K_v}{V h^x S^y}\right)^{\frac{1}{m}} \\ Ra_{req} \ge Ra \\ N_{req} \le K_p \cdot N \end{cases} \quad (1)$$

where $f(S,V)$ – productivity of the finishing turning process, which is determined by the formula $f(S,V) = \frac{1000 \cdot S \cdot V}{\pi \cdot D}$;

S – longitudinal feed, mm/rev; V – cutting speed, m/min; D – diameter of the part processed surface, mm; D_0 – diameter of the workpiece before finishing turning process, mm; S_{min} and S_{max} – respectively the minimum and maximum feedrate, mm/rev; n_{min}, n_{max} – respectively minimum and maximum spindle rotational speed, rpm; P_x – the axial component of the cutting force, N; $P_{\max a.f}$ – the maximum permissible value of the machine axial force, N; C_{P_x}, C_{P_z}, C_v – constants; X_{Px}, Y_{Px}, n_{Px}, X_{Pz}, Y_{Pz}, n_{Pz}, x, y, m – exponents, K_{Px}, K_{Pz}, K_v – correcting coefficients; N_{eng} – power of the machine-tool motor, kW; η – coefficient of efficiency of the machine-tool motor; h – cutting depth, mm; Δ_Σ – total processing error, μm; TD – dimension limit, μm; Ra_{req} – required part surface roughness, μm; Ra – machined part surface roughness, μm; N_{req} – required value of part fatigue life, cycles; N – calculated value of the 40Kh steel parts fatigue life, cycles; K_p – coefficient of generalized properties of the materials structural chrome alloy steel group.

After running the optimization program for execution, the main window of the program appears on the monitor. In this window, the user specifies in the relevant field the information about the tool, the part, the used equipment, the acceptable errors, parameters of finishing turning process recommended by normative data, the operational cycle stress and the required fatigue life [16].

After entering the initial data, clicking on the button "Calculate" a window with drop-down lists to select the structural material and the material of the cutting part of the instrument and the required qualification of the surface's accuracy.

The program initiates a window (Fig. 1) with a graphical representation of the admissible solutions area and the determined optimal cutting condition of finishing turning process by pressing the OK button.

The intermediate results of the optimization task solution and its final result are presented in the table 1, which includes the finishing turning process cutting condition, productivity and fatigue life. The results of the optimization task are saved as a text file at the specified address.

Table 1. Results of solving the optimization problem

Feed rate, S	0.11 mm/rev
Cutting speed, *V*	145.6 m/min
Cutting depth, *h*	0.2 mm
Productivity of the finishing turning process, *f*	217.488 m/min
Calculated value of the 40Kh steel parts fatigue life, N	80906 cycles

Thus, the created optimization method makes it possible, on the basis of experimental research and the processing of their results [2], to solve the problem of determining the optimal finish turning process cutting conditions of parts made from the materials of the structural alloyed chromium steels group, with a visual representation and preservation in a text file.

Implementation of the methodology for providing the necessary fatigue life of the parts at their finish turning process is carried out in the following sequence.

1. Analysis of the part construction and conditions of its operation.
2. Determination of maximum stresses in the part material during operation by finite-element analysis.
3. Determination of the initial data of the finishing turning process:
   - chemical composition, physical and mechanical properties of the processed material;
   - constructive data on the part and its workpiece;
   - instrument material the and its geometrical parameters;
   - the main technical characteristics of the processing machine.
4. Determination of the optimal finish turning process cutting conditions by a mathematical model of the finish turning process, which ensures the maximum performance of the finishing turning process of the part with its required fatigue life.

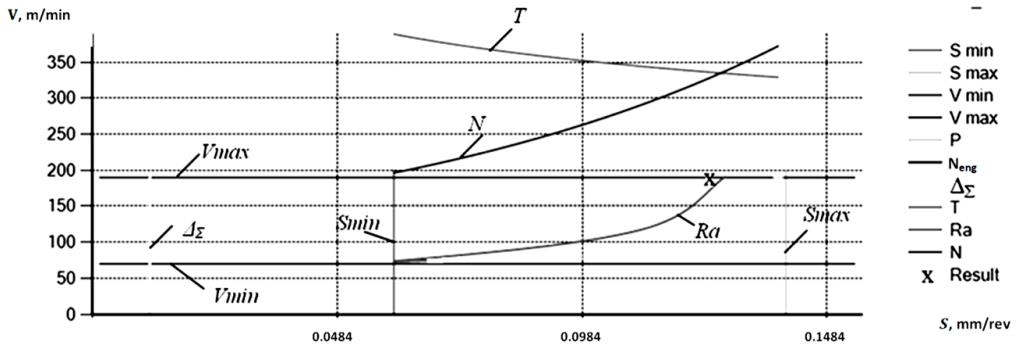


Fig. 1. Graphical representation of the admissible solutions area

3. An example of determining the optimal finish turning process cutting conditions of parts working in varying loads conditions

The part "Shaft" of a rotary mechanism operating in conditions of varying loads is considered as example of the application of the aforementioned methodology of providing the necessary fatigue life of the parts. The sketch of a shaft, made of steel 40Kh DSTU 7806:2015.

Production conditions of the rotary mechanism provide a abrupt change in the size and direction of loading on the output gear. This leads to the formation of cracks and subsequent fatigue failure of the shaft at the point of the shaft connection with a gear pin through the key. Consequently, the shaft is constantly in a complex stress-strain state.

The study of the operating conditions of the work of the part is made in the FEMAP 10.2.0 software program [3]. The boundary conditions that simulate the shaft fastening in the mechanism (boundary conditions of the 1st kind) on the model of the shaft are represented as triangles with a digital value, indicating a limited number of freedom degrees (Fig. 2). The torque acting on the conic surface of the shaft (boundary conditions of the 2nd kind) was determined by a pair of forces. The value of the operational load was 1000 H. The loading on the model is presented in the form of arrows in the direction of loading and its numerical value (Fig. 2). In Fig. 3 shows the finite-element model of the shaft in the deformed state.

A graph representing of equivalent tensile stresses or von Mises stresses for the for bulky KEs is presented on the right side of the software graphical interface. The distribution and stress values in the keyhole zone are shown in Fig. 3. The major stresses on a cylindrical surface were 400 MPa.

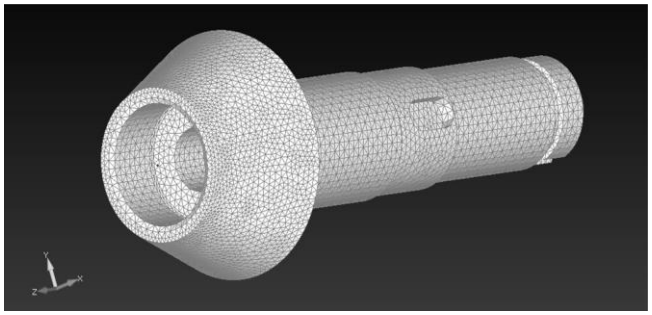


Fig. 2. General view of the finite-element model of the shaft

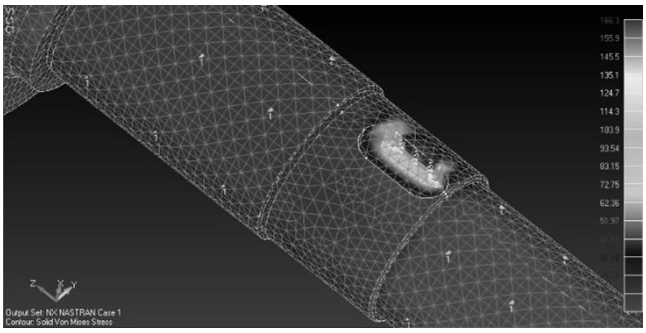


Fig. 3. Stress-strain state model of the shaft under operating conditions

The chemical composition, physical and mechanical properties of the part's material (steel 40Kh) are given in table 2. The required part fatigue life is $8.5 \cdot 10^4$ cycles. The finishing turning of the part was carried out at the Haas ST-20 CNC turning center, using the T15K6 tool material. The relative dimensional wear of the cutter is 0.7 μm/km. The required period of tool stability is 60 min [9]. The geometrical parameters of the tool cutting part are presented in table 3. The main technical characteristics of the machine-tool and part and workpiece data are induced in tables 4 and 5. The value of composite accuracy components of the turning process [9] is presented in table 6.

A set of initial data is used when forming the limitations of a mathematical model in a program for optimizing cutting conditions turning processing (Fig. 4).

Clicking the "OK" button initiates a window with the results of solving the optimization problem. The optimal cutting condition of the "Shaft" component (S = 0.138 mm/rev; V = 187.6 mm/min; h = 0.1 mm) are defined as a result of the use of the developed algorithm and the program for optimizing the finishing turning process by sliding admission and determined operational stresses in the part material, providing the necessary fatigue life (at least N = 16 104 cycles) during its operation and maximum production productivity (the value of the estimated target function is f = 217 m/min).

The given example shows that the proposed methodology of technological support allows determining the optimal cutting conditions, which provides the necessary fatigue life of the part with maximum productivity of its manufacturing process. According to this method, it is possible to predict the number of load cycles when operating the part before its destruction, which will avoid emergency situations.

Table 2. The chemical composition, physical and mechanical properties of the steel 40Kh

C, %	Mn, %	S, %	P, %	Cr, %	Si, %	Ni, %	Cu, %	N, %
0.44	0.8	0.035	0.035	1.1	0.37	0.3	0.3	0.008
E, MPa	G, MPa	ρ, kg/m3	σT, MPa	σB, MPa	δ, %	ψ, %	KCU, kJ/m2	HB
214000	85000	7850	785	980	10	45	590	217

Table 3. Geometrical parameters of the tool cutting part

Plan approach angle φ°	72.5
Face cutting edge angle φ_1°	72.5
Rake angle γ°	-14
Cutting edge inclination λ°	5
Tool nose radius r	0.4 mm

Table 4. Main technical characteristics of the Haas ST-20 CNC turning center

Maximum spindle power, $N_{шп}$	14.9 kW
Useful effect main drive of machine-tool, η	0.8
Maximum permissible value of the machine X axial force, $P_{x\,max}$	20.46 kN
Maximum spindle rotational speed, n_{min}	4000 rpm
Minimum spindle rotational speed, n_{max}	400 rpm
Maximum federate, S_{max}	6 mm/rev
Maximum system compliance, W_{max}	0.04 mm/kN
Minimum system compliance, W_{min}	0.035 mm/kN

Table 5. Details of the workpiece and workpiece

Diameter of the processing surface, D	∅ 38 mm
Surface roughness, Ra	1.25 μm
Diameter of processing workpiece, D_0	∅ 40 mm
Length of the processing surface, l_p	30.5 mm
Total length of tool passage in the direction of feed, L	40 mm
Quality accuracy processing surface	7
Batch size, n	25 p.

Table 6. Data for calculating the total error

Error of workpiece installation in the membrane cartridge, $\Delta\varepsilon_y$	5 μm
Error of the technological system adjustment, $\Delta_н$	4 μm

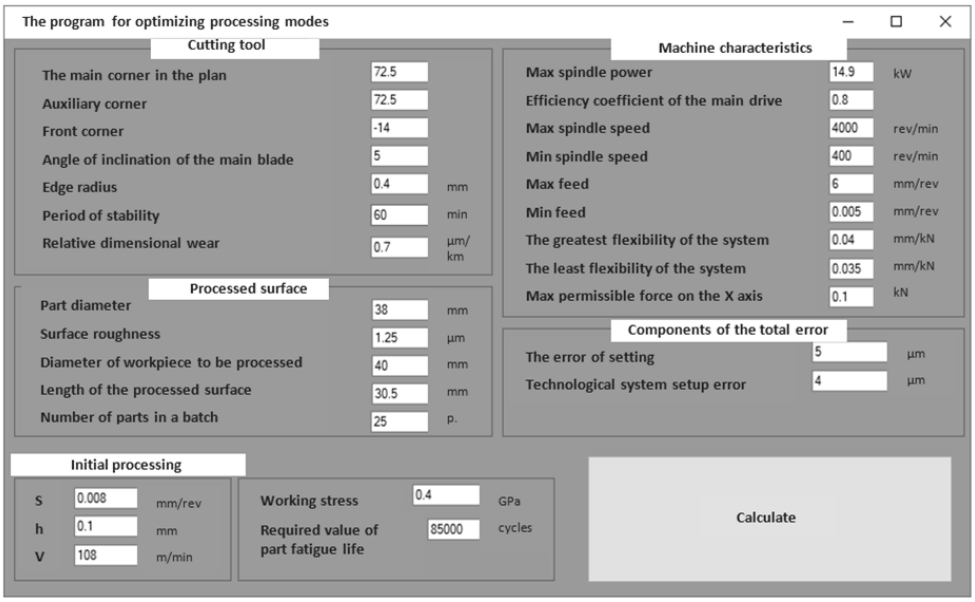


Fig. 4. The main window of the program optimal cutting conditions definition with the input data

4. Conclusions

1. The method of technological support of fatigue life of a part working in conditions of alternating loads is developed, by solving the problem of optimization of turning processing regimes, which allows to determine in a computerized mode the cutting mode which gives the maximum value of the productivity of the process of the finishing turning of the part in the necessary for its cyclic durability in the field of admissible solutions and takes into account the actual characteristics of the material of the part.
2. The sequence of the methodology implementation for providing the fatigue life of the parts at the stage of preproduction planning, which allows us to determine part's optimal cutting condition, is presented.
3. The application of the finite element method allows us to determine the location of localization and the values of stresses and deformations that arise under the action of loads during the operation of parts, and to use them in solving the problem of cutting condition optimization.
4. An example of practical use of the methodology of technological support of the necessary value of cyclic durability and maximum productivity of manufacturing of the "Shaft" component was given, which allowed to determine the rational regimes of its turning process, which are: S = 0.138 mm/rev; V = 187.6 mm/min; h = 0.1 mm, providing the required fatigue life of at least N = 16·104 cycles, which satisfies the requirements of the product developers regarding the work life of part , and the maximum productivity of manufacturing.

References

[1] Barandych K. S., Vysloukh S. P.: Technological Supportof Parts' Fatigue Life by Modeling Their Turning Process. Naukovi Visti NTUU KPI 2, 2018, 61–69.
[2] Barandych K. S, Vysloukh S. P, Antonyuk V. S.: Ensuring Fatigue Life of Parts During Finish Turning with Cubic Boron Nitride Tools. Journal of Superhard Materials 3(40), 2018, 206–215.
[3] Barandych C. S., Vysloukh S. P.: Generating finite-element model of the shaft and settlement boundary value problem of stress-strain state. Zbirnyk Naukovykh Prats', Seriya: Haluzeve Mashynobuduvannya, Budivnytstvo 2, 2014, 228–232.
[4] Bernyk P. S. et al.: Tekhnolohichni metody zabezpechennya nadiynosti mashyn. Kyiv 2004.
[5] Javidi A.: Influence of Machining on the Surface Integrity and Fatigue Strength of 34CrNiMo6 Steel. Thesis in partial fulfilment of the requirements for the degree of a Doctor of ontanistic Sciences (Dr. mont.), 2008.
[6] Karpenko H. V. et al.: Malotsyklovaya ustalost' staly v rabochykh sreda. Kyiv 1977.
[7] 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.
[8] Lavrenko Ya. I.: Do pitannya viznachennya resursu konstruktivnykh yelementiv pri zminnykh navantazhennyakh, Visnyk NTUU "KPI" Seriia Mashinostroyeniye 56, 2009, 88–92.
[9] Melnychuk P. P. et al.: Tekhnolohyya mashynobuduvannya, Zhytomyr 2005.
[10] M'Saoubi R., et al.: A review of surface integrity in machining and its impact on functional performance and life of machined products. Int. J. Sustainable Manufacturing 1, 2008, 203–236.
[11] Polishchuk L. Bilyy O., Kharchenko Y.: Prediction of the propagation of crack-like defects in profile elements of the boom of stack discharge conveyor. Eastern-European Journal of Enterprise Technologies 1(6), 2016, 44–52.
[12] Polishchuk L., Mamyrbayev O., Gromaszek K.: Mechatronic Systems 2: Applications in Material Handling Processes and Robotics (1st ed.). Routledge, 2021 [http://doi.org/10.1201/9781003225447].
[13] Pramanik A. et al.: Littlefair Fatigue life of machined components. Adv. Manuf. 5, 2017, 59–76 [http://doi.org/10.1007/s40436-016-0168-z].
[14] Pysarenko V. H., Bernyk P. S., Syvak R. I.: Tekhnolohichni metody pidvyshchennya nadiynosti detaley mashyn, Vinnytsia 2008.
[15] Vislouh S. P.: Informatsioni tekhnologii v zadachakh tekhnologichnoyi pidgotovki priladota Mashinobudivnogo virobnitstva. Kyiv 2011.
[16] Wójcik W., Pavlov S., Kalimoldayev M.: Mechatronic Systems 1: Applications in Transport, Logistics, Diagnostics, and Control (1st ed.). Routledge, 2021 [http://doi.org/10.1201/9781003224136]

Ph.D. Kateryna Barandych
e-mail: Barandichk@ukr.net

Ph.D. of Engineering Sciences, docent, associate professor at the Department of Computer-integrated technologies of device production National technical university of Ukraine „Igor Sikorsky Kyiv Polytechnic institute". She has published more than 90 scientific works. The main direction of her scientific activity is technological support of components performance characteristics.

http://orcid.org/0000-0003-0331-3216

Ph.D. Sergii Vysloukh
e-mail: vsp1@ukr.net

Ph.D. of Engineering Sciences, docent, associate professor at the Department of Computer-integrated technologies of device production National technical university of Ukraine „Igor Sikorsky Kyiv polytechnic institute". He published more than 150 scientific works. The main directions of scientific activity: information technologies in engineering and medicine, mathematical modeling of processes and systems, optimization of processes and systems, data analysis.

http://orcid.org/0000-0002-2204-2602

D.Sc. Eng. Grygoriy Tymchyk
e-mail: deanpb@kpi.ua

Doctor of Engineering Sciences, professor, Dean of the Instrumentation Engineering Faculty National technical university of Ukraine „Igor Sikorsky Kyiv polytechnic institute". He published more than 570 scientific works. The main directions of scientific activity: creation of computer-integrated, laser, optoelectronic, vibroacoustic and electromagnetic systems for diagnosing the state of technological equipment and managing product quality.

http://orcid.org/0000-0003-1079-998X

Ph.D. Oleksandr Murashchenko
e-mail: oleksandr.murashchenko@gmail.com

Candidate of Science (Engineering), associate professor, Department of Computing Equipment, Vinnytsia National Technical University. He has published 26 scientific papers.
Research interests include Digital to Analog Converters (DAC) with weight redundancy, usage of redundant positional number systems in order to reduce glitches in DAC; mathematical modeling of glitches during digital-to-analog conversion.

http://orcid.org/0009-0005-9314-2095

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 confer ence 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

Ph.D. Saule Kumargazhanova
e-mail: SKumargazhanova@gmail.com

She is associate professor on 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.

http://orcid.org/0000-0002-6744-4023

http://doi.org/10.35784/iapgos.5328 *received: 11.07.2023 | revised: 19.08.2023 | accepted: 05.09.2023 | available online: 30.09.2023*

RESEARCH THE EFFECT OF THE FRACTIONAL NUMBER SLOTS OF POLE ON WIND TURBINE GENERATION USING THE ENHANCED SPOTTED HYENA OPTIMIZATION ALGORITHM

Ibrahim M. Aladwan[1], Hasan Abdelrazzaq AL Dabbas[2], Ayman. M. Maqableh[3], Sayel M. Fayyad[4], Oleksandr Miroshnyk[5], Taras Shchur[6], Vadym Ptashnyk[7]

[1]Al-Balqa Applied University, Department of Mechatronics Engineering, Al Salt, Jordan, [2]Philadelphia University, Department of Mechanical Engineering, Amman, Jordan, [3]Luminus Technical University College, Electomechanical Engineering Department, Amman, Jordan, [4]Al-Balqa Applied University, Department of Mechanical Engineering, Al Salt, Jordan, [5]State Biotechnological University, Department of Electricity Supply and Energy Management, Kharkiv, Ukraine, [6]Cyclone Manufacturing Inc, Mississauga, Ontario, Canada, [7]Lviv National Environmental University, Department of Information Systems and Technologies, Lviv, Ukraine

Abstract. *The design of machines with permanent magnets is actively developing day by day and is often used in wind energy. The main advantages of such variable speed drives are high efficiency, high power density and torque density. When designing a wind generator with two rotors and permanent magnets, it is necessary to solve such a problem as the correct choice of the number of poles and slots to increase efficiency and minimize the cost of the machine. In this work, an improved spotted hyena optimization algorithm is used to obtain the optimal combination of slots and poles. This optimization algorithm makes it possible to obtain the number of fractional slots per pole and evaluate the operating efficiency of a wind generator with a double rotor and ferrite magnets. At the first stage of machine design, various combinations of slots are installed. Next, the optimal combination is selected from various slot-pole combinations, taking into account the Enhanced Spotted Hyena Optimization (ESHO) algorithm, in which a multi-objective function is configured. Accordingly, the multi-objectives are the integration of reverse electromotive force, output torque, gear torque, flux linkage, torque ripple along with losses. Analysis of the results obtained shows that the proposed algorithm for determining the optimal slot combination is more efficient than other slot combinations. It has also been found that the choice of slot and pole combination is critical to the efficient operation of permanent magnet machines.*

Keywords: wind turbine generation, optimal slot, pole, ESHO algorithm

BADANIE WPŁYWU UŁAMKOWEJ LICZBY SZCZELIN BIEGUNÓW NA GENERACJĘ TURBINY WIATROWEJ PRZY UŻYCIU ULEPSZONEGO ALGORYTMU OPTYMALIZACJI CĘTKOWANEJ HIENY

Streszczenie. *Projektowanie maszyn z magnesami trwałymi aktywnie rozwija się z dnia na dzień i jest często wykorzystywane w energetyce wiatrowej. Głównymi zaletami takich napędów o zmiennej prędkości są wysoka sprawność, wysoka gęstość mocy i gęstość momentu obrotowego. Podczas projektowania generatora wiatrowego z dwoma wirnikami i magnesami trwałymi konieczne jest rozwiązanie takiego problemu, jak prawidłowy dobór liczby biegunów i szczelin w celu zwiększenia wydajności i zminimalizowania kosztów maszyny. W niniejszej pracy zastosowano ulepszony algorytm optymalizacji hieny plamistej w celu uzyskania optymalnej kombinacji szczelin i biegunów. Ten algorytm optymalizacji umożliwia uzyskanie liczby ułamkowych szczelin na biegun i ocenę wydajności operacyjnej generatora wiatrowego z podwójnym wirnikiem i magnesami ferrytowymi. Na pierwszym etapie projektowania maszyny instalowane są różne kombinacje szczelin. Następnie wybierana jest optymalna kombinacja spośród różnych kombinacji szczelin i biegunów, biorąc pod uwagę algorytm Enhanced Spotted Hyena Optimization (ESHO) (ulepszony algorytm optymalizacjihieny cętkowanej hieny), w którym skonfigurowana jest funkcja wielocelowa. W związku z tym, celami wielozadaniowymi są integracja odwrotnej siły elektromotorycznej, wyjściowego momentu obrotowego, momentu obrotowego przekładni, połączenia strumienia, tętnienia momentu obrotowego wraz ze stratami. Analiza uzyskanych wyników pokazuje, że proponowany algorytm określania optymalnej kombinacji szczelin jest bardziej wydajny niż inne kombinacje szczelin. Stwierdzono również, że wybór kombinacji szczelin i biegunów ma kluczowe znaczenie dla wydajnej pracy maszyn z magnesami trwałymi.*

Słowa kluczowe: generacja turbiny wiatrowej, optymalna szczelina, biegun, algorytm ESHO

Introduction

Permanent magnet (PM) motors are intensively studied during the last few years [5, 13, 14, 19–21]. Presently, concerning the higher power/torque density along with high efficiency, fractional slot PM machines (FSPMMs) are extensively engaged in industrial automation, electrical vehicle and power generation including several other applications [14, 19, 21]. Basically, the PM machines (PMM) are modelled in a manner so that in contrast with the magnets, the air gap flux density (FD) regarding the armature reaction is lower. This will avert the PM's demagnetization in the event of defect currents in the stator windings [20]. Even though conventional fractional slot PMM having the entire teeth wound windings demonstrates better effectiveness, the coils localized at the final part of every single section could be effortlessly spoiled by simply being in the open air [5]. Alternatively, the high torque requisite on direct drive PM machines constantly results in a larger machine size together with material consumption [13]. PMMs have been broadly utilized in numerous applications. The PMMs are established as a generator in a wind turbine (WT) in the perspective of renewable energy systems [23, 27]. With specific reference to vehicle traction drives, and the direct-drive wind power generators, the consequence of fractional-slot of winding (FSCW) PMMs is concentrated broadly and is spotlighted here [25]. The entire functions necessitate high torque quality [12, 18]. Naturally, an electric machine is somewhat a conventionalist drive part that is improving slower than electronic components along with program control logic. This unit is the motor that verifies the drive's energetic data and also the weight-size parameters completely and to a great extent and that's why it is extremely essential to optimize this unit [8]. FEA software is utilized in the procedure of modelling the machine [24, 29, 30]. The well-known models of FEA tools are QuickField, Altair Flux, and as well FEMM. The first two models are proprietary software and the last model is free of software [31]. Rare-Earth (RE), also the non-RE, is the PM material's classification. Terbiums (Tb), Samarium (Sm), Dysprosium (Dy) along with Neodymium (Nd) are certain RE Elements (REEs) of RE-PMs. The two major categories of non-RE PMs are Alnico and Ferrite (Fe), which date to the earlier 1930s and 1950s, respectively [40]. Proportionate to the proficiency and the consistency, the PMSM could be preferred. The difficulties in PMSM are the high price and the flux's weakening on limiting the electric loading [22]. PMSMs could be organized as Surface-mounted PM Machines (SPMs) as well Interior PM Machines (IPMs) regarding the PM's positions [34]. The features and the proficiency of synchronous SPM machines are extremely exaggerated by the slot-pole combination [7]. The consequence of phase number on the winding factors, CT frequency and net radial forces are systematically scrutinized in various previous researches, which offers guiding principles to choosing the optimal slot/pole number combinations for multi-phase PMM [10]. The numeral count of stator slots per rotor poles per phase ratio directly influences

the PMSM's proficiency [1, 4, 33]. Subsequently, the foremost step in the PMM is selecting the slot-pole combinations. In this study, the outcome of the fractional number of SPP in PMM is being evaluated by choosing the slot-pole.

The authors of the article [33] evaluated the consequence of various pole-slots combinations on the generator's performance and at last established a set of methods that are comparatively better. The simulation outcomes validated that this methodology efficiently minimized the higher harmonic of no-load counter electromotive force in addition to that it enhanced the induced voltage distortion rate along with torque vibration and considerably enhanced the proficiency of the motor [37, 38]. And this scheme was utilized as the guiding signification for the structural optimization of superconducting motors.

In the paper [3] proffered a fresh modular fractional slot PMM having redundant teeth. The manufacturing of modular PMM with 42-slots/32-poles (42S/32P) combination was detailed elaborately like an instance, and that was handled as the amalgamation of six 6S/5P segments as well as a redundant 6S/2P device. Extra security was offered for coils positioned in these areas because the end part of every single section in the modular PMM was enclosed by half a redundant tooth in production. Progressively, the machine's fault-tolerant capacity was enhanced by the partition of six segments. By deploying the dual three-phase winding the modular machine's efficacy was improved and it was established by the electromagnetic performance fore-casted by FEA.

At paper [6] executed an absolute sensitivity exploration of the optimal parameters for the axial flux PM synchronous machines running in the field weakening (FW) area. The two design purposes of the methodology were to enlarge the power density (Power density is the ratio of fraction of maximal to rated speed: n_{max}/n_r and the associated inductance parameter maintaining the proficiency at the aimed speed beyond 90%) and to deliberate the optimal combination for every single phase, various slots/poles/phases combinations were deliberated. The outcome of the ratio of a number of stator slots to the number of rotor poles on the Power density and the nmax/nr was also calculated. It was exposed that the bettered Power density was attained by the factors having the low values. Regarding the "2" design goals, the outcome of the ratio of the outer diameter as well the inner to the outer diameter was illustrated. Consequently, the finite, also the theoretical infinite speed designs, were contrasted. The analytical design's potency was recognized by studying the entire 3D finite element (FE).

Autor of paper [35] proffered Fractional Slot PM (FSPM) machines by the fractional numbers of the slots for each part in a single periodicity of the machines. Alter-natively, the phase back-EMF signals were well-proportioned. Furthermore, the multi-layer winding methodology was deployed with the target of attaining a slot/pole com-bination by a small Unbalance Magnetic Force (UMF) also the armature MMRF's low THD value. Additionally, for discovering the multilayer winding's arrangement, a simple methodology was proffered by utilizing an extremely small UMF.

At work [15] examined the interior PM (IPM) machine's terminal voltage distortion by fractional slot concentrated windings (FSCW), also a specific prominence on the outcome of stator slot (Ns) together with rotor pole ($2p$) combinations. Initially, the 12-slot/10-pole machine was deployed for certifying the happening of voltage distortion. The methodology was then scrutinized by deploying the frozen permeability (FP) technique. The outcome of this examination exposed that the disparity in flux paths owing to rotor saliency was the cause of such a phenomenon, particularly whilst the present advancing angle reached 90°. Consequently, regarding the unavoidable device saturation, a design trade-off was utilized by choosing appropriate Ns/2p combinations for minimizing the effect. At last, for certifying the evaluation, prototypes were formulated.

In the paper [17] proffered an 18-slot/26-pole FSPM surface-mounted device (FSPMSM) having a coil-pitch of 2 slot-pitches. The machine's torque creation was done taking into account the fractional slot machine's principle together with the magnetic gear's outcome. Its proficiency was evaluated and correlated by an 18-slot/10-pole FSPMSM. The outcomes confirmed that the 18-slot/26-pole FSPMSM had better flux weakening capacity, smaller torque ripple, larger torque along with torque density, also a higher efficacy region in contrast to its 10-pole counterpart. Additionally, the possible slot/pole combinations of FSPMSMs having coil-pitch of 2 slot-pitches regarding the similar working rule were employed [2, 9, 36, 39].

1. Materials and methods

The PM generator that is directly operated by WT has the remunerations of higher efficiency, the simple structure and the reliable operation, also the characteristics of multi-pole, low speed, and large size. The disputes to be resolved while structuring a direct-driven PM wind generator are sufficient utilization of the structural dimension, enhancing the proficiency, and minimizing the machine's price by appropriate selection of the pole number and the Ns number [11, 16, 28]. The consequence of the fractional number of SPP and the optimal slot-pole combination in WT PMM is evaluated in this research. In the 1st phase, the wind generator and their magnetic pressure is obtained. Next, the slot-pole combination is initiated. Afterward, by deploying the ESHO algorithm the optimal slots along with poles combination are chosen. Figure 1 displays the present technique's block diagram [26, 41].

Fig. 1. Block diagram for the presented research [16]

1.1. Design of wind turbine generator

The WT is modelled to produce the highest power at a large spectrum of wind speeds. Nevertheless, during initialization, the significant model, and the decision in selecting the appropriate site meant for a WT. The WT's mechanical output power (O_p) is specified as

$$Q_p = \frac{1}{2}\rho\pi L^3 P_C(\omega,\delta)u_w^3 \tag{1}$$

The ω in equation (1) is the proportion of the blade tip's speed to the wind's speed which is expressed as

$$\omega = \frac{\eta_t L}{u_w} \tag{2}$$

where ρ – the air density in (kg/m^3); L – is the radius of the rotor's swept area (m); P_c – the power coefficient. This coefficient is a function of ω also the blade's pitch angle δ in (deg). The power coefficient is proffered as the division of the turbine mechanical power by the power accessed by the wind; u_w – the wind speed (m/sec); η_t – denotes the rotor angular speed (rad/sec) together with that the WT's aerodynamic mechanical torque (Mt) which is formulated as

$$M_t = \frac{\frac{1}{2}\rho\pi L^3 P_C(\omega,\delta)u_w^3}{\omega} \tag{3}$$

In DRPM wind generators, between 2 concentric rotors, the up stator is inserted, along with that the surface-mounted FM is extensively polarized in various directions. For that reason, the magnetic circuits in the DRPM wind generator's inner as well as in outer parts are in parallel, with that they divide up a common stator yoke. The circularly wounded winding is implemented, hence the insulation of the layer inside the slot isn't required having the benefits like less copper loss, easy maintenance, and as well short end winding. The fluxes flowing via the inner and the outer parts are similar, accordingly, the DRPM wind generator can be witnessed as an inner rotor generator occupied within an outer rotor one with single-layer winding.

In the DRPM WT generator, a magnetic pressure is deployed. As of the magnetic flux (MF) density, the electromagnetic pressure is measured utilizing the Maxwell stress tensor. The radial pressure brings out a vibration response, for this reason, the peripheral elements of the MF density are devalued. The radial pressure is formulated as:

$$R_r(\theta,t) \backsimeq \frac{\beta_r(\theta,t)}{2\gamma_0} \tag{4}$$

where β_r – denotes the radial MF regarding and its determined using the following expression

$$\beta_r(\theta,t) = \Lambda(\theta,t)F_r(\theta,t) \tag{5}$$

The MMF $F_r(\theta,t)$ is presented as

$$F_r(\theta,t) = \sum_{k=1}^{\infty} A_k \cos((\varepsilon_k t - kl\theta) \tag{6}$$

where θ – the tangential position; t – time; R_r – the radial pressure; γ_0 – the magnetic permeability of the air.

The MF is then obtained by multiplying the MMF by the air-gap permeance: $\Lambda_{slot}(\theta,t)$ – specifies the air-gap permeance; $F_r(\theta,t)$ – indicates the magnetomotive force (MMF); A_k – denotes the k-th MMF wave's amplitude; ε_k – the rotational frequency (rad/sec), which is $\varepsilon_k = k{\cdot}l{\cdot}v$ for the k-th harmonic; v – the rotor's rotational frequency (rad/sec); l – the number of pole pairs.

If the machine's slots are considered, the air-gap permeance variation pertained to the slots must be regarded, which is generally expressed by a Fourier series:

$$\wedge_{slot}\ (\theta,t) = \wedge_0 + \sum_{k=1}^{\infty} \wedge_h \cos(hS_2\theta) \tag{7}$$

So, the MF density is derived by substituting equation (7) in equation (5)

$$\beta_r(\theta,t)|_{ecc=\xi} = F_r(\theta,t)[\wedge_0 + \sum_{k=1}^{\infty} \wedge_h \cos(hS_2\theta)] \tag{8}$$

where h – the harmonic; S_2 – the number slots; Λ_0 – the air-gap permeance without slots; Λ_h – the coefficients of the Fourier series.

If the outcome of the slots is added, a novel spatial harmonic is attained consequently by the amalgamation of the MMF's harmonics and the slots. Consequently, the harmonics of spatial orders $k{\cdot}l \pm h{\cdot}s$ are attained, and the harmonics of the MMF of spatial order $k{\cdot}l$. In equation (4), the radial pressure is attained subsequently the MF's harmonics are added to and subtracted from one another for indicating the pressure's harmonics

1.2. Initialize the slot-pole combination

This stage initiates the number of slot-pole combinations. The number *SPP* per phase SPP_q is a fraction and not an integer, moreover this information is signified as "fractional slot winding". The total phases are denoted by *m*. The doubling-up of these phases is contrasted to a three-phase system that reduces *SPPq* into half. It relied upon the number of slots S_s and the number of pole pairs *l*.

$$SPP_q = \frac{S_s}{2lm} \tag{9}$$

A six-phase machine is nothing but an amalgamation of two three-phase systems. A division of the combinations to three-phase windings is the feasible combination of S_s and the number of poles 2*l*. In this work, the DRPM wind generator's slot as well the pole combination is selected like 12-slot/10-pole, 12-slot/14-pole, 18-slot/16-pole, 18-slot/20-pole, 24-slot/22-pole, and 24-slot/26-pole.

1.3. Optimal slot-pole combination selection using ESHO algorithm

Various slot-pole combinations are initiated and then by deploying the ESHO methodology the optimal slot-pole combination is elected. This selection process is very significant for developing the DRPM WT generator. Initially, the population is initiated. In this, the initial population is the slot-pole combination which is specified as (Z_i, i = 1, 2, 3, … n). The fitness is estimated after the initialization. The fitness is measured regarding the amalgamation of CT, EMF, OT, torque pulsation, FL and losses.

If the phase currents are zero a torque occurs between stator slots and rotor magnets, and this torque is termed CT. This influences vibrations, torque fluctuations, along with acoustic noise in the machine. The generator's proficiency as well as their lifetime may get reduced because of these negative features. Moreover, in WTs, the huge CT with low speed controls the rotor's rotation therefore the electricity production is prevented. Henceforth, in situations like this, the lessening of these effects is essential; by decreasing the CT's value. CT(J_{cog}) with the variables of rotor's position and air gap can be expressed mathematically with Equation:

$$I_{cog} = \frac{1}{2}\varpi_{air}^2 \frac{dw}{d\theta} \tag{10}$$

where w – the air gap reluctance; $\bar{\omega}_{air}$ – the air gap flux amount; θ – the rotor's position.

The CT's value is a function of the change in the air gap reluctance. The air gap value changes at regular intervals, subsequently, the CT's value changes as well. In this circumstance, CT is expressed with Fourier series and Equation:

$$I_{cog} = \sum_{k=1}^{\infty} J_{mk} \sin(hzk\theta) \tag{11}$$

where J_{mk} – the Fourier coefficient; h – the frequency; z – the least multiple of 2p number; k = 1, 2, 3 cycle number; θ – the rotor position.

Faraday's law of electromagnetic induction is deployed for assessing the back EMF in relation to the time variation of MF (12). Through the MF density (13) the MF is attained on the cross-section surface in the coil, which has to be vertical to the MF path

$$e_f = -N_t \frac{d\Omega}{dt} \tag{12}$$

$$\Omega = \iint BB_t ds \tag{13}$$

where e_f – back-electromotive force; N_t – the number of turns; $\frac{d\Omega}{dt}$ – the MF's time variation inside the coil; BB_t – the flux density.

While measuring the MF's time variation the constant angular velocity is being considered.

The complete power of the nacelle and the rotor hub must be favoured by the turbine tower, thus large PMGs are not desired in WT design. Nevertheless, the torque (and output power) may be augmented by other means. The sizing constant K is expressed as

$$K = \vartheta K_{wl}\left(a_{BB_t}\right)A_{ll} \tag{14}$$

where ϑ – the constant parameter; K_{wl} – the basic harmonic winding factor; a_{BB_t} – the rotor surface's average FD; A_{ll} – the electrical loading.

An unstable torque pulsation is formed by the CT's combination along with torque. Consequently, undesirable and additional vibrations along with noise have occurred while operating the machine. The mathematical derivation of torque pulsation J_p is

$$I_p = J_{cog} + J_{rip} \tag{15}$$

where J_{rip} – the torque ripple.

The interaction of a magnetic field with a material such as what would take place when a magnetic field goes via a coil of the wire causes FL. FL is proffered by the number of windings and the flux, where the instantaneous value of a time-varying flux is specified as $\vec{f}$. The $\vec{f}$ equation is specified as

$$\vec{f} = \overrightarrow{Nf} \tag{16}$$

where $\overrightarrow{Nf}$ – the total flux. In equation (17) the metrics combination is presented:

$$F_{it} = j_{cog} + e_f + K + J_p + PR_{gr} + \vec{f} \quad (17)$$

where F_{it} – the fitness function.

The encircling phase is processed after the fitness is being evaluated. Other searching factors will be considering their best position pertained to the prey or target as the best response and revise it. This action is mathematically given as:

$$A_y = \left|\Psi \cdot E_p(x) - \varsigma\left(E(x)\right)\right| \quad (18)$$

$$E_p(x+1) = E_p(x) - \tau \cdot A_y \quad (19)$$

where A_y – the distance between the spotted hyena and the prey; E_p – indicates the vector position pertained to the prey; E – the spotted hyena's vector position; x – the current iteration; ς – specifies the position vector's mean value that averts the convergence problem; ψ, τ – the coefficient factor vectors:

$$\Psi = 2 \cdot rand_1 \quad (20)$$

$$\tau = 2 \cdot g \cdot rand_2 - g \quad (21)$$

$$g = 5 - \left(Iteration \cdot \frac{s}{MAX_{Iteration}}\right) \quad (22)$$

g – the declines linearly from 5 to 0 to stabilize the exploration and the exploitation while having the maximum iterations in the process. This system enhances the development as the iterations $MAX_{iteration}$ increase. Where $rand_1$ and $rand_2$ are random vectors in [0, 1].

Next, the hunting process is executed. While hunting, the hyenas get information and the prey tracking capacity from their trusted friends. The hyena that has the finest statistics regarding the prey's position is proffered as the best search agent. Consequently, the best search agent is pursued by the other agents to revise their positions

$$A_y = \left|\Psi \cdot E_y - E_x\right| \quad (23)$$

$$E_k = E_y + E_{y+1} + \cdots E_{k+y} \quad (24)$$

$$I_k = E_k + E_{k+1} + \cdots E_{k+N} \quad (25)$$

where E_y – the spotted hyena's best position regarding the prey; E_k – together with this the spotted hyena's other position; N – the spotted hyena's total number, and then it is measured as:

$$N = Count_{nos}\left(E_y + E_{y+1} + \cdots E_{y+LO}\right) \quad (26)$$

LO – the denotes a random vector having a range of [0.5, 1]; nos – the count of answers, (the reference answers are counted); E_y – the group of N optimal answers.

Next, the attacking prey phase is processed. The vector value y is minimized for denoting the attacking prey mathematically. The differentiation in the vector τ is also minimized for changing the value in vector y which can lessen from 5 to 0 during iterations. The swarm of spotted hyenas hit the prey when $|\tau| < 1$. The attacking prey phase can mathematically be formulated as:

$$E(x+1) = \frac{I_y}{N} \quad (27)$$

where $E(x+1)$ – revises the other search agent's position to conclude the best search agent's position and the best solution.

The SHO algorithm permits its search agents to revise their positions and the attack towards the prey.

The search for prey is the closing phase. The exact answer is detected if $\tau \geq 1$, regarding equation (26). Vector ψ is the section of the SHO methodology where exploration is possible. Vector ψ holds random values that offer the prey's random weights regarding the equation (27). If the $\psi > 1$, vector has priority over the $\psi<1$ vector to display more random features of the SHO methodology and the distance's effect. Subsequently, the fitness value is being calculated. If the pre-specified threshold of the fitness value is not attained, then the process is continued till it attains the pre-specified threshold value. The optimal slot-pole combination is elected from the initialized slot-pole combinations by employing this methodology.

The pseudo-code for the ESHO methodology is represented in figure 2. In this pseudo-code, the initialization of the slot-pole combinations, fitness evaluation, and the updating procedure is given.

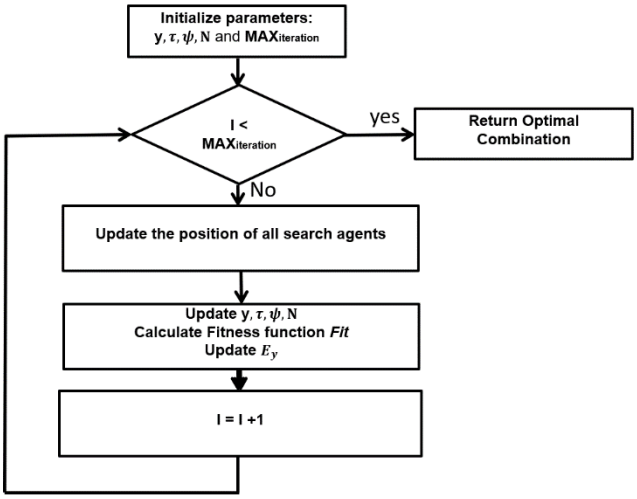


Fig. 2. Flow chart of the ESHO algorithm

2. Results and discussion

The slot-pole combination's effect is assessed in this section. MATLAB Simulink platform is deployed for executing the proffered work.

2.1. Performance analysis

Proportionate to the CT, FL, losses, back-EMF and OT; the usefulness of the optimal selection and without optimal slot-pole combination selection is evaluated. While considering the initialized slot-pole combination, a better slot-pole combination is attained in the optimal selection. And the average value of the entire initialized slot-pole combination was attained without optimal selection.

The FL analysis with optimal selection and without optimal selection is presented in figure 3. FL is small and takes place in electric machines. Flux lines take the path of least reluctance in most of the parts. Reluctance is proffered as the resistance to the flow of flux lines. These materials have a higher permeability. The number of slot-pole combination's optimal selection has less FL leakage than the slot-pole combinations without optimal selection. The evaluation is done regarding the variation in the rotor's position. If the rotor position is 180, the optimal selection has 0.4 Wb FL whereas without optimal selection it has 0.9 Wb. Accordingly, it points out that the optimal se-lection from the initialized slot-pole combination offers the finer result than the usual selection process.

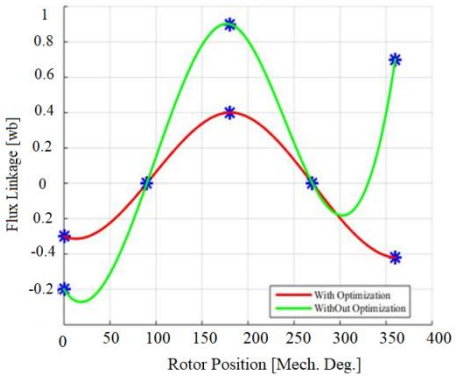


Fig. 3. FL analysis with optimal selection

CT of the slot-pole combination is presented in figure 4. It is significant to assess the CT. CT is one of the common challenges for machines. The harmonics in the back-EMF are pertained by the torque ripple in CT. Regarding the variation of the rotor position, the CT is evaluated. The optimal selection of slot-pole combination has less CT in correlation with the without

optimal selection option of the average value of the entire initialized slot-pole combination. For example, if the rotor position is 1, then with optimal selection-based slot-pole combination has 0.2 Nm whereas the without optimal selection has the 0.4 Nm. Correspondingly, for a range of rotor positions, with optimal selection has the finer outcome than the without optimal slot-pole combination.

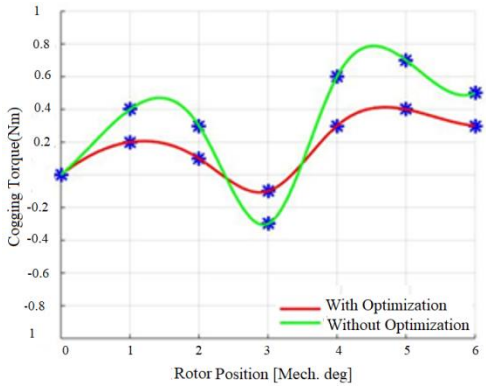


Fig. 4. Analysis of CT

The OT estimation of the slot-pole combination is assessed with optimum selection from the initialized slot-pole combination and the without optimal selection and then it is presented in figure 5. In general, the OT is centred on the rotor torque along with speed. Proportionate to the disparity in various electrical angles, the OT is evaluated. The model is specified as a better system if it has high OT. In this, the slot-pole combi-nation with optimal selection has higher OT than that without optimal selection. If the electrical angle is 360, with optimal selection has the 64 Nm OT and the without optimal selection has 43 Nm. The optimal slot-pole selection offers higher OT for the other electrical angles as well. Therefore, the best outcome is obtained using the recommended optimal selection.

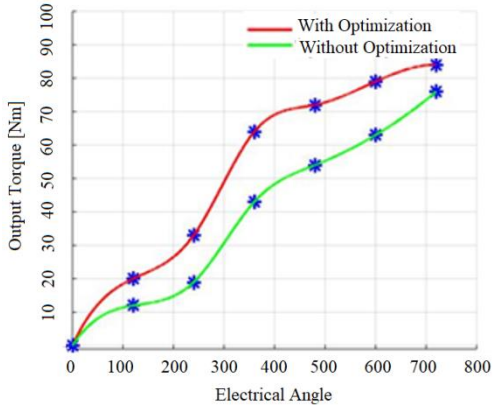


Fig. 5. Graphical representation of OT analysis

Figure 6 represents the losses regarding the slot-pole combination. The usefulness of the system and its reliability can be forecasted by having a better understanding of the features of the high-speed machine, particularly; a better understanding of its losses. The rotor losses could be high for high-speed machines, and in certain situations, these losses may harm the machine by the magnet's demagnetization. Hence, in machines, the assessment of losses is a major process. In relation to the rotor position angle, the loss is varied. In contrast, the with optimal slot-pole combination selection achieves lesser losses than the without optimal selection. For the rotor position 270, the optimal selection has 65 W whereas the without optimal slot-pole combination has the 78 W. A better result can also be obtained for the other rotor position by considering the with optimal slot-pole combinations.

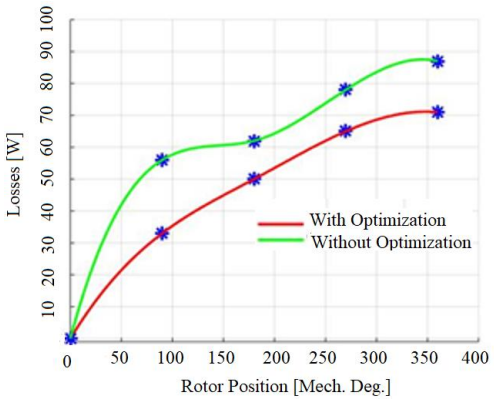


Fig. 6. Losses analysis

Figure 7 displays the Back-EMF of the slot-pole combination. In electric motors, the back EMF voltage occurs whilst there is a relative motion between the stator windings and the rotor's magnetic field. The figure of the back-EMF waveform is recognized by the rotor's geometrics features. In this, the efficiency is evaluated in relation to the rotors' position. For 180 rotor position, with optimal selection has 60 V whereas the without optimization has 80 V. Hence, a better proficiency is conquered with optimal selection.

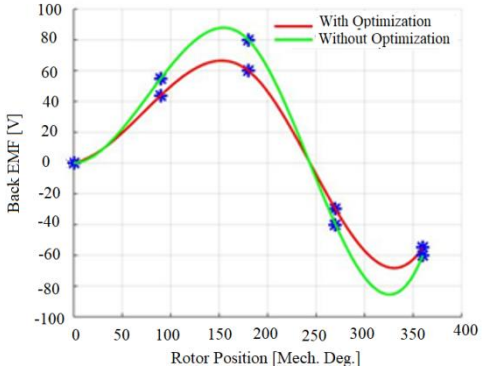


Fig. 7. Pictorial representation of the back-EMF analysis

The fitness vs iteration of the proffered ESHO methodology with the other methodologies, like SHO, Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), and Genetic Algorithm (GA) is represented in Fig. 8. For every single iteration, the fitness value varies. The proposed ESHO has 47 fitness for the 25th iteration. And for the same iteration, the fitness value for the previous methodologies is less than that of the proposed methodology. Likewise, for the 20th iteration, the ESHO has 34 fitness, whereas the previous methodologies have less fitness value than the proposed methodology. Subsequently, it is proved that a better outcome is achieved by the ESHO method-based selection than the without optimal selection and the other methods.

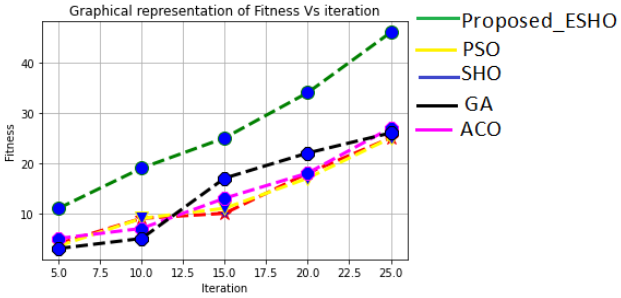


Fig. 8. Graphical representation of Fitness Vs iteration

3. Conclusions

The outcome of the SPP's fractional number in the DRPM WT generator with optimal slot-pole combination selection is assessed in this research. Mostly, small WTs utilize direct-driven PM generators, which have the characteristics of low speed and higher efficacy. Very simple controls are desired for small WTs as they are self-started. In the WT generator, the slot-pole combination obtains the main place and hence the significance of the SPP's fraction number. At first, the slot-pole combination is initialized and from that combination, the optimal selection is made by the ESHO. FLs, EMF, CT, OT, torque pulsation and losses are the features that are referred to during the selection process. The proficiency of the optimal slot-pole combination is evaluated with the entire initialized slot-pole combinations in relation to those factors. In this, the optimal selection of SPP's fraction number obtains better proficiency than the without optimal slot-pole combination selection. Thus, this method regarding the optimal slot-pole combination selection is crucial for the DRPM WT generator. In the future, the proposed methodology can be expanded by considering a variety of factors for a better optimal slot-pole combination selection.

References

[1] Abdelmoula R., Benhadj N., Chaieb M., Neji R.: Finite element comparative analysis software of a radial flux synchronous motor for electric vehicle drive. Proceedings of the International Conference on Recent Advances in Electrical Systems, 2016, 62–67.

[2] Al_Issa H. A., Qawaqzeh M., Khasawneh A., Buinyi R., Bezruchko V., Miroshnyk O.: Correct Cross-Section of Cable Screen in a Medium Voltage Collector Networkwith Isolated Neutral of a Wind Power Plant. Energies 14, 2021, 3026 [http://doi.org/10.3390/en14113026].

[3] Ambekar R., Ambekar S.: Design investigation for continual torque operative performance of PMSM for vehicle. Sādhanā 45, 2020, 120 [http://doi.org/10.1007/s12046-020-01360-y].

[4] Andrade K. M., Santos H. E., Wellington M. V., Almeida T. E., Paula G. T.: PeMSyn – a free matlab-femm based educational tool to assist the design and performance assessment of synchronous machines. Eletron. Poten., Fortaleza 25(2), 2020, 163–172 [http://doi.org/10.18618/REP.2020.2.0009].

[5] Chakir A., Tabaa M., Moutaouakkil F., Medromi H., Alami K.: Control System for a Permanent Magnet Wind Turbine Using Particle Swarm Optimization and Proportional Integral Controller. International Review of Automatic Control (IREACO) 13(5), 2020 [http://doi.org/10.15866/ireaco.v13i5.18482].

[6] Chen X., Wang J.: Magnetomotive force harmonic reduction techniques for fractional-slot non-overlapping winding configurations in permanent-magnet synchronous machines. Chinese Journal of Electrical Engineering 3(2), 2017, 103–113 [http://doi.org/10.23919/CJEE.2017.8048416].

[7] Demir Y., Yolacan E., El-Refaie A., Aydin M.: Investigation of Different Winding Configurations and Displacements of a Nine-Phase Permanent-Magnet-Synchronous Motor with Unbalanced AC Winding Structure. IEEE Transactions on Industry Applications 55(4), 2018, 3660–3670 [http://doi.org/10.1109/TIA.2019.2913156].

[8] Dutta R., Pouramin A., Rahman M.: A novel rotor topology for high-performance fractional slot concentrated winding interior machine. IEEE Transactions on Energy Conversion 36(2), 2020, 658–670 [http://doi.org/10.1109/TEC.2020.3030302].

[9] Edhah S., Alsawalhi J., Al-durra A.: Multi objective optimization design of fractional slot concentrated winding synchronous machines. IEEE Access 7, 2019, 162874–162882 [http://doi.org/10.1109/ACCESS.2019.2951023].

[10] Gandzha S., Sogrin A., Kiessh I.: The comparative analysis of electric machines with integer and fractional number of slots per pole and phase. Procedia Engineering 129, 2015, 408–414 [http://doi.org/10.1016/j.proeng.2015.12.137].

[11] Hemeida A., Taha M., Abdallh A., Vansompel H., Dupre L., Sergeant P.: Applicability of fractional slot axial flux synchronous machines in the field weakening region. IEEE Transactions on Energy Conversion 32(1), 2016, 111–121 [http://doi.org/10.1109/TEC.2016.2614011].

[12] Iegorov O., Iegorova O., Miroshnyk O., Savchenko O.: Improving the accuracy of determining the parameters of induction motors in transient starting modes. Energetika 66(1), 2020, 15–23 [http://doi.org/10.6001/energetika.v66i1.4295].

[13] Ismagilov F., Vavilov V., Yamalov I., Karimov R.: Fault-Tolerant Electric Motors with Permanent Magnets and Electromagnetic Shunting. International Review of Aerospace Engineering (IREASE) 13(2), 2020, 51–58 [http://doi.org/10.15866/irease.v13i2.17751].

[14] Kolsi H., Ben Hadj N., Chaieb M., Neji R.: Design of Permanent Magnet Synchronous Motor by Means of Power Density Optimization For e-Vehicle Applications. International Review on Modelling and Simulations (IREMOS) 15(3), 2022 [http://doi.org/10.15866/iremos.v15i3.21739].

[15] Li G., Ren B., Zhu Z.: Design guidelines for fractional slot multi-phase modular machines. IET Electric Power Applications 11(6), 2017, 1023–1031 [http://doi.org/10.1049/iet-epa.2016.0616].

[16] Li X., Zhu Z., Thomas A., Wu Z., Wu X.: Novel modular fractional slot machines with redundant teeth. IEEE Transactions on Magnetics 55(9), 2019, 1–10 [http://doi.org/10.1109/TMAG.2019.2918190].

[17] Liu Y., Zhu Z.: Electromagnetic performance comparison of 18-slot/26-pole and 18-slot/10-pole fractional slot surface-mounted machines. 20th International Conference on Electrical Machines and Systems (ICEMS) 2017, 11–14 [http://doi.org/10.1109/ICEMS.2017.8056383].

[18] Liu Y., Zhu Z.: Influence of gear ratio on the performance of fractional slot concentrated winding machines. IEEE Transactions on Industrial Electronics 66(10), 2019, 7593–7602 [http://doi.org/10.1109/TIE.2018.2885728].

[19] Lounthavong V., Sriwannarat W., Seangwong P., Siritaratiwat A., Khunkitti P.: Optimal Stator Design to Improve the Output Voltage of the Novel Three-Phase Doubly Salient Permanent Magnet Generator. International Journal on Energy Conversion (IRECON) 8(4), 2020, 118–125 [http://doi.org/10.15866/irecon.v8i4.19302].

[20] Makhad M., Zazi K., Zazi M., Loulijat A.: Smooth Super Twisting Sliding Mode Control for Permanent Magnet Synchronous Generator Based Wind Energy Conversion System. International Journal on Energy Conversion (IRECON) 8(5), 2020, 171–180 [http://doi.org/10.15866/irecon.v8i5.19362].

[21] Murali N., Mini V. P., Ushakumari S.: Modified V-Shaped Interior Permanent Magnet Synchronous Motor Drive for Electric Vehicle. International Review on Modelling and Simulations (IREMOS) 14(6), 2021 [http://doi.org/10.15866/iremos.v14i6.20884].

[22] Nur T., Mawar S.: Improvement of Cogging Torque Reduction by Combining the Magnet Edge Shaping and Dummy Slot in Stator Core of Fractional Slot Number in Permanent Magnet Machine. IOP Conference Series Materials Science and Engineering 807(1), 2020, 012023, 29–30 [http://doi.org/10.1088/1757-899X/807/1/012023].

[23] Ouiddir F., Benouzza N., Gherabi Z.: Stator Current Square Analysis to Discriminate Between Eccentricity and Demagnetization Faults in PMSMs. International Review of Electrical Engineering (IREE) 17(1), 2022, 11–19 [http://doi.org/10.15866/iree.v17i1.20950].

[24] Pazyi V., Miroshnyk O., Moroz O., Trunova I., Savchenko O, Halko S.: Analysis of technical condition diagnostics problems and monitoring of distribution electrical network modes from smart grid platform position. IEEE KhPI Week on Advanced Technology (KhPIWeek), 2020, 20168725, 57–60 [http://doi.org/10.1109/KhPIWeek51551.2020.9250080].

[25] Peng B., Wang X., Zhao W., Ren J.: Study on shaft voltage in fractional slot machine with different pole and slot number combinations. IEEE Transactions on Magnetics 55(6), 2019, 1–5 [http://doi.org/10.1109/TMAG.2019.2898566].

[26] Pezhman J., Taghipour S., Khoshtarash J.: Expansion of the feasible slot pole combinations in the fractional slot PM machines by applying three-slot pitch coils. IEEE Transactions on Energy Conversion 34(2), 2018, 993–999 [http://doi.org/10.1109/TEC.2016.2614011].

[27] Qawaqzeh M., Szafraniec A., Halko S., Miroshnyk O., Zharkov A.: Modelling of a household electricity supply system based on a wind power plant. Przegląd Elektrotechniczny 96, 2020, 36–40 [http://doi.org/10.15199/48.2020.11.08].

[28] Qawaqzeh M., Zaitsev R., Miroshnyk O., Kirichenko M., Danylchenko D., Zaitseva L.: High-voltage DC converter for solar power station. International journal of power electronics and drive system 11(4), 2020, 2135–2144 [http://doi.org/10.11591/ijpeds.v11.i4.pp2135-2144].

[29] Røkke A., Nilssen R.: Analytical calculation of yoke flux patterns in fractional-slot machines. IEEE Transactions on Magnetics 53(4), 2017, 1–9 [http://doi.org/10.1109/TMAG.2016.2623583].

[30] Savchenko O. A., Miroshnyk O. O., Dyubko S., Shchur T., Komada P., Mussabekov K.: Justification of ice melting capacity on 6-10 kV OPL distributing power networks based on fuzzy modeling. Przeglad Elektrotechniczny 95(5), 2019, 106–109.

[31] Shen J., Wang C., Miao D., Jin M., Shi D., Wang Y.: Analysis and optimization of a modular stator core with segmental teeth and solid back iron for pm electric machines. IEEE International Electric Machines & Drives Conference (IEMDC), 2011, 1270–1275 [http://doi.org/10.1109/IEMDC.2011.5994787].

[32] Szafraniec A., Halko S., Miroshnyk O., Figura R., Zharkov A., Vershkov O.: Magnetic field parameters mathematical modelling of windelectric heater. Przeglad elektrotechniczny 97(8), 2021, 36–41.

[33] Tahanian H., Aliahmadi M., Faiz J.: Ferrite Permanent Magnets in Electrical Machines: Opportunities and Challenges of a Non-Rare-Earth Alternative. IEEE Transactions on Magnetics 56(3), 2020, 1–20 [http://doi.org/900120.10.1109/TMAG.2019.2957468].

[34] Tessarolo A., Mezzarobba M., Barbini N.: Improved four-layer winding design for a 12-slot 10-pole machine using unequal tooth coils. 42nd Annual Conference of the IEEE Industrial Electronics Society – IECON 2016, 1686–1691 [http://doi.org/10.1109/IECON.2016.7793399].

[35] Torreggiani A., Bianchini C., Davoli M., Bellini A.: Design for Reliability: The Case of Fractional-Slot Surface Permanent-Magnet Machines. Energies 12, 2019, 1691 [http://doi.org/10.3390/en12091691].

[36] Torrent M., Perat J. I., Jiménez J. A.: Permanent Magnet Synchronous Motor with Different Rotor Structures for Traction Motor in High Speed Trains. Energies 11, 2018, 1549 [http://doi.org/10.3390/en11061549].

[37] Trunova I., Miroshnyk O., Savchenko O., Moroz O.: The perfection of motivational model for improvement of power supply quality with using the one-way analysis of variance. Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu 6, 2019, 163–168 [http://doi.org/10.29202/nvngu/2019-6/24].

[38] Tymchuk S., Miroshnyk O.: Assess electricity quality by means of fuzzy generalized index. Easternt-European Journal of enterprise technologies 3/4(75), 2015, 26–31 [http://doi.org/10.15587/1729-4061.2015.42484].

[39] Wang Q., Li Y., Deng G., Zhang H., Li Y., Xu J., Wang X.: Optimization Study of Poles-Slots Combination of Large Capacity Offshore HTS Wind Generator Based on Ansys Maxwel. Journal of Physics Conference Series 1754, 2021, 012042 [http://doi.org/10.1088/1742-6596/1754/1/012042].

[40] Zhu Z., Wu D., Ge X.: Investigation of voltage distortion in fractional slot interior machines having different slot and pole number combinations. IEEE Transactions on Energy Conversion 31(3), 2015, 1192–1201 [http://doi.org/10.1109/TEC.2016.2553140].

[41] Zou T., Qu R., Li D., Jiang D.: Synthesis of fractional-slot vernier permanent magnet machines. International Conference on Electrical Machines (ICEM) 2016, 911–917 [http://doi.org/10.1109/ICELMACH.2016.7732634].

Ph.D. Eng. Ibrahim M. Aladwan
e-mail: Ibrahim.aladwan@bau.edu.jo

Since 2014 he has been working at AL Balqa Applied University, Mechatronics Engineering Department as instructor for several courses related with Mechatronics Engineering. Dr. Ibrahim occupied several administrative positions, including Dean of Student Affairs, and Director of NTTI (National Training of Trainees Institute). In addition, Dr. Aladwan has a comprehensive training experience. His research interests include: Electric Drive, CNC machine, Renewable Energy.

http://orcid.org/0000-0002-7305-2413

Ph.D. Eng. Hasan Abdelrazzaq AL Dabbas
e-mail: drdabbas@engineer.com

Since 2014 he has been working at Philadelphia University, Faculty of Engineering, Department of Mechanical Engineering as instructor for several courses related with mechanical Engineering.
His research interests include: Materials Engineering, Mechanical Engineering Manufacturing Engineering, Mechanical Behavior of Materials Machining CNC Machining.

http://orcid.org/0000-0003-1301-4279

Ph.D. Eng. Ayman. M. Maqableh
e-mail: a.maqableh@ltuc.com

Maqableh has a Ph.D. in Mechanical Engineering from the University of Nottingham, UK. He has extensive experience in workforce development, technical and vocational education and public-private partnerships. He is currently the Dean of LTUC College-Jordan, managing more than 300 academic and administrative employees.

http://orcid.org/0000-0003-1301-4279

Ph.D Eng. Sayel M. Fayyad
e-mail: fswes@bau.edu.jo

Sayel M. Fayyad, holding Ph.D. in mechanical engineering, and B.A., Master degrees in mechanical engineering, computer science, education technology, and MBA, B.Sc. in mechanical engineering, physics and computer science. About 80 publications, working in Al-Balqa Applied University, Al Salt, Jordan since 2007 till now. Research work: control systems, vibrations, renewable energy and dynamics.

http://orcid.org/0000-0002-7305-2413

D.Sc. Eng. Oleksandr Miroshnyk
e-mail: omiroshnyk@btu.kharkiv.ua

Graduated from Kharkiv State Technical University of Agriculture, Kharkiv, Ukraine, in 2004, and was qualified as an Electrical Engineer. He received his Ph.D. in electric engineering (power stations, systems and networks) from the National Technical University "Kharkiv Politechnic Institute", Ukraine, in 2009 and Doctor of Technical Sciences degree in 2016.
His research interests are related to research in the field of smart grid, renewable energy and computer numerical control machines.

http://orcid.org/0000-0002-6144-7573

Ph.D. Eng. Taras Shchur
e-mail: shchurtg@gmail.com

Graduated from Kharkiv State Technical University of Agriculture, Kharkiv, Ukraine, in 2004, and was qualified as an mechanical engineer. He received his Ph.D. in mechanical engineering (Machines and means of mechanization of agricultural production) from the Kharkiv State Technical University of Agriculture, in 2009. His research interests are related to research in the field of smart grid, renewable energy and computer numerical control machines.

http://orcid.org/0000-0003-0205-032X

Ph.D. Vadym Ptashnyk
e-mail: ptashnykproject@gmail.com

Graduated from Lviv Polytechnic National University, Ukraine, in 2010, and was qualified as an applied physicist. He received his Ph.D. in ecological safety from the Sumy State University, Ukraine, in 2014.
His research interests are related to research in the field of Internet of Things, microprocessing technology and water quality control methods.

http://orcid.org/0000-0002-1018-1138

http://doi.org/10.35784/iapgos.5327 received: 26.07.2023 | revised: 16.08.2023 | accepted: 18.09.2023 | available online: 30.09.2023

NEW SURFACE REFLECTANCE MODEL WITH THE COMBINATION OF TWO CUBIC FUNCTIONS USAGE

Oleksandr Romanyuk[1], Yevhen Zavalniuk[1], Sergii Pavlov[1], Roman Chekhmestruk[2], Zlata Bondarenko[1], Tetiana Koval[3], Aliya Kalizhanova[4], Aigul Iskakova[5]

[1]Vinnytsia National Technical University, Vinnytsia, Ukraine, [2]3D GENERATION UA, Vinnytsia, Ukraine, [3]Mykhailo Kotsiubynskyi State Pedagogical University, Vinnytsia Ukraine, [4]University of Power Engineering and Telecommunications; Institute of Information and Computational Technologies MES CS RK, Almaty, Kazakhstan, [5]Kazakh National Research Technical University named after K. I. Satpayev, Almaty, Kazakhstan

Abstract*. In the article the model of light reflection based on the combination of two cubic bidirectional reflectance distribution functions is developed. The main components of color and the main requirements for reproducing the object`s glares are analyzed. The usage characteristics of Cook-Torrance, Bagher, Oren-Nayar, coupled Shirley reflection models are described. The advantages and disadvantages of the highly productive Blinn-Phong model are considered. The necessity of approximating the Blinn-Phong model by a function of low degree is justified. The characteristics of the cubic polynomial approximation of the Blinn-Phong model are determined. It was established that the main drawback of this approximation is a significant deviation of the function from the reference function in the glare's attenuation zone. The combined function that combines two cubic functions is proposed. The first cubic function reproduces the glare's epicenter, and the second replaces the specified function in the attenuation zone. A system of equations for calculating the coefficients of the second function was created. The formula for the connection point of two cubic functions is obtained. A graph of the developed combined model based on cubic functions is obtained. For the combined and original cubic functions a comparison of the maximum relative errors in the glare's epicenter zone, the maximum absolute errors, and the relative errors at the inflection point was made. A three-dimensional plot of the absolute error of the combined cubic model from the Blinn-Phong model depending on the shininess and the angle value is built. Visualization results based on the combined and the original cubic functions are compared. It is confirmed that the proposed reflection model increases the realism of glare formation in the attenuation zone. The resulting combined reflection model provides a highly accurate approximation of the Blinn-Phong model and is highly efficient because the third power function is used.*

Keywords: bidirectional reflectance distribution function, cubic function, reflectance model, shading, combined function

NOWY MODEL ODBICIA ŚWIATŁA OD POWIERZCHNI WYKORZYSTUJĄCY KOMBINACJĘ DWÓCH FUNKCJI SZEŚCIENNYCH

Streszczenie*. W artykule opracowano model odbicia światła oparty na kombinacji dwóch sześciennych dwukierunkowych funkcji rozkładu odbicia. Przeanalizowano główne składniki koloru i główne wymagania dotyczące odtwarzania odblasków obiektu. Opisano charakterystykę użytkowania modeli odbicia Cooka-Torrance'a, Baghera, Orena-Nayara i Shirleya. Rozważono zalety i wady wysoce wydajnego modelu Blinn-Phong. Uzasadniono konieczność aproksymacji modelu Blinna-Phonga funkcją niskiego stopnia. Określono charakterystykę wielomianu sześciennego aproksymującego model Blinna-Phonga. Ustalono, że główną wadą tej aproksymacji jest znaczne odchylenie funkcji od funkcji odniesienia w strefie tłumienia olśnienia. Zaproponowano funkcję kombinowaną, która łączy dwie funkcje sześcienne. Pierwsza funkcja sześcienna odtwarza epicentrum olśnienia, a druga zastępuje określoną funkcję w strefie tłumienia. Stworzono układ równań do obliczania współczynników drugiej funkcji. Uzyskano wzór na punkt połączenia dwóch funkcji sześciennych. Uzyskano wykres opracowanego połączonego modelu opartego na funkcjach sześciennych. Dla połączonych i oryginalnych funkcji sześciennych dokonano porównania maksymalnych błędów względnych w strefie epicentrum olśnienia, maksymalnych błędów bezwzględnych i błędów względnych w punkcie przegięcia. Zbudowano trójwymiarowy wykres błędu bezwzględnego połączonego modelu sześciennego z modelu Blinna-Phonga w zależności od połysku i wartości kąta. Porównano wyniki wizualizacji oparte na połączonych i oryginalnych funkcjach sześciennych. Potwierdzono, że proponowany model odbicia zwiększa realizm powstawania odblasków w strefie tłumienia. Wynikowy połączony model odbicia zapewnia bardzo dokładne przybliżenie modelu Blinna-Phonga i jest bardzo wydajny, ponieważ używana jest funkcja trzeciej potęgi.*

Słowa kluczowe: dwukierunkowa funkcja rozkładu współczynnika odbicia, funkcja sześcienna, model współczynnika odbicia, cieniowanie, funkcja łączona

Introduction

Modern three-dimensional rendering systems must meet the requirements of high performance and high realism. In order to create realistic three-dimensional graphic images, it is necessary to take into account the features of light reflection from the objects' surfaces in the scene. For this, during the surface shading, the three components of color [10, 11] are taken into account - ambient, diffuse and specular. The process of finding the specular [17] component of color is the most time-consuming, as it involves the calculation of normalized vectors to the observer, to the light source, and the normal vector. After this the lighting model, which has a high degree, is calculated and the color intensities for the RGB components are determined.

The main requirements for the reproduction of glares on the surfaces of objects are highly productive performance, highly accurate reproduction of the epicenter zone, and realistic reproduction of the attenuation zone. The existing models of light reflection do not comprehensively meet all the three requirements. Therefore, it is necessary to develop effective new models of the surface reflectivity, which provide a more highly productive implementation of rendering [15, 18].

1. Literature overview

The characteristics of the spatial reflection of light [14] from the surface are presented using the bidirectional reflectance distribution function (BRDF) [13]. BRDF is a function of the zenith and azimuthal angles (respectively, θ, ϕ) of the vectors to the light source and the observer (Fig. 1). In order to calculate the BRDF, the radiance of reflected light and the irradiance of light that fell on the surface are determined.

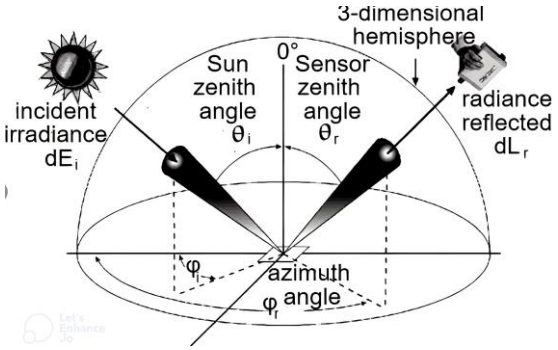


Fig. 1. Data for BRDF calculation [9]

BRDF is calculated according to the formula [14]

$$\frac{dL_r(\theta_i,\phi_i,\theta_r,\phi_r,\lambda)}{dE(\theta_i,\phi_i)} \quad (1)$$

where L_r – the radiance of light, E – irradiance of light, λ – the wave length.

BRDFs are divided into physically accurate and empirical.

Physically accurate [6, 7] BRDFs are more accurate and usually take into account the division of the object's surface into facets. The physically accurate BRDFs include the Cook-Torrance [5, 6], Bagher [1], Oren-Nayyar [6, 12], coupled Shirley [3, 6] models [1, 16, 19].

In the Cook-Torrance model [6] it is taken into account that the reflectivity of the surface is affected by micro-facets [6], oriented along the vector $\vec{H} = (\vec{L}+\vec{V})/\left|\vec{L}+\vec{V}\right|$, where $\vec{L}$ and $\vec{V}$, respectively, are the vectors to the light source and the observer. The model includes a diffuse component, represented by Lambertian reflection, and a specular component. During the calculation of the specular component, the Fresnel factor F, the geometric shading factor G, and the distribution of micro-facets D are used. The specular component of this BRDF is calculated according to the formula [5, 6]

$$\frac{FDG}{\pi(\vec{N}\cdot\vec{L})(\vec{N}\cdot\vec{V})} \quad (2)$$

Bagher model [1] is an improvement of the Cook-Torrance model and lies in the usage of the SGD distribution (shifted gamma distribution) of micro-facets. Due to the usage of this distribution, compliance of the model with the measured characteristics of the most materials is ensured. SGD distribution is calculated according to the formula

$$\frac{\chi[0;\pi/2](\theta_m)}{\pi\cos(\theta_m)^4}\left(\frac{\alpha^{p-1}}{\Gamma(1-p,\alpha)}\right)\left(\frac{e^{\frac{-\alpha^2+\tan(\theta_m)^2}{\alpha}}}{(\alpha^2+\tan(\theta_m)^2)^p}\right) \quad (3)$$

where θ_m – the angle between surface and micro-facet normals, $\chi[0;\pi/2]=1$ if $\theta_m<\pi/2$, $\chi[0;\pi/2]=0$ if $\theta_m>=\pi/2$, Γ – incomplete gamma-function, p – the parameter of model, α – the roughness.

The Oren-Nayard model [6, 12] is an improvement of the Lambert model. It is calculated according to the formula

$$\frac{\rho}{\pi}\left((1-0.5\frac{\alpha^2{}_m}{\alpha^2{}_m+0.33})+\right.$$
$$+0.45\frac{\alpha^2{}_m}{\alpha^2{}_m+0.09}\max(0,\cos(\phi_{w_i}-\phi_{w_o}))\cdot \quad (4)$$
$$\left.\cdot\sin(\max(\theta_{w_o},\theta_{w_i}))\tan(\min(\theta_{w_o},\theta_{w_i}))\right)$$

where $\alpha^2{}_m$ – object's surface rougness, w_i, w_o – the light source and observer vectors respectively.

The coupled Shirley BRDF [3, 6] connects the diffuse and specular color components. The latter component is applied to polished surfaces of dielectrics. The model is energetically plausible. It is calculated according to the formula

$$[R_0+(1-\cos(\omega_o))^5(1-R_0)]f_{r,s}(w_o,w_i)+$$
$$+kR_m[1-(1-\cos(\omega_o))^5][1-(1-\cos(\omega_i))^5] \quad (5)$$

where k – normalization constant, R_m – the reflectance of matte component, R_0 – parameter [0.03, 0.06].

The disadvantage of this group of BRDFs is significant computational costs, therefore, physically accurate BRDFs are rarely used in highly productive graphic systems.

Among the empirical models, characterized by high productivity and approximation accuracy of light reflection, the Blinn and Phong BRDFs are the most often used.

For the calculation of Phong BRDF [4], a vector to the observer and a specular reflection vector $\vec{R}$ are used. Phong BRDF is calculated by the formula $\cos(\psi)^n$, where ψ – is the angle between $\vec{V}$ and $\vec{R}$, n is the shininess coefficient of the surface.

One of the disadvantages of the model is the possibility of a larger ψ value than 90º, which can lead to the appearance of image artifacts.

In the Blinn model [2] the angle γ between the vectors $\vec{N}$ and $\vec{H}$, which does not exceed 90º, is used instead of ψ. $\vec{N}$ is a normal to the surface.

When the *n* values are big the Blinn and Phong BRDFs do not meet the requirements of highly productive graphics systems. Therefore, for the Blinn-Phong model the approximating BRDFs are used.

For the Blinn-Fong model approximation by a cubic function [8] the calculated coefficients *A, B, C,* which depend on the shininess coefficient, are used.

The cubic model of light reflection is calculated according to the formula [8]

$$A\cos(\gamma)^3+B\cos(\gamma)^2+C\cos(\gamma) \quad (6)$$

The values of the Q,G points of the ordinate axis are used to calculate the coefficients of the model. The point Q is usually located near the level of the inflection point of the function. The point G is located near the zero of the ordinate axis and is used to control the size of the glare's attenuation zone.

The coefficient *A* is calculated according to the formula

$$\frac{LR(R-L)+GR(1-R)+LQ(L-1)}{L^3(R-R^2)+L^2(R^3-R)-LR^2(R-1)} \quad (7)$$

where $R=\cos(t)$, $L=\cos(u)$, t,u – abscissa axis points that correspond to ordinate axis points Q,G.

The coefficient *B* is calculated according to the formula

$$\frac{GR(R^2-1)+LR(L^2-R^2)+QL(1-L^2)}{L^3(R-R^2)+L^2(R^3-R)-LR^2(R-1)} \quad (8)$$

The coefficient *C* is calculated according to the formula

$$\frac{GR(R-R^2)+L^2Q(L-1)+LR(LR^2-L^2R)}{L^3(R-R^2)+L^2(R^3-R)-LR^2(R-1)} \quad (9)$$

For this function it was established that the optimal Q,G values are 0.5, 1/18 [8].

Fig. 2 shows the plot of *A, B, C* coefficients values when $Q=0.5$, $G=1/18$ and $n\in[4,256]$.

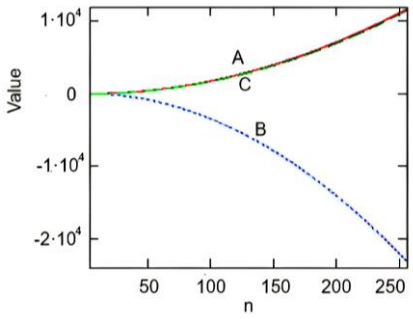


Fig. 2. The plot of A, B, C coefficients values

As shown in the figure, the coefficients *A* and *C* have similar values.

Let us denote the considered cubic function F_{KUB1}, which is an approximation of the Blinn-Fong BRDF (F_B)). Fig. 3 shows the graphs of F_{KUB1} and F_B when *n* = 50.

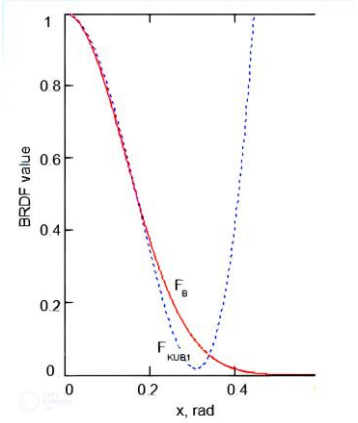


Fig. 3. The plots of F_B, F_{KUB1} when n = 50

The disadvantage of F_{KUB1} is an insufficiently accurate approximation of F_B in the attenuation zone, which leads to an unnatural reproduction of glare`s attenuation. Therefore, the F_{KUB1} improvement is necessary to ensure a more accurate F_B approximation.

2. Aim of the research

The aim of the paper is to develop the new lighting model through using the combination of two cubic BRDFs in order to provide the more accurate glare reproduction in the attenuation zone.

3. The development of combined BRDF based on polynomial cubic functions

To ensure a highly accurate F_B approximation, we combine F_{KUB1} with another cubic function, which will allow us to more accurately reproduce the attenuation zone of the glare.

For the connection of F_{KUB1} with the new cubic function F_{KUB2}, we will choose the level of the ordinate axis 0.5, below which the values of F_{KUB1} and F_B are noticeably different.

Let's determine the coefficient formulas for F_{KUB2} using a system of equations. The first equation corresponds to the level $\cos(E)^n = 0.5$, the second equation corresponds to the level $\cos(t)^n = Q$, the third to the level $\cos(u)^n = G$, the fourth to the level $\cos(\pi/2)^n = 0$. The system of equations is defined as

$$\begin{cases} A2\cdot cos(E)^3 + B2\cdot cos(E)^2 + C2\cdot cos(E) + D2 = 0.5 \\ A2\cdot cos(t)^3 + B2\cdot cos(t)^2 + C2\cdot cos(t) + D2 = Q \\ A2\cdot cos(u)^3 + B2\cdot cos(u)^2 + C2\cdot cos(u) + D2 = G \\ D2 = 0 \end{cases} \quad (10)$$

where $E = a\cos(e^{-0.693/n})$.

From the system we find that the coefficient $A2$ is calculated according to the formula

$$0.5\frac{LR(R-L)+GR(2E^2-R2E)+LQ(L2E-2E^2)}{L^3(RE^2-R^2E)+L^2(R^3E-RE^3)-LR^2(RE^2-E^3)} \quad (11)$$

where $E = e^{-0.693/n}$.

$B2$ is calculated according to the formula

$$0.5\frac{GR(R^2 2E-2E^3)+LR(L^2-R^2)+QL(2E^3-L^2 2E)}{L^3(RE^2-R^2E)+L^2(R^3E-RE^3)-LR^2(RE^2-E^3)} \quad (12)$$

$C2$ is calculated according to the formula

$$0.5\frac{GR(R2E^3-R^2 2E^2)+L^2Q(L2E^2-2E^3)+LR(LR^2-L^2R)}{L^3(RE^2-R^2E)+L^2(R^3E-RE^3)-LR^2(RE^2-E^3)} \quad (13)$$

Fig. 4 shows the plots of F_{KUB2} $A2$, $B2$, $C2$ coefficients values when $Q = 0.1$, $G = 1/40$ and $n \in [4,256]$.

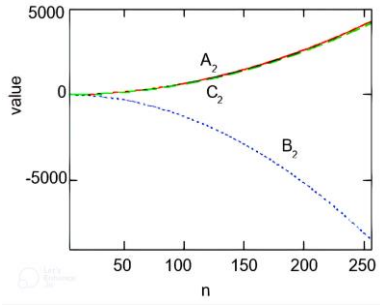


Fig. 4. The plots of $A2$, $B2$, $C2$ coefficients values

It is advisable to store the calculated F_{KUB1} and F_{KUB2} coefficients in a block of permanent memory, since they are the same for all cases of calculation of BRDF.

Fig. 5 shows the plots of F_{KUB2}, F_{KUB1} and F_B when *n* = 50. The selected values for Q, G F_{KUB2} are 0.1, 1/40.

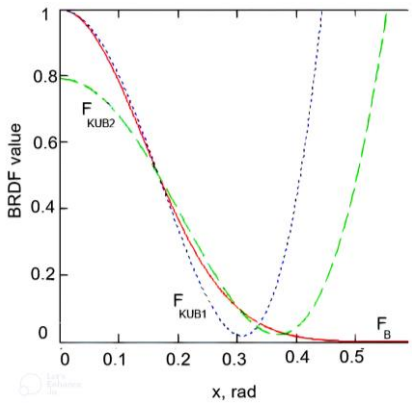


Fig. 5. The plots of F_B, F_{KUB2}, F_{KUB1} when n = 50

As can be seen from the figure, the combination of F_{KUB2}, F_{KUB1} at the point 0.5 provides a highly accurate approximation of F_B in the epicenter and attenuation zones.

Let's find the abscissa of the connection point of F_{KUB1} and F_{KUB2}. We equate the calculation formulas of F_{KUB1} and F_{KUB2}

$$\begin{aligned} &A\cdot\cos(x)^3 + B\cdot\cos(x)^2 + C\cdot\cos(x) = \\ &= A2\cdot\cos(x)^3 + B2\cdot\cos(x)^2 + C2\cdot\cos(x) \end{aligned} \quad (14)$$

From the equation, we find the formula for the point of connection of F_{KUB1} and F_{KUB2} $connect(n)$

$$connect(n) = a\cos(e^{-0.693/n}) \tag{15}$$

With the usage of MS Excel the simplified calculation expressions of $connect(n)$ ($connect\_appr(n)$) depending on n were obtained

$$\begin{cases} (\frac{1}{2^9}-\frac{1}{2^{12}})n^2-(\frac{1}{2^5}+\frac{1}{2^6}+\frac{1}{2^7})n+(\frac{1}{2^1}+\frac{1}{2^2}), n \ge 4 \wedge n \le 16 \\ (\frac{1}{2^{15}}+\frac{1}{2^{16}}+\frac{1}{2^{18}})n^2-(\frac{1}{2^7}-\frac{1}{2^{10}})n+(\frac{1}{2^2}+\frac{1}{2^3}), n > 16 \wedge n \le 70 \\ (\frac{1}{2^{19}}-\frac{1}{2^{22}})n^2-(\frac{1}{2^{10}}-\frac{1}{2^{13}})n+(\frac{1}{2^3}+\frac{1}{2^4}), n > 70 \wedge n \le 256. \end{cases} \tag{16}$$

Fig. 6 shows the plots of $connect\_appr(n)$ and $connect(n)$ depending on the values $n \in [4,256]$.

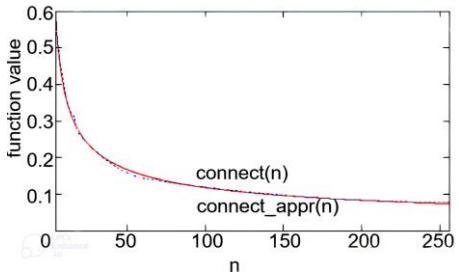


Fig. 6. The plots of original and approximated formulas of calculation of the cubic functions connection point

Therefore, a high-precision approximation of the original formula of the cubic functions connection point is ensured.

To ensure the smoothness of the connection of two cubic functions, a 0.5 level was chosen. Finding the derivatives of functions F_{KUB1} and F_{KUB2} is not necessary, since the values of functions are similar at a sufficiently large interval near $connect(n)$ (Fig. 7 shows the absolute deviation between F_{KUB1} and F_{KUB2} for $n \in [50,200]$).

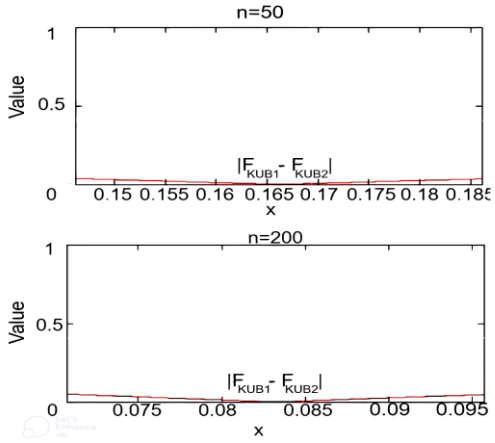


Fig. 7. The plots of absolute deviation between F_{KUB1} and F_{KUB2} near their connection point

We denote the developed combined function as F_{KUB3}.

Fig. 8 shows the graphs of F_{KUB1}, F_{KUB3} and F_B when n = 50.

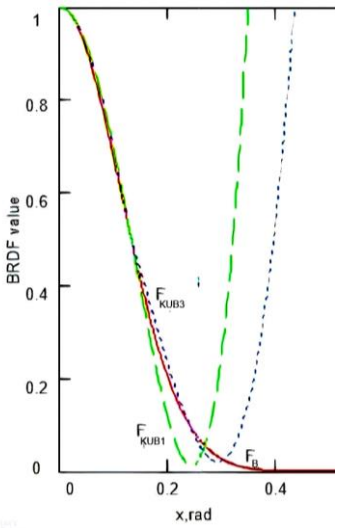


Fig. 8. The plots of F_{KUB1}, F_{KUB3} and F_B when n = 50

Through using the combination of two cubic functions in F_{KUB3}, an increase in the accuracy of the approximation of the glare's attenuation zone has been achieved.

Since the F_{KUB1} and F_{KUB3} coincide in the zone of the glare's epicenter, their maximum relative errors δ of F_B approximation in the epicenter are equal (2.95%, Fig. 9).

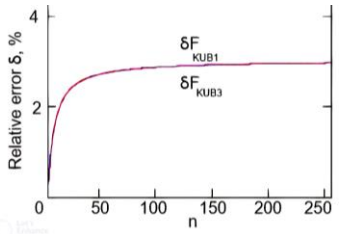


Fig. 9. The plots of maximum relative errors of F_{KUB1} and F_{KUB3} in glare`s epicenter

Fig. 10 shows the plots of the relative errors of F_{KUB1} and F_{KUB3} from F_B at the inflection point that is separating the epicenter and attenuation zones. The plot is built relative to the values of $n \in [4,256]$.

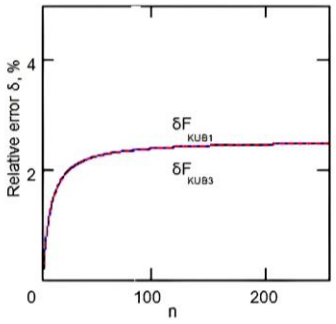


Fig. 10. The plots of relative errors of F_{KUB1} and F_{KUB3} at the F_B inflection point

The maximum relative error of F_{KUB1} and F_{KUB3} from F_B in its inflection point is 2.5%.

Fig. 11 shows the plots of maximum absolute errors Δ of F_{KUB1} and F_{KUB3} from F_B depending on $n \in [4,256]$.

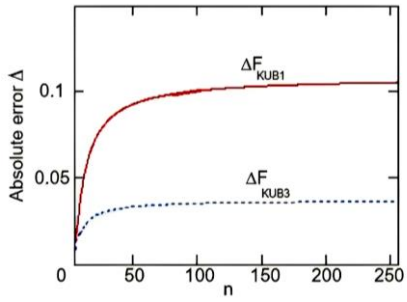


Fig. 11. The plots of absolute errors of F_{KUB1} and F_{KUB3} from F_B

It is worth noting that the maximum absolute error of F_{KUB1} from the known function is 0.11, it's unacceptable. Maximum absolute error of F_{KUB3} from F_B is 0.035. Therefore, F_{KUB3} in comparison with F_{KUB1} provides more accurate F_B approximation.

Fig. 12 shows the plot of absolute errors between F_{KUB3} and F_B depending on $n \in [4,256]$ and $x \in [0;\pi/2]$ values.

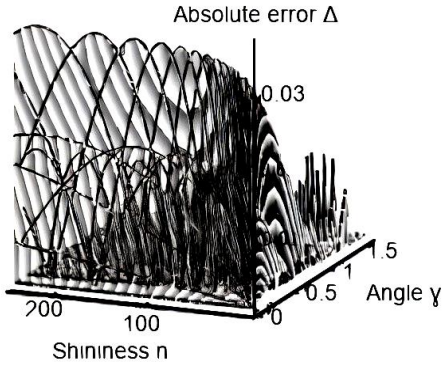
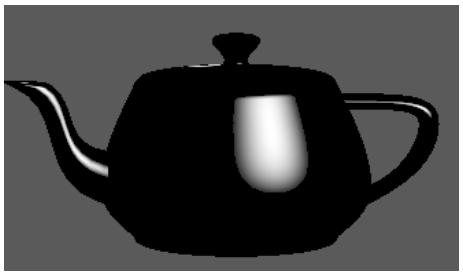


Fig. 12. The plot of absolute errors between F_{KUB3} and F_B depending on angle value and shininess

Based on F_{KUB1} and F_{KUB3} the test figure "Teapot" was visualized in BRDF Explorer. The visualization result is given in Fig. 13.

Fig. 13 shows that F_{KUB3} provides a more realistic formation of the glare's attenuation zone.



F_{KUB1}

F_{KUB3}

Fig. 13. The visualized teapots based on F_{KUB1} and F_{KUB3}

4. Conclusions

The article proposes a model of surface reflectance based on the combination of two cubic functions. The developed model is an improvement of a separate cubic model [8] and provides an increase in the accuracy of the approximation of the widely used Blinn-Fong model. The model combines the original cubic function and the new cubic function with calculated formulas of coefficients.

The developed combined model was compared to the original cubic model. It is shown that the developed model provides highly accurate reproduction of both the glare's epicenter and its attenuation.

The developed model is intended for use in highly realistic graphic systems.

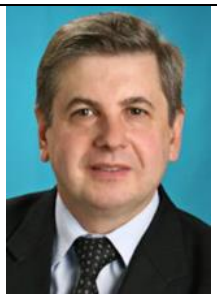
References

[1] Avrunin O. G., Tymkovych M. Y., Abdelhamid I. Y., Shushliapina N. O., Nosova Y. V., Semenets V. V.: Features of image segmentation of the upper respiratory tract for planning of rhinosurgical surgery. IEEE 39th International Conference on Electronics and Nanotechnology, ELNANO 2019, 485–488.
[2] Bagher M.: Accurate fitting of measured reflectances using a Shifted Gamma micro-facet distribution. Computer Graphics Forum 31(4), 2012, 1509–1518.
[3] Han Yu. et al.: Learning a 3D Morphable Face Reflectance Model from Low-cost Data. arXiv: 2303.11686, 2023.
[4] Jakob W. et al.: A Comprehansive Framework for Rendering Layered Materials. ACM Transactions on Graphics 33(4), 2014, 1–14.
[5] Kurt M.: Real-Time Shading with Phong BRDF Model. DEUFMD 21(63), 2019, 859–867.
[6] Liu H. et al.: Development of a Face Recognition System and Its Intelligent Lighting Compensation Method for Dark-Field Application. IEEE Transactions on Instrumentation and Measurement 70, 2021, 1–16.
[7] Montes R., Urena C.: An Overview of BRDF Models. University of Granada, 2012.
[8] Romanyuk A. The Bidirectional Reflectance Distributive Function Classification Scientific Papers of Donetsk National Technical University 9, 2008,145–151.
[9] Romanyuk O., Chornyy A.: Vysokoproduktyvni metody ta zasoby zafarbovuvannya tryvymirnykh hrafichnykh ob"yektiv. UNIVERSUM-Vinnytsya, Vinnytsya, 2006.
[10] Romanyuk O.: Komp'yuterna hrafika: Navchal'nyy posibnyk. VDTU, Vinnytsya 1999.
[11] Romanyuk O. et al.: The Concept and Means of Adaptive Shading. 12th International Conference on Advanced Computer Information Technologies (ACIT), Ruzomberok, 2022, 33 –38.
[12] Romanyuk S., Pavlov S., Wójcik W. et al.: Using lights in a volume-oriented rendering. Proc. SPIE 10445, 2017, 104450U.
[13] Schill S. et al.: Temporal Modeling of Bidirectional Reflection Distribution Function (BRDF) in Coastal Vegetation. GIScience & Remote Sensing 41(2), 2004, 116–134.
[14] Tan P.: Phong Reflectance Model. Computer Vision, 2020, 1–3.
[15] Thiele S. et al.: A Novel and Open-Source Illumination Correction for Hyperspectral Digital Outcrop Models. IEEE Transactions on Geoscience and Remote Sensing 60, 2022, 1–12.
[16] Wójcik W., Pavlov S., Kalimoldayev M.: Information Technology in Medical Diagnostics II. Taylor & Francis Group, CRC Press, Balkema book, London 2019.
[17] Wójcik W., Smolarz A.: Information Technology in Medical Diagnostics. CRC Press, 2017.
[18] Zavalniuk Ye. K. et al.: The development of the modified schlick model for the specular color component calculation. Information technology and computer engineering 55(3), 2022, 4–12.
[19] Zou Ya. et al.: Developmental Trends in the Application and Measurement of the Bidirectional Reflection Distribution Function. Sensors 22 (5), 2022, 1739–1763.

Prof. Oleksandr N. Romanyuk
e-mail: rom8591@gmail.com

Doctor of Technical Sciences, professor, Head of Department of Software Engineering of Vinnytsia National Technical University.
Research field: formation and processing of graphic images.

http://orcid.org/0000-0002-2245-3364



Ph.D. Zlata Bondarenko
e–mail : bondarenko@vntu.edu.ua

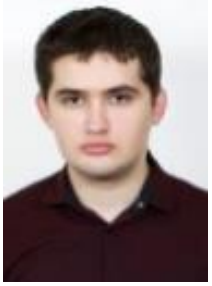
Associate professor of the Department of Higher Mathematics, Vinnytsia National Technical University.
Scientific direction – innovative technologies for training future specialists in technical specialties.

http://orcid.org/0000-0003-3339-0570

M.Sc. Yevhen K. Zavalniuk
e-mail: qq9272627@gmail.com

Ph.D. student, Department of Software Engineering of Vinnytsia National Technical University.
Research interests include formation and processing of graphic images.

http://orcid.org/0009-0005-1202-4653



Ph.D. Tetiana Koval
e-mail: tanyabyzja@ukr.net

Candidate of Pedagogic Sciences, associate professor at the Department of Artistic Disciplines, Preschool and Elementary Education, Vinnytsia Mykhailo Kotsiubynskyi State Pedagogical University, Vinnytsia Ukraine
Research field: information technologies in pedagogic

http://orcid.org/0000-0002-3190-1181

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 hemodynamic cardiovascular system of comprehensive into account scattering effects.
http://orcid.org/0000-0002-0051-5560

Ph.D. Aliya Kalizhanova
e-mail: kalizhanova_aliya@mail.ru

Candidate of physical and mathematical sciences, professor at University of Power Engineering and Telecommunications, the chief researcher of the Institute of Information and Computational Technologies of the Ministry of Education and Science CS of the Republic of Kazakhstan.
Scientific interests of the leader: mathematical modeling of systems, models of transport systems network analysis, optimization methods, technologies for developing sensor systems for signals receive - transmit, mathematical modeling of Bragg fiber gratings.

http://orcid.org/0000-0002-5979-9756

PhD Roman Y. Chekhmestruk
e-mail: Rc.ua@3dgeneration.com

Candidate of Technical Sciences. Technical Director 3D GENERATION UA, research field: formation and processing of graphic images.

http://orcid.org/0000-0002-5362-8796

M.Sc. Eng. Aigul Iskakova
e-mail: Iskakova1979@mail.ru

Lecturer of the Kazakh National Research Technical University named after K. I. Satpayev.
Total number of scientific and methodological developments and articles – more then 45.

http://orcid.org/0000-0001-8043-819X

http://doi.org/10.35784/iapgos.5331 *received: 08.08.2023 | revised: 09.09.2023 | accepted: 17.09.2023 | available online: 30.09.2023*

THE CONCEPT OF ELECTRONIC CONTROL UNIT FOR COMBUSTION ENGINE IN HYBRID TANDEM

Tomasz Zyska[1], Marcin Powązka[2], Bartłomiej Forysiuk[2]
[1]Lublin University of Technology, Department of Electronics and Information Technology, Lublin, Poland, [2]Radiotechnika Marketing Sp. z o.o., Warsaw, Poland

Abstract. *The article presents the concept of a hybrid propulsion unit for ultralight aircraft. A combustion unit and an electric motor/generator were presented. As part of the research, a dedicated combustion engine controller and a two-way converter for energy conversion were built. The prototype drive unit was installed on a test stand and its initial characteristics were determined. The results of research and conclusions related to the construction of a hybrid drive unit are presented.*

Keywords: combustion engine, axial motor-generator, engine controller, hybrid engine

KONCEPCJA ELEKTRONICZNEJ JEDNOSTKI STERUJĄCEJ SILNIKA SPALINOWEGO W ZESPOLE HYBRYDOWYM

Streszczenie. *W artukule przedstawiona została koncepcja hybrydowej jednostki napędowej dla ultralekkich statków powietrznych. Przedstawione zostały jednostka spalinowa a także silnik/generator elektryczny. W ramach badań zbudowany został dedykowany sterownik silnika spalinowego i dwukierunkowa przetwornica do konwersji energii. Prototypowy zespół napędowy zabudowano na stanowisku badawczym i wyznaczono jego wstępne charakterystyki. Przedstawiono wyniki badań i wnioski związane z budową hybrydowej jednostki napędowej.*

Słowa kluczowe: silnik spalinowy, zespół silnik-generator, kontroler silnika, zespół hybrydowy

Introduction

The latest proposals of the European Commission as part of the Fit for 55 package include: taxation of aviation fuel, which has so far been exempt from fees, the withdrawal of free CO_2 emission allowances and a greater share in driving with low-emission fuels [2, 3]. An alternative to traditional propulsion could be synthetic fuels or biofuels.

Electric or hybrid aircraft may also become the future of aviation. Hybrid drives in the aviation industry are a novelty that gives the possibility of environmentally friendly flights with less noise.

The hybrid connection of a gasoline combustion engine and an electrical motor offers numerous advantages, which make it an ideal solution for fuel-efficient and environmentally-friendly appllications. Some of the major advantages are shown below:

1. Fuel Efficiency: The hybrid connection provides improved fuel efficiency by utilizing the electric motor during low-speed and light-load conditions. This results in significant fuel savings and reduced emissions.
2. Performance: The hybrid connection provides improved performance by utilizing the electric motor to assist the gasoline engine during high-speed and heavy-load conditions. This results in improved acceleration and better handling.
3. Regenerative Braking: The hybrid connection utilizes regenerative braking, which converts the kinetic energy into electrical energy during braking. This energy is stored in the battery pack and can be utilized to power the electric motor.
4. Reduced Emissions: The hybrid connection reduces emissions by utilizing the electric motor during low-speed and light-load conditions. This results in reduced emissions of greenhouse gases and other pollutants.

The future of hybrid electric aircraft in aviation is promising, as the industry seeks to reduce emissions and fuel consumption, while maintaining safety and performance standards.

1. Hybrid connection of gasoline combustion engine and electrical motor

The hybrid connection of a gasoline combustion engine and an electrical motor can be achieved through different methods. The most common approach is through a series-parallel hybrid system, which utilizes both series and parallel configurations to optimize the performance of the powertrain. In this system, the gasoline engine is coupled with an electric motor, which is powered by a battery pack. The battery pack provides additional power to the electric motor, which in turn assists the gasoline engine during acceleration or high load conditions.

During low-speed and light-load conditions, the gasoline engine can be turned off, and the vehicle can operate solely on the electric motor. This is known as the electric-only mode, which provides a significant reduction in fuel consumption and emissions. On the other hand, during high-speed and heavy-load conditions, the gasoline engine is turned on, and the electric motor assists the engine to provide additional power. This is known as the hybrid mode, which optimizes the performance of the powertrain and provides improved fuel efficiency.

1.1. Serial hybrid system

A serial hybrid system, also known as a range-extended electrical device (REED), utilizes an electric motor as the primary source of power and a gasoline engine as a secondary source of power. In a serial hybrid system, the gasoline engine does not drive the machine directly, but instead, it is used to generate electricity, which powers the electric motor. The electric motor provides the majority of the mechanical power, while the gasoline engine is used only to recharge the battery or to provide additional power when needed.

The serial hybrid system offers several advantages, such as improved fuel efficiency, reduced emissions, and improved performance. The electric motor provides the majority of the power needed to drive machine, which results in improved fuel efficiency and reduced emissions. The gasoline engine is only used to recharge the battery or to provide additional power when needed, which further improves fuel efficiency and reduces emissions. The serial hybrid system also offers improved performance, as the electric motor provides instant torque, which results in improved acceleration and better handling.

One limitation of the serial hybrid system is the cost of the battery and electric motor. The battery and electric motor are more expensive than a gasoline engine, which results in a higher cost for the vehicle. The serial hybrid system also has limited electric-only range, as the battery is only used to power the electric motor and not to drive the wheels directly [3].

1.2. Parallel hybrid system

A parallel hybrid system utilizes both a gasoline engine and an electric motor to drive the wheels directly. In a parallel hybrid system, the gasoline engine and the electric motor are

connected to the same transmission and can work together or independently to provide power to the receiver. The gasoline engine provides power during high-speed and heavy-load conditions, while the electric motor provides power during low-speed and light-load conditions.

The parallel hybrid system offers several advantages, such as improved fuel efficiency, improved performance, and regenerative braking. The electric motor provides power during low-speed and light-load conditions, which results in improved fuel efficiency and reduced emissions. The gasoline engine provides power during high-speed and heavy-load conditions, which further improves fuel efficiency and performance. The parallel hybrid system also utilizes regenerative braking, which converts the kinetic energy of the vehicle into electrical energy during braking. This energy is stored in the battery and can be used to power the electric motor.

One limitation of the parallel hybrid system is the complexity of the powertrain. The parallel hybrid system has more components than a traditional gasoline-powered vehicle, which results in increased maintenance and repair costs. The parallel hybrid system also has limited electric-only range, as the electric motor is not the primary source of power and is only used during low-speed and light-load conditions [3].

2. Proposition of hybrid system for ultralight aircrafts

Block diagram of presented solution is show on Fig. 1.

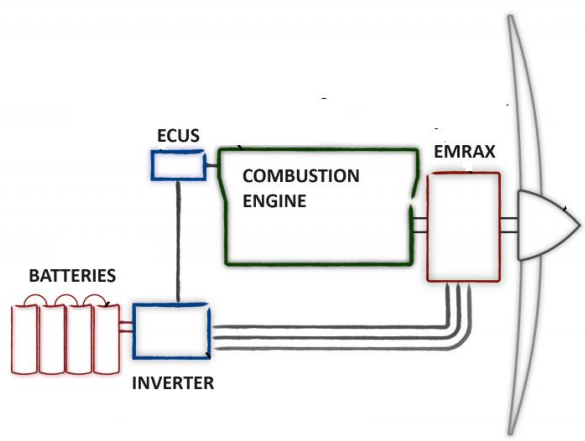


Fig. 1. Block diagram of proposed hybrid system

The diagram (Fig. 1) shows the basic components of a serial hybrid aircraft engine.

The combustion engine is connected together with the electric motor through a shaft and further drives the propeller. The electric motor is also a generator that allows both to supply mechanical power to the main shaft of the unit and to recharge the batteries during the flight [1].

The main converter is a bi-directional converter that allows energy to flow to and from the batteries. The converter also determines the operating state of the internal combustion engine.

This solution requires a specific design of the electronic controller of the internal combustion engine.

Due to the generally accepted standards for aviation, the internal combustion engine controller was built as two independent units. The first unit is responsible for generation the fuel injection sequence and the second unit is responsible for generation the ignition sequence [4].

All three controllers – the main converter, the injection controller and the ignition controller are connected by a common CAN bus.

On figure 2 laboratory stand is presented.

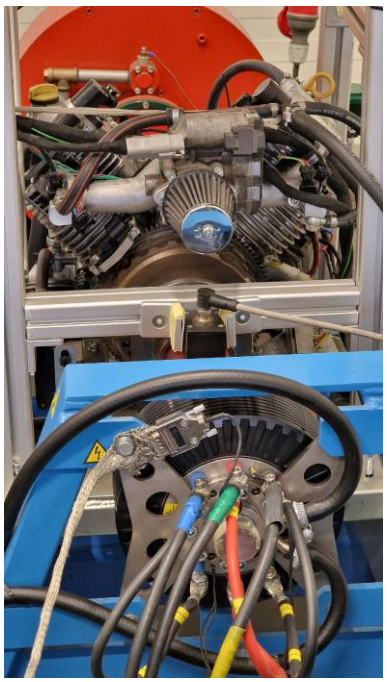


Fig. 2. Laboratory stand for hybrid system testing

2.1. Mechanical part of system

The object of the research is a two-cylinder internal combustion engine in the system V. It is a spark-ignition, air-cooled, non-supercharged four-stroke engine with a maximum power of 35 HP at a rotational speed of 4640 rpm. Engine view and some important technical informations are presented below.

Fig. 3. Vaxell VX 40I engine (Vaxell datasheet)

Table 1. VX 40I basic technical information (Vaxell datasheet)

Construction	2 cylinder V2/90°
Continuous power	35 hp at 4 640 rpm
Dry weight	34.6 kg
Cooling	air
Reduction gear unit	1:1.77
TBO	1500 h
Propeller RPM	2620 1/min

This engine is coupled with the EMRAX motor / generator by direct connection through the drive shaft. The choice of this element was caused by its low weight, high power and ease of installation.

EMRAX view and some important technical informations are presented below.

Fig. 4. EMRAX Axial flux motor / generator (EMRAX datasheet)

Table 2. EMRAX basic mechanical information (EMRAX datasheet)

Type:	Axial flux motor / generator
Casing diameter:	208 mm
Axial length:	85 mm
Dry mass:	9.4 kg (AC) / 10.0 kg (CC) / 10.3 kg (LC)
Stator cooling:	air (IP21) / combined (IP21) / liquid (IP65)
Mounting:	Front: 6x M8 threaded holes Back: 16x M8 threaded holes
Stacking:	Two motors can be stacked together to achieve doubled power / torque

Table 3. EMRAX basic electrical information

Maximal battery voltage:	580 (HV) / 390 (MV) / 140 Vdc (LV)
Peak power(at 6000 RPM):	86 kW
Continuous power*:	up to 56 kW
Peak torque:	150 Nm
Continuous torque*:	up to 90 Nm
Efficiency:	92-98%

*Subject to drive cycle, thermal conditions and controller capability.

2.2. Injection electronic control unit

The electronic unit controlling the operation of the internal combustion engine is designed to generate the appropriate impulses to stimulate the injectors on the basis of strictly specific input signals. The inputs for such a signal are:

1. RPM – hall effect engine speed sensor, informing at the same time about the location of the first cylinder TDC,
2. TPS – potentiometric throttle position sensor, indicating the "intent" of the pilot,
3. MAP – air pressure sensor in the intake pipes,
4. MAT – air temperature sensor in the intake pipes,
5. FPT – fuel temperature and pressure sensor, in the fuel manifold.

Based on signals from the above sensors the dose of fuel applicable in the next one is determined.

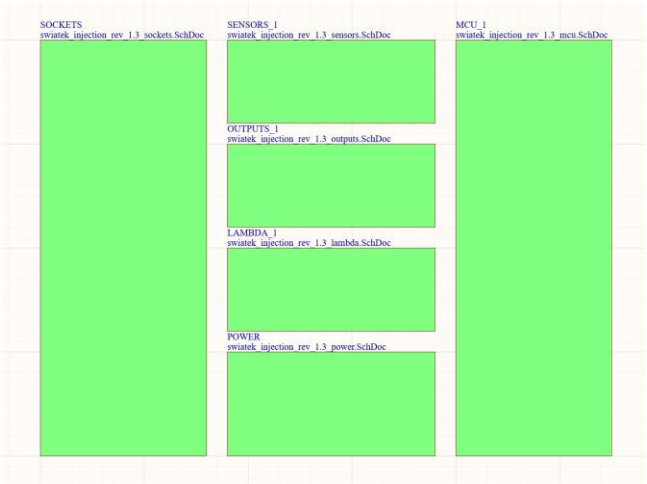


Fig. 5. Block diagram of Injection ECU

The experimental injection control unit is based on the Texas Instruments RM48L540DPGETR DSP processor. The injection sequence is built on the basis of input signals from the sensors. The injectors are controlled by current keys made in MOSFET technology. The injection control algorithm takes into account both the static states of engine operation - needed to determine the characteristic ones operating points and testing the possibility of replacing the generated mechanical power with electric power generated by the electric part of the hybrid unit.

Block diagram of Injection Controller is presented on Fig. 5.

The injection controller PCB is shown on Fig. 6.

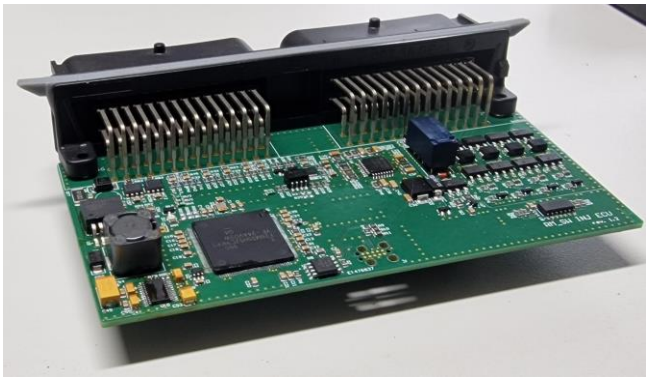


Fig. 6. Injection ECU – PCB view

Exemplary measurements of the generated sequence are presented below (Fig. 7). The yellow line shows the pulses coming from the RPM sensor. The orange line and the red line represent injector control for cylinder one and cylinder two, respectively.

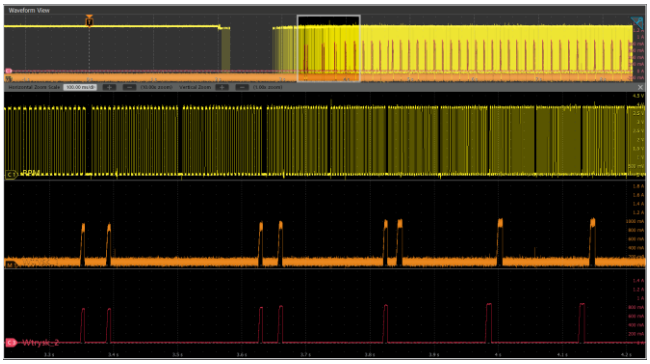


Fig. 7. Example of injection pulses sequence

The first two pulses after the rotational speed occur are caused by fuel injection and engine start-up facilitation. Later in the waveform after the start-up, an out-of-phase sequence is observed for both cylinders.

2.3. Ignition electronic control unit

The MAP and TPS sensors were used as input signals in the ignition system. In addition, a correction for the supply voltage was introduced because the supply voltage intensively affects the amount of energy stored in the ignition coil. As in the previous injection system, the same processor was used with the same necessary environment for its operation.

Block diagram of Ignition Controller is presented on Fig. 8. Main difference is between both controllers is placed in actuators section. Injectors and spark coils needs specific electronic circuits for current control.

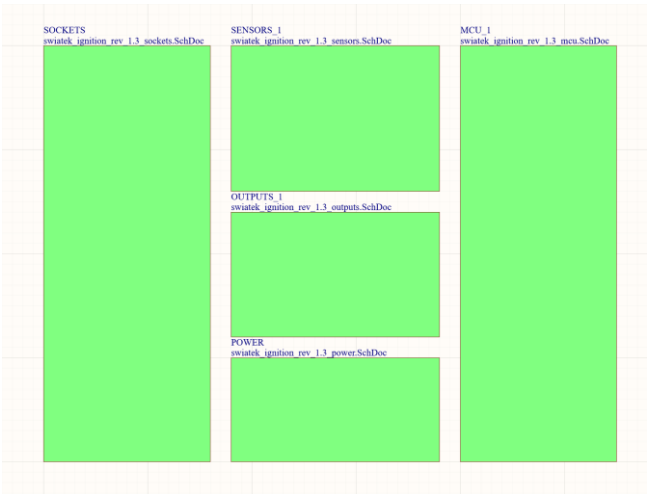


Fig. 8. Block diagram of Ignition ECU

The ignition controller PCB is shown on Fig. 9.

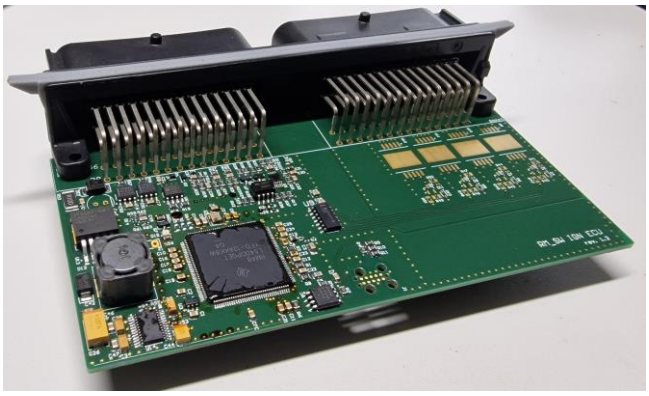


Fig. 9. Injection ECU – PCB view

The ignition coil control sequences are shown on Fig. 10.

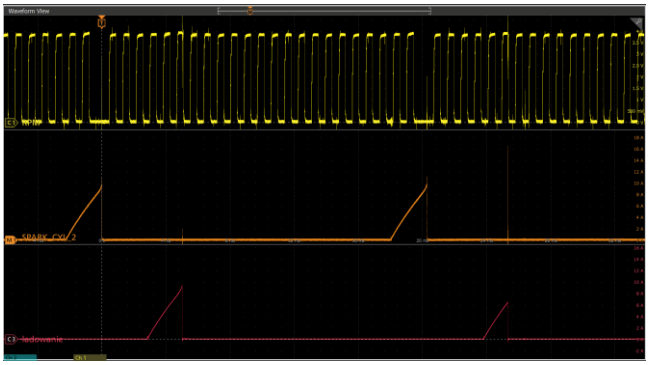


Fig. 10. Example of ignition pulses sequence

The waveform from the rpm sensor is marked with a yellow line. The orange and red lines show the ramming of the ignition coils.

The angular displacement between both cylinders of the controlled internal combustion engine is clearly visible. The duration of the control pulse depends on the supply voltage and the falling edge causes a physical spark in the engine cylinder.

3. Conclusions

The publication presents a hybrid concept and a unit that in the future may be used as a drive for an ultralight aircraft.

The current work focuses on the control of the internal combustion engine in order to achieve the proper efficiency of the process of converting chemical energy into mechanical energy.

Both injection and ignition control systems based on the input signals of the measured engine have been presented.

The optimal control process of the internal combustion engine was achieved through many hours of testing on the engine dynamometer.

The engine was loaded both with the electrodynamical brake and the EMRAX motor / generator, which in the next step in the next works will constitute the hybrid part of the proposed drive.

The most important conclusion from the work carried out is that it is possible to construct a hybrid system, it is possible to build algorithms for controlling the internal combustion engine and prepare a stand for pre-implementation tests.

Acknowledgements

Dane przedstawione w artykule powstały podczas realizacji Projektu współfinansowanego przez Unię Europejską ze środków Europejskiego Funduszu Rozwoju Regionalnego w ramach Programu Inteligentny Rozwój, Działanie 1.1 Projekty B+R przedsiębiorstw, Poddziałania 1.1.1 Badania przemysłowe i prace rozwojowe realizowane przez przedsiębiorstwa.
Projekt realizowany w ramach konkursu Narodowego Centrum Badań i Rozwoju: Szybka Ścieżka.

References

[1] Gabler J., Köhler D., Heinzler T., Lauff U.: ECU function development. ATZ Elektron Worldw 5, 2010, 36–39 [http://doi.org/10.1007/BF03242287].
[2] Rendón M. A., Sánchez R. C. D., Gallo M. J. et al.: Aircraft Hybrid-Electric Propulsion: Development Trends, Challenges and Opportunities. J Control Autom Electr Syst 32, 2021, 1244–1268 [http://doi.org/10.1007/s40313-021-00740-x].
[3] Voskuijl M., van Bogaert J., Rao A. G.: Analysis and design of hybrid electric regional turboprop aircraft. CEAS Aeronaut J 9, 2018, 15–25 [http://doi.org/10.1007/s13272-017-0272-1].
[4] Zyska T., Czarnigowski J., Jakliński P.: The concept of an electronic control for injection system into inlet manifolds of aircraft piston engine. Przegląd Elektrotechniczny 86(7), 2010, 235–237.

Ph.D. Eng.Tomasz Zyska
e-mail: t.zyska@pollub.pl

Graduate of the Faculty of Electrical Engineering, Lublin University of Technology in 2002. Since 2003, a research and teaching employee of the Department of Electronics and Information Technology at Lublin University of Technology. He defended his doctoral dissertation in 2011. His areas of interest include electronic systems for processes control and connected metrological asspects.

http://orcid.org/0000-0002-3911-0237

M.Sc. Marcin Powązka
e-mail: mpowazka@radiotechnika.com.pl

R&D Department Manager in Radiotechnika Marketing Sp. z o.o.

http://orcid.org/0009-0002-9180-2481

M.Sc. Bartłomiej Forysiuk
e-mail: bforysiuk@radiotechnika.com.pl

Software Development Department Manager in Radiotechnika Marketing Sp. z o.o.

http://orcid.org/0009-0002-0583-463X

http://doi.org/10.35784/iapgos.4051 received: 09.07.2023 | revised: 06.09.2023 | accepted: 27.09.2023 | available online: 30.09.2023

TESLA SWITCH OF 4 BATTERIES BASED ON THE ARDUINO UNO BOARD

Mykola Polishchuk[1], Serhii Grinyuk[1], Serhii Kostiuchko[1], Anatolii Tkachuk[2], Pavlo Savaryn[3]
[1]Lutsk National Technical University, Faculty of Computer and Information Technologies, Department of Computer Engineering and Cyber Security, Lutsk, Ukraine, [2]Lutsk National Technical University, Faculty of Computer and Information Technologies, Department of Electronics and Telecommunications, Lutsk, Ukraine, [3]Lutsk National Technical University, Faculty of Digital, Educational, and Social Technologies, Department of Digital Educational Technologies, Lutsk, Ukraine

Abstract. *The paper considered the theoretical information of generators for several power sources, schematic their solutions, and the main disadvantages and advantages. A generator for 4 batteries based on field-effect transistors with optical galvanic isolation and a clock generator built on the basis of the Atmega328 microcontroller, which is part of the Arduino Uno device, was studied. The result of the study was the confirmation of the existence of the "Tesla-switch" effect. A program code was developed for the alternate switching of six transistor switches of the optical decoupling. The structural diagram of the device was studied, and the influence of the generator frequency and the consumed power of the load on the output parameters of the device were determined. The considered idea of energy saving and environmental friendliness of power supply systems is relevant, in particular in cases of blackouts.*

Keywords: generator, batteries, transistor, microcontroller

PRZEŁĄCZNIK TESLI DLA 4 AKUMULATORÓW OPARTY NA MODULE ARDUINO UNO

Streszczenie. *W artykule rozważono informacje teoretyczne dotyczące generatorów dla kilku źródeł zasilania, schematy ich rozwiązań oraz główne wady i zalety. Zbadano generator dla 4 baterii oparty na tranzystorach polowych z optyczną izolacją galwaniczną oraz generator zegara zbudowany na bazie mikrokontrolera Atmega328, który jest częścią urządzenia Arduino Uno. Wynikiem badań było potwierdzenie istnienia efektu "Tesla-switch". Opracowano kod programu do naprzemiennego przełączania sześciu przełączników tranzystorowych odsprzęgacza optycznego. Zbadano schemat strukturalny urządzenia oraz określono wpływ częstotliwości generatora i mocy pobieranej przez obciążenie na parametry wyjściowe urządzenia. Rozważana idea oszczędzania energii i przyjazności dla środowiska systemów zasilania jest istotna, w szczególności w przypadku awarii zasilania.*

Słowa kluczowe: generator; baterie; tranzystor; mikrokontroler

Introduction

One of the biggest challenges in the renewable energy field is the electricity production that depends on natural resources that humans cannot control. For example, solar power is only generated when sunlight is available and shuts off at night; Wind energy also depends on the wind presence, so if the wind speed is very low, the turbine does not turn, resulting in zero electricity flow to the grid. On the other hand, too much wind can damage the generator, so a delicate balance must be maintained to maintain constant energy production. Uncertainties in energy production in renewable energy technologies make integration difficult [10].

Constantly high-quality electricity is necessary to ensure stability and network high efficiency. The power supply quality allows the system to work well with high reliability and lower costs. On the other hand, poor power quality can have serious negative consequences for the power grid and industrial processes. This can lead to high costs and equipment failure. Power quality issues include frequency disturbance, voltage/current harmonics, low power factor, voltage fluctuations, and transmission line transients [14].

Most renewable energy installations that share their energy with the grid require large areas. In most cases, renewable energy sources depend on location, which can turn off users. First, some renewable energy sources are simply not available in different regions. Second, the distance between the renewable energy source and the grid is a major consideration in cost and efficiency terms. In addition, renewable energy sources depend on weather, climate, and geographical location, so one type of energy production is not suitable for a region. Although the field is improving, there is a lack of information and awareness about renewable energy benefits and needs. Investments and capital incentives were available for the renewable energy sources implementation. There is a clear need for government agencies to assist and advise applicants and potential recipients on how to apply for renewable energy incentives. The high initial installation cost is one of the main obstacles to the development of renewable energy. While the coal-fired power plant development requires about $6 per megawatt, wind and solar power plants are known to have required large investments as well. In addition to this, the storage systems of the produced energy are expensive and present a real challenge in megawatt production terms [7].

1. Literature review

Elektrodine Corporation tested the Tesla circuit with 4 batteries for 3 years [9]. They found that at the end of testing the battery showed no signs of any wear and tear, while using conventional lead-acid batteries. The systems were powered by lighting, heaters, televisions, small engines, and a 30 hp electric motor. If the batteries were discharged to the lowest value and then connected with a load, the batteries recharged in less than 1 min. Left without intervention, each of the batteries received a up to 36 V charge. When using mechanical switches, it was concluded that with a switching frequency of less than 100 Hz, the scheme is ineffective, and more than 800 Hz can be dangerous. The charging process is illustrated in Fig. 1 and Fig. 2 [15].

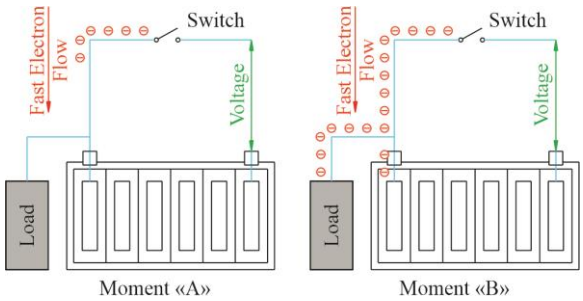


Fig. 1. Connecting the charger to the battery and the load. Moment «A» – the beginning of the electrons movement. Moment «B» – the stop of electrons at the entrance to the battery and the electrons movement into the load [9, 17]

At moment «A», the switch is closed, connecting the current source (battery, charged capacitor, other) with the lead-acid battery. Electrons begin to move outside the connecting conductor. Being very light, and not encountering significant resistance, they move very quickly (inside the conductor, electrons move very slowly, because movement through the conductor is difficult). All goes well until the «B» moment, when the electrons reach the lead plates inside the battery. Here they face a problem, because the current movement through the plates is carried out by lead ions. The latter perfectly cope with their task, but due to their great weight, they have a fraction of a second to start moving. This fraction of a second is very important. In that split second, the electrons pile up as they continue to come down the connecting conductor at tremendous speed. Therefore, at the moment «C» a significant amount of them accumulates [17].

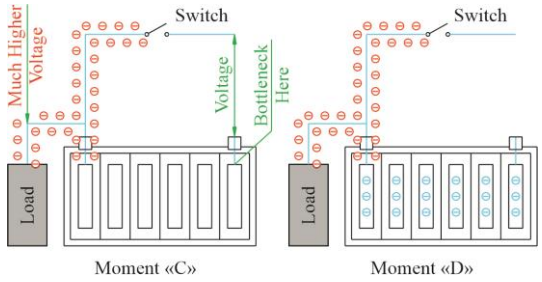


Fig. 2. Connecting the charger to the battery and the load. Moment «C» – increase in voltage at the inputs to the battery and load. Moment «D» – the beginning of the electrolyte ions movement in the battery – the charging beginning [9, 17]

The electron's large number accumulation is similar to the sudden connection of much higher voltage sources, giving capable a much larger current. This situation is very short-lived but has several very important consequences. First, at the moment «D» the battery receives a current that is much higher than the expected value from the source [13].

Second, excess energy enters the battery, charging it more than expected, and at the same time, some of the excess energy moves into the load, doing useful work. By load, we mean a lamp, a motor, an inverter, a pump, a drill, or whatever [1].

Therefore, excess energy is collected and used to perform useful work and simultaneously charge the batteries. Instead of being discharged by powering the load, the battery is charged by powering the load! That is why this scheme allows you to rotate the motor even from a discharged battery. This is made possible by the fact that the battery plates are lead composed, which creates an «electron plug» (a electrons plug), forcing the battery to charge and power the load at the same time. In principle, the more the battery is discharged, the faster it charges. John Bedini, who is an undisputed expert in this field, set up experiments in which the motors rotated continuously for three years or more, while the batteries did not discharge at all, despite the motors performing useful work [4].

2. Researches methodology

In order to ensure the necessary electrons accumulation, the switch closing must be very clear and instantaneous. A thyristor or field-effect transistor is suitable for these purposes, since, when turned on, it switches sharply and completely. There is an assumption that the Tesla candle with 4 batteries is based on this principle, and worked in the frequency range of 100…800 Hz [8].

This scheme can be improved by sharply disconnecting the electron flow from the output voltage source before the electron accumulation process is completed. This, in turn, causes a sudden (very short) increase in additional power, which increases the voltage and, as a result, allows you to increase the return of useful power to the load and accelerate the battery charge [2].

An even greater effect can be achieved if the next pulse reaches the battery and loads before the effect of the previous pulse dissipates. Suppose that this is what was found to be "unsafe" during experiments at Electrodyne Corporation at frequencies above 800 Hz. Maybe it's not that the battery or the load is "not ready" to accept excess energy, but rather that the components used are not designed for high currents and voltages. When the frequency is further increased, some circuit components are guaranteed to fail because they are not designed to work with high currents and voltages (note that the used output capacitors were designed for 100 V, which is eight times the nominal batteries voltage). This can hardly be called a problem, considering that they have a 12-volt battery, if necessary, they could perfectly withstand a 36 V charge voltage of [6, 11].

The Tesla candle from 4 batteries demonstrated the seemingly impossible thanks to:

- intercepting the current from the load and directing it to charge the second battery, instead of letting it dissipate;
- using the inertia of heavy lead ions with the help of short, sharp pulses (switching) [12].

John Bedini showed that the electric effect of Tesla with 4 batteries can be realized with three batteries (Fig. 3) [3, 9].

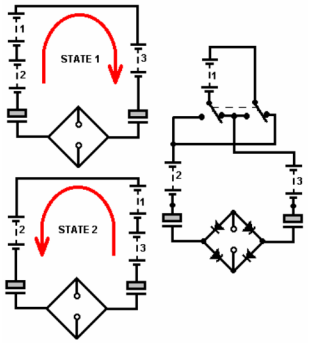


Fig. 3. Bedini generator for 3 batteries: operating cycles and connection method with mechanical switches [9]

In this circuit, the current in battery "1" never changes. This may not matter, since the energy supplied is taken from the zero-oscillation energy domain, not batteries. The flowchart used by John when he designed it is in Fig. 4.

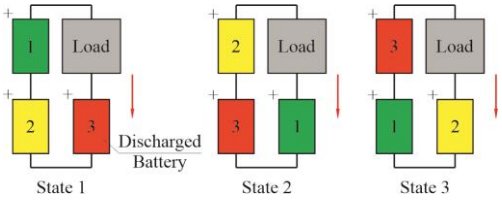


Fig. 4. Block diagram for a three-stroke Bedini generator [9]

This is a more difficult version of switches to implement. One battery "3" in the diagram is never charged. If mechanical switches were used (but they are not suitable because they are too slow), then the block diagram would look like this (Fig. 5).

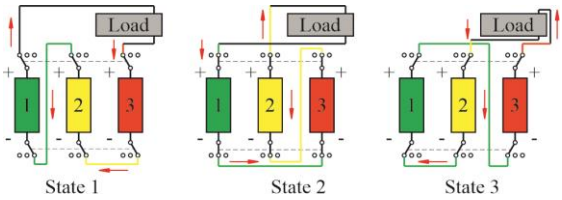


Fig. 5. The three-stroke generator with mechanical switches connection diagram [4]

The problem is to implement this scheme in an electronic version. Bedini said his device produces excess energy by creating resonance in the batteries, which are treated as capacitors. He warns that sometimes batteries explode – probably due to exceeding the limit frequency. He also noted that if the battery is used as a capacitor in a resonant circuit, it releases energy many times more than in normal mode (Fig. 6).

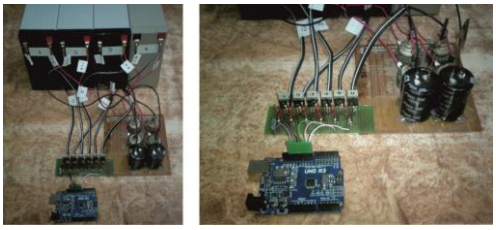


Fig. 6. Scheme of the installation for experiments (constructed by the authors)

Bedini was invited to participate in a symposium on the Tesla's centenary occasion in Colorado Springs on August 11, 1984. The symposium was dedicated to the Nikola Tesla's centenary arrival in the USA. It was organized by the Tesla Committee, Institute of Electronics and Electrical Engineers, Pikes Peak Section, and Ford Aerospace & Communications Corporation, Colorado Springs Operation. At the symposium, Bedini unexpectedly demonstrated a Tesla-type converter, the cigarettes pack size, that he had recently been building. During all demonstrations, which lasted 24 years, in the symposium framework, constant loads were connected to the working system. Despite this, the converter kept the nickel-cadmium battery fully charged [5, 16].

3. Results

The idea of a generator for 4 batteries is not new and arose long before the semiconductor radio electronic components appearance. Mechanical relays and contactors were used in the first generators for rapid battery switching. However, they could not provide stable contact, switching frequency and no distortion. It was also difficult to avoid contact losses.

With the element base development, it became easier to perform the creating task new schematic solutions. Semiconductor transistors have become an excellent substitute for mechanical switches when building circuits with high-speed switching. However, at high frequencies, the stabilizing problem the switching frequency remained. Control generators circuits with a quartz resonator successfully cope with this type complexity.

As a rule, circuits of arithmetic and logical devices are built with frequency stabilization on the basis of quartz. A microcontroller, like nothing else, will be suitable for the Tesla generator control device. When building the device, the microcontroller use will allow:

- configure system parameters;
- get a stable frequency at the output;
- connect additional sensors to monitor system operation;
- enter control elements into the device, which will be controlled by the microcontroller.

The relatively high voltages that will be operating in the system, there is a damaging high risk the control circuit and the microcontroller. On the other hand, the microcontroller itself is not designed to control power systems. In this situation, an optoelectric conversion device – an optocoupler or an optocoupler – will come in handy. It will allow stable and high-speed microcontroller controlled generator power circuits. The generator circuit and the control unit galvanic isolation will also be carried out, which will protect the microcontroller from damage by high voltage.

Considering all of the above, the Bedini-Tesla generator scheme was chosen with the following modifications:

- bipolar transistors were replaced by field ones;
- optocouplers were used for galvanic decoupling;
- a programmable microcontroller was chosen as the control device.

The generator power part is shown in Fig. 7

Figure 7 schematically shows the optocouplers parts, after they are all connected to the microcontroller according to the same scheme and it does not matter how they are physically connected, after the controlled output of the microcontroller (Fig. 8) they are freely programmable.

To carry out research, we need to alternately switch 6 transistor switches using galvanic (optocoupler) decoupling. Since the connection in the scheme and optocoupler control is no different from the connection and light diode control, the program code for the microcontroller will be very similar to the code for controlling 6 LEDs. The trios of LEDs (optocouplers) will alternately turn on and off at equal time intervals.

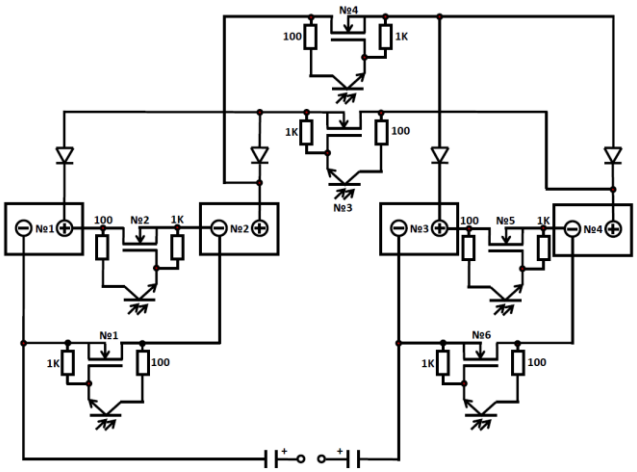


Fig. 7. The generator power part scheme (constructed by the authors)

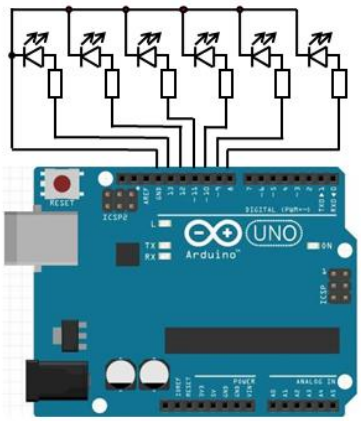


Fig. 8. The generator control part scheme (constructed by the authors)

So, the optocoupler control program will look like this (after the double slash, comments for each line are given):

```
                    // the setup block is executed once at startup
                     // reboots
voidsetup() {
pinMode(8, OUTPUT);     // digital pin 8 is triggered as an output
pinMode(9, OUTPUT);     // digital pin 9 is triggered as an output
pinMode(10, OUTPUT);   // digital pin 10 is triggered as an output
pinMode(11, OUTPUT);   // digital pin 11 is triggered as an output
pinMode(12, OUTPUT);   // digital pin 12 is triggered as an output
pinMode(13, OUTPUT);   // digital pin 13 is triggered as an output
}
                    // the loop block executes in a circle over and
over
voidloop() {
digitalWrite(8, LOW);    // on output 8 sets the potential LOW
digitalWrite(9, LOW);    // on output 9 sets the potential LOW
digitalWrite(10, LOW);   // on output 10 sets the potential LOW
digitalWrite(11, HIGH);  // on output 11 sets the potential HIGH
digitalWrite(12, HIGH);  // on output 12 sets the potential HIGH
digitalWrite(13, HIGH);  // on output 13 sets the potential HIGH
delay(5);              // delay in milliseconds
digitalWrite(13, LOW);   // on output 13 sets the potential LOW
digitalWrite(12, LOW);   // on output 12 sets the potential LOW
digitalWrite(11, LOW);   // on output 11 sets the potential LOW
digitalWrite(10, HIGH);  // on output 10 sets the potential HIGH
digitalWrite(9, HIGH);   // on output 9 sets the potential HIGH
digitalWrite(8, HIGH);   // on output 8 sets the potential HIGH
delay(5);              // delay in milliseconds
}
```

During the research, the program will not change except for the delay line (5). The number in parentheses is the delay in milliseconds between generator switches. By changing this number in both lines, you can change the generator frequency.

Battery parameters were measured at a 20 W load and a generator 100 Hz frequency. Measurements were made every 15 minutes, during 3 hours. Zero point measurements were performed with the generator turned off. After 3 hours of measurements, each battery was charged for one hour.

After measuring the voltage on the batteries in the generator mode at different frequencies, the generator was turned off, and the batteries were disconnected from the electronic keys. A load was connected to each battery in turn and measurements were made similar to the measurements of the generator (table 6).

According to the received data (table 1–5), graphs of the voltage change depending on the generator frequency and over a certain time period were drawn, Fig. 9 – Fig. 13.

Table 1. Results of voltage measurements on batteries at frequency 100 Hz

Battery number	Timestamps in minute												
	0	15	30	45	60	75	90	105	120	135	150	165	180
1	12.72	12.62	12.61	12.61	12.61	12.61	12.61	12.61	12.61	12.61	12.61	12.61	12.61
2	12.72	12.41	12.41	12.40	12.39	12.38	12.36	12.35	12.33	12.31	12.30	12.28	12.26
3	12.70	12.62	12.62	12.62	12.61	12.61	12.61	12.61	12.61	12.61	12.61	12.61	12.61
4	12.96	12.55	12.55	12.55	12.55	12.54	12.52	12.51	12.49	12.48	12.47	12.46	12.44

Table 2. Results of voltage measurements on batteries at frequency 200 Hz

Battery number	Timestamps in minute												
	0	15	30	45	60	75	90	105	120	135	150	165	180
1	12.71	12.50	12.49	12.48	12.47	12.46	12.45	12.44	12.43	12.42	12.41	12.36	12.35
2	12.52	12.19	12.16	12.12	12.08	12.05	12.00	11.95	11.90	11.82	11.69	9.39	9.51
3	12.69	12.51	12.50	12.49	12.48	12.48	12.48	12.47	12.46	12.45	12.45	12.43	12.42
4	12.73	12.39	12.36	12.34	12.31	12.29	12.26	12.24	12.22	12.20	12.17	12.17	12.17

Table 3. Results of voltage measurements on batteries at frequency 300 Hz

Battery number	Timestamps in minute												
	0	15	30	45	60	75	90	105	120	135	150	165	180
1	12.60	12.25	12.15	12.12	12.05	12.02	11.96	11.89	11.80	11.72	11.53	11.18	9.77
2	12.21	9.35	8.79	8.76	8.25	8.19	8.14	8.11	8.07	8.00	7.89	7.78	7.17
3	12.62	12.35	12.29	12.28	12.26	12.25	12.24	12.22	12.20	12.18	12.16	12.14	12.12
4	12.45	12.16	12.12	12.11	12.08	12.07	12.05	12.02	12.01	11.98	11.96	11.95	11.94

Table 4. Results of voltage measurements on batteries at frequency 400 Hz

Battery number	Timestamps in minute												
	0	15	30	45	60	75	90	105	120	135	150	165	180
1	12.43	12.16	12.10	12.05	11.98	11.90	11.81	11.64	9.35	8.52	7.48	7.38	7.33
2	12.43	12.06	11.96	11.83	9.32	8.22	7.93	7.35	7.18	5.91	5.31	4.00	3.93
3	12.31	12.01	11.98	11.94	11.88	11.67	11.45	9.76	9.55	9.24	8.44	7.31	7.26
4	12.16	11.82	11.77	11.72	11.73	11.70	11.68	11.67	11.61	11.62	11.60	11.60	11.60

Table 5. Results of voltage measurements on batteries at frequency 500 Hz

Battery number	Timestamps in minute												
	0	15	30	45	60	75	90	105	120	135	150	165	180
1	12.38	12.12	12.07	11.99	11.93	11.90	11.60	10.41	9.29	9.29	8.27	7.88	7.18
2	12.45	12.07	12.01	11.92	11.78	11.52	8.18	7.42	7.39	7.17	5.40	5.08	3.95
3	12.28	12.10	12.07	12.02	11.98	11.96	11.86	11.79	11.68	9.46	9.13	7.55	7.34
4	12.21	11.95	11.89	11.82	11.75	11.72	11.72	11.70	11.67	11.62	11.63	11.62	11.64

Table 6. The measurements results when connected are loaded to each battery separately when the generator is turned off

Battery number	Timestamps in minute												
	0	15	30	45	60	75	90	105	120	135	150	165	180
1	12.18	1.96	0										
2	12.26	4.35	0.58	0									
3	12.19	2.83	0										
4	12.18	11.68	11.46	11.08	5.13	2.34	0.36	0					

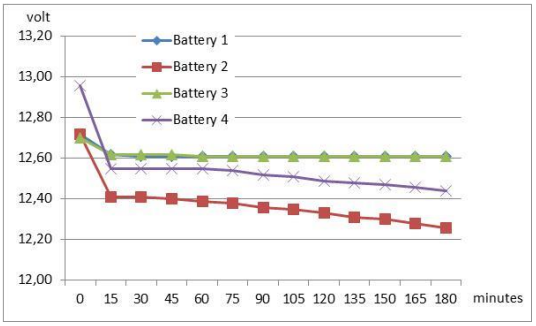


Fig. 9. The generator operation at the frequency 100 Hz, load 20 W

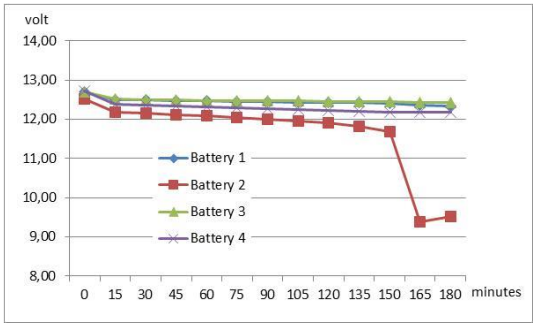


Fig. 10. The generator operation at the frequency 200 Hz, load 20 W

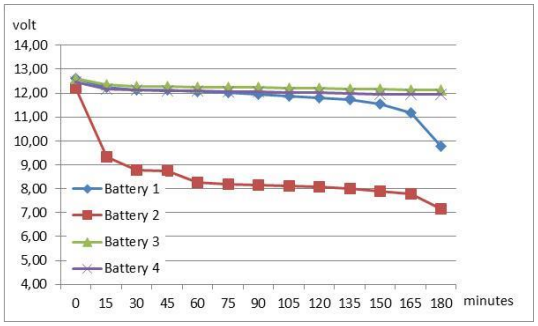


Fig. 11. The generator operation at the frequency 300 Hz, load 20 W

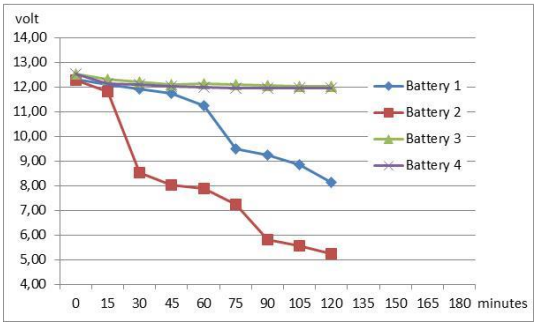


Fig. 12. The generator operation at the frequency 400 Hz, load 20 W

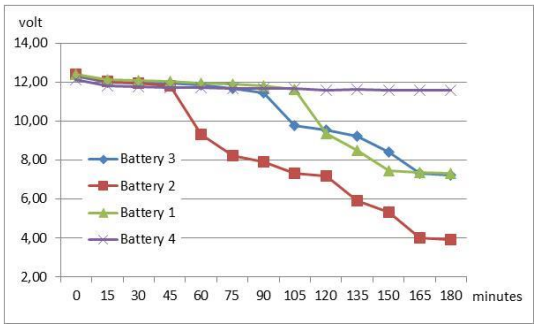


Fig. 13. The generator operation at the frequency 500 Hz, load 20 W

Since in the generator mode, 4 batteries worked on the load, and in the control measurement mode, each battery was connected to the same load separately, it would be wrong to compare the results without adjusting the time intervals. If when measuring the voltage, the battery has a certain voltage, then in the generator mode – 4 batteries – the same voltage should be obtained in a time period that is 4 times greater.

Thus, it is logical to compare voltage measurements on the battery, for example, at the 15th minute in mono mode with measurements on the same battery in the generator mode at the 60th minute of operation. Table 7 and Fig. 14 displays the time-corrected measurement results of each battery when the load is directly connected.

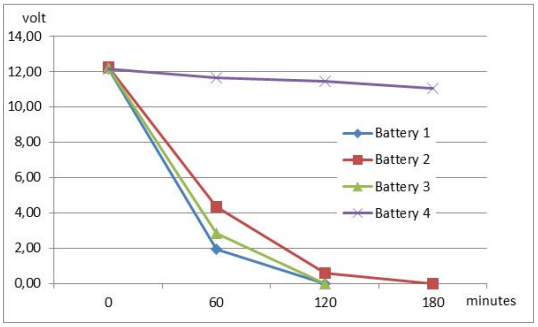


Fig. 14. The battery parameters measurements in mono mode

Table 7. The measurements results when connected are loaded to each battery separately when the generator is turned off

Battery number	Timestamps in minute			
	0	60	120	180
1	12.18	1.96	0	
2	12.26	4.35	0.58	0
3	12.19	2.83	0	
4	12.18	11.68	11.46	11.08

Analyzing figure 14, we can conclude that charging the batteries in mono mode is performed by discharging batteries 1, 2, and 3. In this way, the use of the proposed circuit 4 of the Tesla battery generator makes it possible to distribute the charge between the batteries and increase the time of their effective operation by approximately 30%.

4. Conclusions

In the research course, a device based on Nikola Tesla's 4-battery generator was created.

After analyzing the existing options for building similar systems, a structural scheme was determined. This device has a construction structure that allows you to fully define the requirements for the problems solved by the device, as well as the specified technical requirements. Also, in the performing work process, appropriate technical indicators were achieved that meet the technical task requirements. And also the product quality level is ensured, which, thanks to its simplicity and flexibility, allows for further development, changes in schematic and technical implementation and conducting research in the future.

Advantages:

- the scheme simplicity
- ease of programming the microcontroller
- possibility to change the frequency
- the control system protection due to the galvanic separation between the generator and the field transistors
- longer battery life within one charge and discharge cycle.

In the future, it is possible to improve this device, namely to change the frequency by changing the microcontroller software code. It is also possible to additionally connect various sensors, executing devices and information display devices. It is also possible to reconfigure the generator power part to reduce the temperature of the transistor keys and, as a result, increase the device efficiency.

References

[1] Arabsalmanabadi B. et al.: Charging Techniques in Lithium-Ion Battery Charger: Review and New Solution. IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society, Washington, DC, USA, 2018, 5731–5738 [http://doi.org/10.1109/IECON.2018.8591173].

[2] Azahan N. A. N. et al.: Analysis of relationship between acceleration and battery state-of-charging in electric vehicle. 2016 IEEE International Conference on Power and Energy (PECon), Melaka, Malaysia, 2016, 150–154 [http://doi.org/10.1109/PECON.2016.7951550].

[3] Chau K. T.: Overview of Electric Vehicle Machines – From Tesla to Tesla, and Beyond. 2016 International Conference of Asian Union of Magnetics Societies (ICAUMS), Taiwan, 2016, 1–6 [http://doi.org/1010.1109/ICAUMS.2016.8479941].

[4] Fakhrurrazey F. S. et al.: Performance comparison of 4-Pole Neodymium Magnet Bedini SSG free energy generator. IEEE 8th International Power Engineering and Optimization Conference (PEOCO2014), Langkawi, 2014, 573–578 [http://doi.org/10.1109/PEOCO.2014.6814494].

[5] IEEE Recommended Practice for Installation, Maintenance, Testing, and Replacement of Vented Nickel-Cadmium Batteries for Stationary Applications. IEEE Std 1106-2015 (Revision of IEEE Std 1106-2005) , 2015, 1–49 [http://doi.org/10.1109/IEEESTD.2015.7328671].

[6] Kostiuchko S. et al.: The Auxiliary Parametric Sensitivity Method as a Means of Improving Project Management Analysis and Synthesis of Executive Elements. Miraz M. H., Southall G., Ali M., Ware A., Soomro S. (eds): Emerging Technologies in Computing. iCETiC 2021. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 395, 2021, 174–184 [http://doi.org/10.1007/978-3-030-90016-8_12].

[7] Kumar P., Kumar V.: Energy storage options for enhancing the reliability of Power system in the presence of Renewable Energy Sources. Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, 2020, 1071–1076 [http://doi.org/10.1109/ICIRCA48905.2020.9183349].
[8] Marin-Garcia G. et al.: Battery Types and Electrical Models: A Review. IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC), Ixtapa, 2020, 1–6 [http://doi.org/10.1109/ROPEC50909.2020.9258711].
[9] Nikola Tesla's 4-Battery Switch. https://www.scribd.com/document/231006639/Nikola-Tesla-s-4-Battery-Switch?utm_medium=cpc&utm_source=google_pmax&utm_campaign=3Q_Google_Performance-Max_RoW&utm_term=&utm_device=c&gclid=Cj0KCQjwkqSlBhDaARIsAFJANkgE7m09BuFdzYb_vMstgTJNoDd8HDbVEP8W2wqoe8fuKyjoqDuFA20aAmMFEALw_wcB# (available: 08.07.2023).
[10] Sayemul I., Naruttam Kumar R.: Renewables integration into power systems through intelligent techniques: Implementation procedures, key features, and performance evaluation. Energy Reports 9, 2023, 6063–6087 [http://doi.org/10.1016/j.egyr.2023.05.063].
[11] Suarez C., Martinez W.: Fast and Ultra-Fast Charging for Battery Electric Vehicles – A Review. IEEE Energy Conversion Congress and Exposition (ECCE), Baltimore, 2019, 569–575 [http://doi.org/10.1109/ECCE.2019.8912594].
[12] Takayama K. et al.: A massive-ion beam driver for high-energy-density physics and future inertial fusion. Physics Letters A. 384(27), 2020, 126692 [http://doi.org/10.1016/j.physleta.2020.126692].
[13] Tar B., Fayed A.: An overview of the fundamentals of battery chargers. IEEE 59th International Midwest Symposium on Circuits and Systems (MWSCAS), Abu Dhabi, 2016, 1–4, [http://doi.org/10.1109/MWSCAS.2016.7870048].
[14] Wencai S. et al.: Optimal scheduling of thermal-photovoltaic power generation system considering carbon emission. Energy Reports 9, 2023, 1346–1356 [http://doi.org/10.1016/j.egyr.2023.04.205].
[15] Xu C. et al.: A Survey on Mechanical Switches for Hybrid Circuit Breakers. IEEE Power & Energy Society General Meeting (PESGM), Atlanta, 2019, 1–5 [http://doi.org/10.1109/PESGM40551.2019.8973674].
[16] Zhou K. et al.: Research and Development Review of Power Converter Topologies and Control Technology for Electric Vehicle Fast-Charging Systems. Electronics 12, 2023, 1581 [http://doi.org/10.3390/electronics12071581].
[17] Zhu M. et al.: Effect of XLPE/EPDM Interface on Space Charge Behavior in Cable Accessory. IEEE Access 7, 2019, 183554–183564 [http://doi.org/10.1109/ACCESS.2019.2960470].

Ph.D. Mykola Polishchuk
e-mail: polishchuk.kolia@gmail.com

Deputy dean for partnership Faculty of Computer and Information Technologies, associate Professor of the Computer Engineering and Cybersecurity Department Lutsk National Technical University.

http://orcid.org/0000-0002-1218-5925

Ph.D. Serhii Grinyuk
e-mail: sgrunjuk@lutsk-ntu.com.ua

Vice-dean for youth policy and career guidance Faculty of Computer and Information Technologies, Lutsk National Technical University.

http://orcid.org/0000-0002-0080-3167

Ph.D. Serhii Kostiuchko
e-mail: s.kostiuchko@gmail.com

Associate Professor of the Computer Engineering and Cybersecurity Department. Master of Mathematics and Cybersecurity. CISCO network academy trainer. Secretary of scientific journal "Technical news" and international school of mathematical modeling AMSE-UAPL.

http://orcid.org/0000-0002-1262-6268

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

Vice-dean for R&D Faculty of Computer and Information Technologies, Lutsk National Technical University.
Member of European Alliance for Innovation (EAI), International Association for Technological Development & Innovations (IATDI).

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

Ph.D. Pavlo Savaryn
e-mail: savaryn.pasha@lntu.edu.ua

Head of Distance Education and Web Technologies Sector, Associate Professor in Department of Digital Educational Technologies, Lutsk National Technical University.
Author and co-author 48 scientific articles, 1 monograph, 1 textbook and more than 20 methodical editions.

http://orcid.org/0000-0001-7140-7068

http://doi.org/10.35784/iapgos.3679 *received: 25.05.2023 | revised: 19.07.2023 | accepted: 24.09.2023 | available online: 30.09.2023*

REMOTE SOTA ALGORITHM FOR NB-IOT WIRELESS SENSORS – IMPLEMENTATION AND RESULTS

Piotr Szydłowski, Karol Zaręba
Efento sp. z o.o., Cracow, Poland

Abstract. *In this paper we share our experience with remote software updates for NB-IoT devices. The experience was collected over the years, when managing a fleet of tens of thousands of NB-IoT wireless sensors deployed worldwide by our customers. The paper discusses the main concerns that must be taken into account when designing the remote software over the air (SOTA) update mechanism, describes the remote update algorithm developed and used by us and presents the achieved experimental results based on remote software update of 5 000 NB-IoT sensors deployed in 10 European countries.*

Keywords: Internet of Things, wireless sensors, NB-IoT, software over the air

ALGORYTM ZDALNEJ AKTUALIZACJI OPROGRAMOWANIA W BEZPRZEWODOWYCH SENSORACH NB-IOT – IMPLEMENTACJA I REZULTATY

Streszczenie. *W tym artykule dzielimy się naszymi doświadczeniami ze zdalnymi aktualizacjami oprogramowania w urządzeniach NB-IoT. Doświadczenie zbieraliśmy przez lata, zarządzając flotą dziesiątek tysięcy czujników bezprzewodowych, które używane są na całym świecie przez naszych klientów. W artykule omówiono główne zagadnienia, które należy wziąć pod uwagę przy projektowaniu mechanizmu zdalnej aktualizacji oprogramowania (SOTA), opisano algorytm zdalnej aktualizacji opracowany i wykorzystywany przez nas oraz omówiono eksperymentalne wyniki aktualizacji oprogramowania na podstawie aktualizacji 5 000 czujników NB-IoT pracujących w 10 krajach europejskich.*

Słowa kluczowe: internet rzeczy, sensory bezprzewodowe, NB-IoT, zdalna aktualizacja oprogramowania

Introduction

One of the key challenges for IoT solution providers is to develop an efficient remote software update method for hundreds of thousands of battery-powered devices deployed around the world. The lack of secure, remote firmware updates is identified as a key security issue for IoT devices, by both organisations dedicated to security like OWASP [1] and researchers that test particular IoT solutions [3].

This may be one of the key issues that slow down the IoT adoption, as many users, especially large organisations, are not able to use IoT devices that do not meet their security requirements and standards. Moreover, the security breaches resulting from the firmware update may even lead to serious injuries of the devices' users. The most famous example of it dates back to 2015, when the researchers got almost full control over a car by injecting the malicious code into its entertainment system [1]. A recent example of large scale firmware related security breach is Mozi [10] – a botnet composed of over 1.5 million of non-computer devices used to launch DDoS attacks and steal the data.

There are many remote update mechanisms that have been used for years to update operating systems, software or hardware. However, it's impossible to implement the same mechanisms and algorithms in battery powered IoT devices due to their constrained resources (RAM, ROM, CPU) and the fact that these mechanisms require the updated device to stay long in the active state – the state, when the device communicates with the server and consumes a lot of energy.

Recently, there have been few attempts to standardise the IoT devices software updates mechanisms, which resulted in documents describing the best practices, firmware update architecture (e.g. IETF [4]) or documentation describing the updates (OMA [6]).

There are several key concerns that have to be taken into account when designing an update solution for IoT devices:

- The update must be secure – there are many security aspects that should be taken into account, but all of them are grouped in the following categories of the security threats, based on the S.T.R.I.D.E model [5]: spoofing identity, tampering with data, repudiation, information disclosure, denial of service, elevation of privilege.
- The update process must be safe – in case there are any issues with the update process, the device should be able to automatically roll back to the older version of the software.
- The update process must not impact the device's operations – e.g. in case of IoT sensors the update should not block taking the measurements for a long time (or ideally, not block it at all).
- The update package must be small – there are few very important reasons behind this:
  - The smaller the file the shorter the communication time with the server. In case of battery powered devices this has a direct impact on the battery life time.
  - In case of the cellular IoT devices the update package size has a direct impact on the cost of ownership, as the users are billed for the data sent / received over the mobile network.
  - According to different researchers, the number of cellular IoT devices will grow significantly in the coming years [8]. Sending large amount of data required to perform updates on these devices over the cellular networks that are optimised for small data packages (NB-IoT) can cause serious problems on the network side.
  - The update package must be small enough to be processed by an IoT device with constrained resources (memory and processing power).

All the above requirements are considered and satisfied in our remote differential SOTA update algorithm presented in section 2. Section 3 presents experimental results that show all aspects and advantages provided by our approach.

1. Typical method of Software over the air (SOTA) updates

Many IoT solution providers allow remote updates on their devices. The typical software over the air update procedure is simple: a development team compiles the code after the modifications and generates an entirely new software image, which is then distributed to the devices over the air [12]. However, the typical Over the Air (OTA) update is very troublesome, with millions of battery-powered sensors based on cellular networks (NB-IoT and LTE-M). The amount of data transmitted can not only cause destabilisation of the network but also seriously burden those devices that are optimised for short transmission sessions.

2. Proposed solution

2.1. Key requirements

We develop and produce wireless NB-IoT sensors used by customers around the globe. To provide the end users with the best experience and address the potential issues resulting from software bugs we had to implement an effective, remote software over the air update mechanism. On top of the general software over the air updates requirement described in section 1, the update process must be handled by the wireless sensors with the constrained resourced. The sensors are based on 64 MHz Cortex-M4, 512 KB Flash, 64 KB RAM and the only power source used by them are batteries.

2.2. Proposed SOTA update algorithm

We have developed an update process that seems to fit the IoT devices very well. The solution incorporates a differential update mechanism – a software update system based on sending only the changes in the software to the already-deployed devices. This update method has many advantages over the classic OTA updates, including:

- the amount of data transmitted to the device is several times reduced compared to the classic OTA update;
- the network load is minimised;
- longer device battery life is ensured, as due to the small size of the update package, the device spends less time in the active (connected) mode,
- the measuring system is not destabilised by the lack of measurements during the long update process.

The update process was designed to be used with UDP protocol, as it is supported by NB-IoT networks and minimises the data sent between the sensors and the servers. On top of the UDP we use the CoAP as it is one of the most popular and widely used IoT protocols and thanks to its small overhead it can be used by almost any IoT device.

The security threats are taken into the account during the update process. The developed algorithm guarantees that each of the devices will check the authenticity of the software, make sure that the update file was not corrupted during the transmission over the NB-IoT network and, if it was, request the server to resend the missing packages. The full encryption of the software transmission between the server and IoT devices makes it impossible for any third party to interfere the update process. Thanks to the server authentication and the mechanisms that verify the correctness of the received differential software file, the software update process is completely secure.

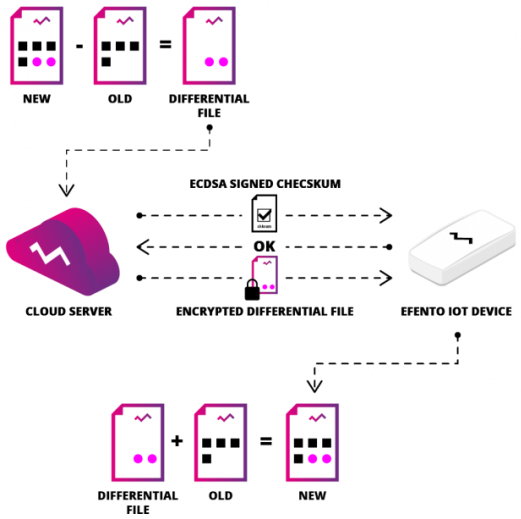


Fig. 1. Remote software over the air update developed by us

The security mechanisms incorporated in the SOTA update mechanism are based on the Elliptic-curve Diffie–Hellman (ECDH) [9] key agreement protocol. Each of the sensors has its own pair of public and private keys, securely generated and flashed on each device during the production. The server's pair of public and private keys are generated for each update session. The differential software file is signed during the compilation using the ECDSA SEC256k1 algorithm [11] with a private key securely stored in our CI infrastructure. The overview of the SOTA update mechanism is presented on Fig. 1.

The new SOTA update algorithm consists of the following steps:

1. After completing the work on the new version of the software, we prepare a differential file that is the difference between the new and the current versions of the software. The software is signed with ECDSA signature.
2. The differential file is sent to the update server and IoT devices are notified of the available update.
3. Sensor sends a CoAP message to the update server with its current software version number and its own public key.
4. The update server calculates a shared secret from its own private key and sensor's public key. The first 16 bytes of the secret will be used as the AES128 encryption key for the communication with the sensor, for a single update session.
5. The update server responds to the sensor with an encrypted CoAP message that contains the new software version number, hash (SHA256) from the new software, ECDSA signature of the new software hash and CRC16 from the whole frame. Along with the encrypted payload the update server sends its public key (not encrypted).
6. The sensor receives the update server's public key and calculates a shared secret from the server's public key and sensor's private key. The first 16 bytes of it will be used as an AES128 encryption key, valid for a single update session.
7. The sensor decrypts the response from the update server and decides, based on the new software version number, whether the software update is needed. If the update is not needed, the sensor terminates the update procedure.
8. If the update is needed the sensor verifies the authenticity of the update file based on the software hash and ECDSA signature. If the authentication fails, the sensor terminates the update procedure
9. If the authentication is successful, the sensor requests the differential software file from the update server.
10. The update server provides the sensor with the differential software file by sending encrypted data packages over raw UDP sockets.
11. Once the download of all the packages is completed, the sensor checks, if none of the packages is missing. If any of the packages is missing, the sensor requests the missing package(s) using a CoAP message. Downloading of the missing packages may be repeated twice. If the sensor was not able to download and build the differential file, the update is terminated and the sensor will try to perform the update again at the next communication.
12. The sensor enters the bootloader. The bootloader calculates the hash (SHA256) from the new software (old software merged with the differential software file received from the update server). If the calculated hash matches the hash sent by the update server, the sensor performs the software update. Otherwise, the update process is terminated and sensor reboots with the old software version.
13. Once the software update is successfully finished, the sensor notifies the update server about that by sending a message with its new software version number.

3. Algorithm testing and results

The results are based on remote software updates performed by us on 5 000 NB-IoT sensors located in 10 European countries, operating in NB-IoT networks of different mobile operators. The network coverage in the places where the sensors operate varies. For the simplicity of presenting the results, we decided to group the devices by their Coverage Enhancement Level (ECL) [2] – a parameter, dynamically set by an NB-IoT device based on the signal quality indicators (RSSI, RSRP, RSRQ). Each ECL determines the number of times downlink and uplink messages can be repeated to reach devices in poor coverage and the number of repetitions in each ECL is predefined by the network. ECL can have one of three values: 0 – used when coverage is good, 1 – used with moderate coverage, 2 – used with poor coverage.

Out of the 5 000 devices, 2 964 were in good coverage (ECL0), 1 424 were in moderate coverage (ECL1) and 612 were in poor coverage (ECL2).

The maximum number of the update attempts (point 11 of the Proposed SOTA update algorithm description in section 2) was set to 3. If a device had not been able to successfully perform the software update three times, the update server will not initiate the process anymore.

We defined the following metrics for the evaluation of the remote SOTA algorithm performance:

- Success rate - how many devices out of the whole test batch (5 000) successfully performed the software update. We also analysed the impact of the signal related parameters on the SOTA update
- Data usage – data required to send the update file to the device over the NB-IoT network.
- Energy consumption – energy consumed by a sensor for the update process.
- Downtime – total time during which the sensor is not able to perform its regular operations (taking the measurements and sending it to the server) due to the update process.

3.1. Success rate

Out of 5 000 devices, 4 988 (99.76%) managed to successfully perform the software update. Some of the devices required more than one attempt to successfully update their software. The 12 devices that failed to update the software were located in poor coverage (ECL2).

Table 1. Results of the remote software update of 5 000 devices

Coverage	Number of devices	Successful update at the 1st attempt	Successful update after repetition(s)	Success rate
ECL0	2 964	2 842	122	100%
ECL1	1 424	1 066	358	100%
ECL2	612	172	428	98%
Total	5 000	4 080	908	99.76%

3.2. Data usage

The size of the differential software file depends on the changes implemented in the software – the larger the changes, the larger the differential file. The total size of the current version of sensors' software is 250 kB. Based on the history of the remote updates performed by us, the smallest differential file was 7 B, the largest 33 kB and the average size of the differential software file was 21 kB. Using the remote update mechanism based on the differential files decreases the data consumption for the remote software update by 229 kB (91.6%) on average. This is a large difference as during its regular operations, with the measurement interval set to 5 minutes and the transmission interval set to 60 minutes, a single sensor consumes 183 kB of data per month.

3.3. Energy consumption

During the tests performed in our office, in good network coverage (ECL0), the total energy consumption during the update was 0.07 mAh for the differential update file of 7 B and 0.5 mAh for the differential file of 33 kB. As the majority of the energy consumed during the update is used to download the update file from the server, the conclusion is clear: the larger the update file, the more energy is required to perform the update. The total capacity of the batteries used in our wireless sensors is 6 300 mAh, so the remote update process based on the differential update files does not drain the battery too much.

The energy consumption during the remote software update process may vary, depending on the network coverage. As in the poor network coverage (ECL2), some packets of the differential file may not be delivered to the sensor at the first attempt and may require repetition(s), the energy consumption will be higher. Due to the specification of the test setup (5 000 devices are deployed in 10 countries around Europe), we were not able to measure to energy consumption of every single device during the update process and asses the network coverage impact on the energy consumption.

3.4. Downtime

As the majority of the remote software update tasks are performed by the main application and only the verification of the software hash and the update itself are performed in the bootloader, the device downtime is minimised. The total downtime (time required to perform the update and restart the device once the update is finished) is 24.5 seconds, no matter what is the update file size.

As the new software file is built by the device once the download of the differential file is completed, the network coverage and the number of repetitions required to deliver the differential file have no impact on the device's downtime.

4. Conclusions

The software over the air algorithm developed and implemented by us meets the security and performance requirements for the remote software update. The conducted tests proved, that the software update mechanism is reliable and the impact on both sensors' batteries and the network is minimised. The proposed software over the air algorithm can be successfully used to perform the updates not only of NB-IoT wireless sensors but to update any type of IoT devices with constrained resources.

The update process based on sending only the differential file to the wireless sensors decreases the data usage, what is beneficial for their owners (lower amount of data sent / received by the devices equals lower fees paid to the mobile operator), but also decreases the impact of the update process on the stability of the NB-IoT network.

We did not notice any impact of the particular operator's NB-IoT network configuration on the update process. The devices that failed to update their software were located in different countries and the common point, that made the update impossible was the poor network coverage.

We think, that it would be possible to achieve 100% success rate (remotely update all the devices, including the 12 devices located in poor coverage that were not able to perform the update), if the maximum number of the allowed update attempts was increased. This however would also increase the battery consumption and decrease the device life time.

Due to the tests specification (devices are deployed in 10 countries) we were not able to fully assess the impact of the network coverage on the energy consumption during the update. In order to get the full picture of that, further tests are required.

Acknowledgments

The research described in this paper was part of project „Opracowanie systemu monitoringu danych fizycznych opartego na technologii NB-IoT / LTE Cat M1 i platformie w chmurze, z wykorzystaniem nowej generacji bezprzewodowych sensorów", co-financed by European Union, program "Regionalny Program Operacyjny Województwa Małopolskiego na lata 2014-2020"

References

[1] Greenberg A.: Hackers Remotely Kill a Jeep on the Highway, With Me in It. [http://www.wired.com] (21.07.2015).
[2] Khan S. M. Z.: Narrowband Internet of Things (NB-IoT): from Radio Network Coverage to Device Energy Consumption Modeling and Energy-Efficient Application. Tallinn University of Technology [http://doi.org/10.23658/TALTECH.7/2022].
[3] Klinedinst D., King C.: On Board Diagnostics: Risks and Vulnerabilities of the Connected Vehicle. Software Engineering Institute, Carnegie Mellon University, 2016.
[4] Moran B. et al.: A Firmware Update Architecture for Internet of Things. RFC 9019. Internet Engineering Task Force, April 2021.
[5] Moran B. et al.: A Manifest Information Model for Firmware Updates in IoT Devices. RFC 9124. Internet Engineering Task Force, 2021.
[6] OMA SpecWorks, OMA LightweightM2M (LwM2M) Object and Resource Registry, 2023.
[7] OWASP IoT Security Team, OWASP Internet of Things Top10, 2018.
[8] Sinha S.: State of IoT 2023: Number of connected IoT devices growing 16% to 16.7 billion globally, IoT Analytics, May 24, 2023.
[9] Standards for Efficient Cryptography, Certicom Research, SEC 1: Elliptic Curve Cryptography, 2009.
[10] Tu T. F. et al.: A comprehensive study of Mozi botnet. International Journal of Intelligent Systems 37, 2022, 6877–6908 [http://doi.org/10.1002/int.22866].
[11] Vanstone S. A.: Responses to NISTs Proposal. Communications of the ACM 35(7), 1992, 50–52.
[12] Zandberg K. et al.: Secure Firmware Updates for Constrained IoT Devices Using Open Standards: A Reality Check. IEEE Access 7, 2019, 71907–71920 [http://doi.org/10.1109/ACCESS.2019.2919760].

M.Sc. Eng. Piotr Szydłowski
e-mail: piotr.szydlowski@efento.pl

Cofounder and CEO of Efento Ltd., a company that designs, develops and produces wireless sensors and platform for device management. Prior funding Efento, he had experience in consulting (management, technology).

He graduated from AGH University of Krakow in 2011 and holds magister degree in computational physics.

http://orcid.org/0009-0007-1384-2011

M.Sc. Eng. Karol Zaręba
e-mail: karol.zareba@efento.pl

Cofounder and CTO of Efento Ltd., a company that designs, develops and produces wireless sensors and platform for device management. Prior funding Efento, he had experience at various software development positions.
He graduated from AGH University of Krakow in 2011 and holds magister degree in automation and robotics.

http://orcid.org/0009-0009-7886-5163

http://doi.org/10.35784/iapgos.3723 *received: 08.06.2023 | revised: 19.07.2023 | accepted: 25.09.2023 | available online: 30.09.2023*

DEVELOPMENT OF A SOFTWARE SYSTEM FOR PREDICTING EMPLOYEE RATINGS

Gulnar Balakayeva, Dauren Darkenbayev, Mukhit Zhanuzakov
Al-Farabi Kazakh National University Almaty, Kazakhstan

Abstract*. The article deals with the problems of increasing the efficiency of enterprises and developing a system for assessing the rating of employees of an enterprise. Aspects of motivation for compliance with standards by employees and structural divisions are also considered; they were studied because employees are the main driving force of any enterprise. In order to properly evaluate it, many companies need rating systems. In this article, the authors conducted a study and developed a software package for assessing employees of an enterprise. This article describes how to develop a ranking prediction system. MongoDB is used as a database, as the fastest way to innovate, providing the flexibility and ease of use of a database.*

Keywords: employee ratings, employee evaluation system, Python, MongoDB

ROZWÓJ OPROGRAMOWANIA DO PRZEWIDYWANIA OCEN PRACOWNIKÓW

Streszczenie*. Artykuł dotyczy problemów zwiększania efektywności przedsiębiorstw i opracowania systemu oceny pracowników przedsiębiorstwa. Uwzględniono także aspekty motywacji do przestrzegania standardów przez pracowników i działy organizacyjne; zostały one zbadane, ponieważ pracownicy są główną siłą napędową każdego przedsiębiorstwa. Aby właściwie ją ocenić, wiele firm potrzebuje systemów oceny. W tym artykule autorzy przeprowadzili badanie i opracowali pakiet oprogramowania do oceny pracowników przedsiębiorstwa. W tym artykule opisano, jak opracować system przewidywania rankingu. MongoDB jest używany jako baza danych, jako najszybszy sposób na wprowadzanie innowacji, zapewniając elastyczność i łatwość użytkowania bazy danych.*

Słowa kluczowe: oceny pracowników, system ocen pracowników, Python, MongoDB

Introduction

The employee rating system is based on parallel motivation, which provides for giving the personnel management system such characteristics that would allow any employee to apply his abilities in the area in which he is strong and receive satisfaction in all categories of needs. The possibility of this proves once again that the contradictions between hierarchical theories and theories of parallel needs can be removed. Undoubtedly, each employee has his own value system, which determines a unique set and ratio of motivating factors. Therefore, the rating system in the organization should provide employees with the widest and most flexible choice of motivating means, within which each employee who has achieved positive results chooses for himself what has the highest value for him. Rating is always an important assessment of individual skills and qualities. Therefore, it is necessary to develop a digital information system based on new technologies and a rating analysis system capable of determining the progress or regression of a user rating [9].

There is a wealth of research on employee evaluation in terms of performance, leadership, and other quality skills. In an organizational context, performance is usually defined as the degree to which a member of an organization contributes to the achievement of organizational goals [12]. To assess these skills, it is necessary to use tools aimed at providing high-quality feedback about the assessed employee. Because the long-term viability and competitiveness of any organization depend on its ability to evaluate the work of employees and study their contribution to the achievement of goals [10].

There are different approaches to employee evaluation: from traditional to advanced models and methods. This work will focus on using new BigData technologies to create a robust software application and model to provide accurate ranking analysis [7].

MongoDB is used to store module data, which makes it more reliable and efficient in terms of speed and storage. It features high availability, horizontal scaling, and geographic distribution. The MongoDB document model is the fastest way to innovate, providing flexibility and ease of use of the database [5].

If relational databases store rows, then MongoDB stores documents. Unlike strings, documents can store information that is complex in structure. A document can be thought of as a repository of keys and values. The key represents a simple label with which a certain piece of data is associated [4].

1. Stages of creating a rating score

What is the rating assessment of the company's employees, which is compiled to assess compliance with its own corporate standards and which we call "internal rating"? Our definition is as follows: a rating is a single system for assigning points and a method for calculating them in a format that is most convenient for visualization and analysis. A comprehensive rating assessment includes several stages:

- development of a methodology for rating and evaluating issues;
- compiling the questions of the questionnaire;
- preparation of rating forms;
- filling out forms based on the results of the study;
- general analysis (determination of the average level);
- recommendations for individual employees (“work on mistakes”).

The basis for the development of the rating methodology is the "mystery shopper" questionnaire, into which selected indicators from the standard are entered. The challenge is to assign weights to different questions. The final rating score is prepared on the basis of fixed standards and non-formalized criteria. When developing a methodology, it should be possible to make subsequent adjustments without losing the original rating data [13].

The development of the rating methodology and the assignment of weights is most often carried out by the company itself, but sometimes external consultants are involved in this, working together with managers. Weighting should be based on qualitative customer research on the importance and perceptions of individual sales and service experiences. It is advisable to compare the obtained data with the results of an expert survey among managers regarding which performance criteria are key and most fully reflect the goals of the organization.

2. User rating analysis. Rating model

Rating is a numeric value representing the employee`s current competence level. It ranges from 0 (lowest value) to 5 (highest value). The rating is a floating-point number (approximation to 2 decimal places).

During each month, the rating of this user will be calculated depending on the following methods of rating assessment [3]:
1. Surveys.
2. Reviews from colleagues.
3. Direct feedback.

A survey is a set of pre-formulated questions created and edited by enterprise managers. All employees of the enterprise at the end of each month participate in a survey where they are asked to evaluate their colleagues. Answer options range from 1 to 5. For example: question 1: "How would you rate the quality of task management of this employee?", response options: "1, 2, 3, 4, 5" (Fig. 1).

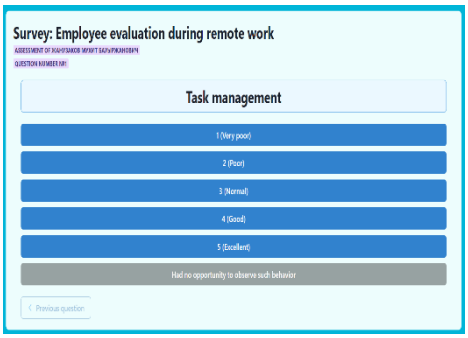


Fig. 1. Survey page

Surveys are optional. Because employees who are forced to take surveys will create biased data as these people usually give unreasonable scores just to complete the survey process as quickly as possible.

The data collected from these surveys will be used to make projections for the future months. That is why it is essential that survey data be as "clean" as possible.

3. Reviews from colleagues

Reviews are the evaluation of an employee in the form of text messages from his colleagues. Feedback can be positive (message type – compliment) or negative (message type – complaint). They are only seen by the managers of the enterprise, because of psychological reasons, as negative feedbacks may affect the employee`s motivation and mental state.

Feedbacks can be anonymous. When negative feedback is created, the system will decrease the current month rating by 10%, otherwise, when it is positive feedback, it will increase it by 15%. Because, positive comments are rarely received and usually harder to get.

If the user`s current rating is "null" (not assessed yet), its rating in case of negative feedback will be 4.5 (10% of 5), and when feedback is positive it will be 5.

User can send feedback by filling up the form in the figure 2.

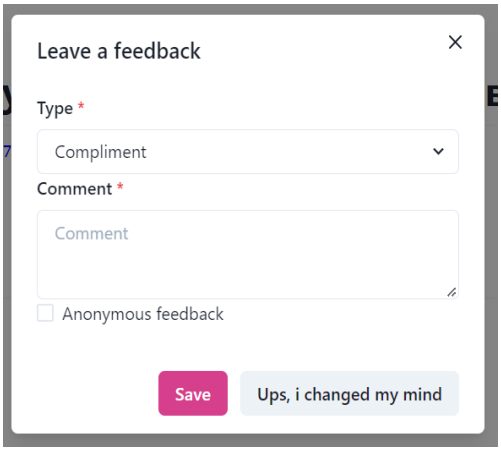


Fig. 2. Feedback form

The feedback type can be "compliment" or "complaint". The number of compliments and complaints received is also displayed, which will allow HR managers to identify unprofessional and positive employee skills in time (figure 3).

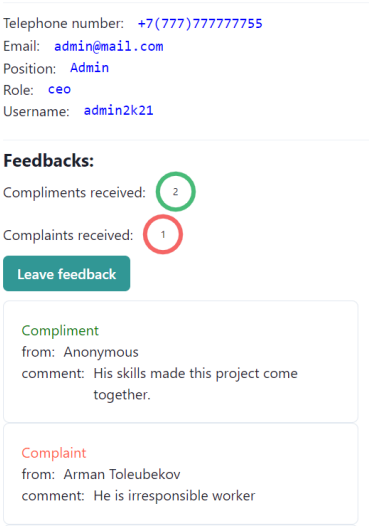


Fig. 3. An employee page with feedback received from colleagues

4. Customers direct assessments

Employees can be assessed directly by anyone, but mostly its clients, by going to the link from generated QR, which is provided after the service or can be obtained directly from the organization itself (figure 4). The link will open up the special form, which will ask the client to assess the current employee`s competence for the provided service from 0 to 5.

Fig. 4. An example of a generated link to a survey page

This value will increase/decrease the current rating by "client Assessment Value" percent. For example, employee`s current value is 3, if the client assessment was 4, then we will increase the current rating by 4%, which is 3.12. However, if the client assessment is below 3, then current rating will be decreased by "client Assessment Value" percent. Let`s say client assessment is 2, then current rating will become 2.94.

If the user`s current rating is "null" (not assessed yet), we will assume that his/her current rating is 1.

5. Development of software application for employee rating analysis

MongoDB Atlas was used as a database for storing ratings and users` data. User`s rating data was stored in "ratings" collections. Rating collection object properties have the following properities [1]:

```
{
        _id: String,
        userId: String,
        createdAt: Date,
        deletedAt: Date,
        data: Array of objects
}
```

Here, the "data" property is an array of individual rating data objects calculated for each month, starting from the month "createdAt" (figure 5).

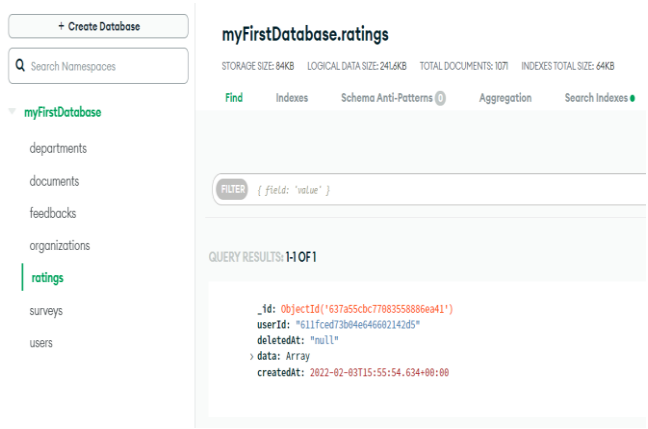


Fig. 5. "Rating" collection

Object "data" contains following props:

{
_id: String,
date: String in the format “month-year”,
value: Number (Float),
}

This data is used to display the rating graph by month. When a new user is created, we create an object in “rating” collection, with the following default values:

{
_id: created automatically by MongoDB,
userId: id of newly created user,
createdAt: current date (new Date()),
deletedAt: null,
data: [
{
_id: created automatically by mongoDB,
date: “01-2022”,
value: null,
}
]
}

6. Information system module architecture

In this work, we have developed a ranking prediction system using machine learning algorithms. In order to use this system to predict users in the system under development, a Python microserver was created.

It receives employee ratings for the current month for each competency from the main web server, passes the data to the employee rating prediction system, and as a result sends the output of the prediction system back to the server (figure 7).

Enterprise managers can view the rating reports of employees in a graph (figure 8).

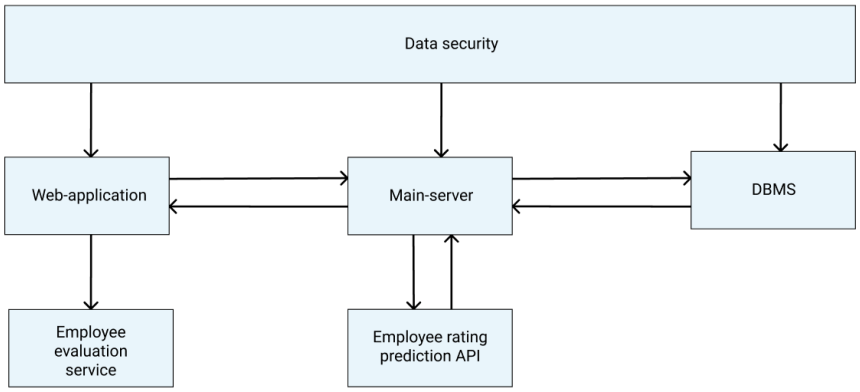


Fig. 7. Architecture of the module of the information system for analyzing the rating of employees

Fig. 8. Rating report for 2023 with predictions

7. Conclusion

The main problem of modern enterprises is the quality assessment of employees` productivity and their professional growth. In this thesis, we provide an information system module to solve this issue [11].

This paper described the developed model and software for assessing enterprise employees. The purpose is to create an information system module for employee rating analysis.

Methods for evaluating and predicting the work of employees based on machine learning algorithms have been developed to provide enterprises with a complete rating analysis using modern technologies. An information system for personnel management has been developed using modern software development tools [2]. A model for predicting the rating of employees was built using machine learning algorithms. A software application has been developed for analyzing and predicting the rating of enterprise employees [8].

A collection has been created in the MongoDB database. Rating assessment methods have been developed. An employee rating predicting service was integrated to the software.

Acknowledgments

This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant #AP09259208).

References

[1] Andersson J.: Enterprise Information Systems Management, [https://www.divaportal.org/smash/get/diva2:9165/FULLTEXT01.pdf].
[2] 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, [http://doi.org/10.35784/iapgos.3463].
[3] 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].
[4] 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.
[5] Balakayeva G. T., et al: Using NoSQL for processing unstructured Big Data. News of the National Academy of sciences of the Republic of Kazakhstan 6(438), 2019, 12–21.
[6] Duffy D.: Domain Architectures. Models and Architectures for UML Applications. Datasim Education BV, Amsterdam 2004.
[7] Fekete M., Rozenberg I.: The practical model of employee performance evaluation. International Conference: Human Capital without Borders: Knowledge and Learning for Quality of Life, 2014.
[8] Gill M.: Issues for consideration in mergers and takeovers from a regulatory perspective. BIS Review 60, 2000 [https://www.bis.org/review/r000721b.pdf].
[9] https://www.kreyonsystems.com/Blog/digitization-with-document-management/
[10] Islama R., Shuib bin Mohd R.: Employee performance evaluation by the ahp: a case study. Asia Pacific Management Review 11(3), 2006, 163–176.
[11] Panetto H. et al.: New Perspectives for the Future Interoperable Enterprise Systems. Computers in Industry 79, 2016, 47–63 [http://doi.org/10.1016/j.compind.2015.08.001].
[12] Panetto H. et al.: New Perspectives for the Future Interoperable Enterprise Systems, [https://www.archives-ouvertes.fr/hal-01142747/document].
[13] Stouffer K. et al.: Guide to Industrial Control Systems (ICS) Security, [https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-82r2.pdf].

Prof. Gulnar Balakayeva
e-mail: gulnardtsa@gmail.com

Doctor of Physical and Mathematical Sciences, professor of the Department of Computer Science, Faculty of Information Technologies, Al-Farabi Kazakh National University, Almaty, Kazakhstan.
Research interests: Development of big data processing systems, modeling of physical and chemical processes.

http://orcid.org/0000-0001-9440-2171

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.

http://orcid.org/0000-0002-6491-8043

Ph.D. student Mukhit Zhanuzakov
e-mail: zhanmuha01@gmail.com

Ph.D. student of KazNU named after al-Farabi.
Research interests: Web application development

http://orcid.org/ 0000-0003-0001-8422

http://doi.org/10.35784/iapgos.5351 *received: 26.07.2023 | revised: 10.09.2023 | accepted: 28.09.2023 | available online: 30.09.2023*

ENGINEERING AND TECHNICAL ASSESSMENT OF THE COMPETITIVENESS OF UKRAINIAN MECHANICAL ENGINEERING ENTERPRISES BASED ON THE APPLICATION OF REGRESSION MODELS

Anna Vitiuk[1], Leonid Polishchuk[2], Nataliia B. Savina[3], Oksana O. Adler[1], Gulzhan Kashaganova[4], Saule Kumargazhanova[5]

[1]Vinnytsia National Technical University, Department of Management, Marketing and Economy, Vinnytsia, Ukraine, [2]Vinnytsia National Technical University, Department of Industrial Engineering, Vinnytsia, Ukraine, [3]National University of Water and Environmental Engineering, Rivne, Ukraine, [4]Turan University, Faculty of Digital Technologies and Art, Department of Computer and Software Engineering, Almaty, Kazakhstan, [5]D. Serikbayev East Kazakhstan Technical University, Department of Information Technologies and Intelligent Systems, Ust-Kamenogorsk, Kazakhstan

Abstract. *The characteristic features of engineering products are revealed. Average industry performance indicators of mechanical engineering enterprises in Ukraine were formed. The competitiveness of mechanical engineering enterprises was studied. The integral indicator of the competitiveness of mechanical engineering enterprises in Ukraine was evaluated. It has been established that the competitiveness industry, despite certain profits received by enterprises, is in a systemic, predictable crisis and only individual enterprises that maintain their own line of economic behavior are successful, increase competitiveness and have prospects for further economic growth.*

Keywords: competitiveness, regression, mechanical engineering, dependence

INŻYNIERYJNO-TECHNICZNA OCENA KONKURENCYJNOŚCI UKRAIŃSKICH PRZEDSIĘBIORSTW BUDOWY MASZYN NA PODSTAWIE ZASTOSOWANIA MODELI REGRESJI

Streszczenie. *Ujawniono charakterystyczne cechy produktów inżynieryjnych. Opracowano średnie wskaźniki wydajności przemysłu przedsiębiorstw inżynierii mechanicznej na Ukrainie. Zbadano konkurencyjność przedsiębiorstw przemysłu maszynowego. Oceniono integralny wskaźnik konkurencyjności przedsiębiorstw przemysłu maszynowego na Ukrainie. Ustalono, że branża konkurencyjności, pomimo pewnych zysków uzyskiwanych przez przedsiębiorstwa, znajduje się w systemowym, przewidywalnym kryzysie i tylko pojedyncze przedsiębiorstwa, które utrzymują własną linię zachowań gospodarczych, odnoszą sukcesy, zwiększają konkurencyjność i mają perspektywy dalszego wzrostu gospodarczego.*

Słowa kluczowe: konkurencyjność, regresja, inżynieria mechaniczna, zależność

Introduction

A feature of mechanical engineering enterprises is the provision of fixed assets for other enterprises, the consequence of which is the further development of other branches of the economy. Therefore, ensuring the active development of mechanical engineering enterprises should be based on the maximum possible use of the conditions of spatial potential: chaotic and systemic trends in the development of the economy, the main directions of development of the corresponding enterprises, infrastructure connectivity, economic ties and partnership relations [10]. In such circumstances, the management of the enterprise, which involves the use of optimal resource provision, rational location of production, development of integration processes and ensuring the effective distribution of tasks will ensure the competitiveness of mechanical engineering enterprises [1, 13].

1. Formulation of the problem

Mechanical engineering is a basic branch of economic development in every country, however, each country has its own conditions and prospects for the development of machine-building industries. The fact remains indisputable that, in order to ensure progressive development, product manufacturers are forced to enter into a competitive struggle for better operating conditions. The result of the progressive development of society was the emergence and formation of the phenomenon of competition, which forces product manufacturers to constantly move and improve themselves, not to stop at the achieved results, that is, to increase their competitiveness [2, 4]. Competitiveness is the most important criterion for the expediency of an enterprise's activity, a condition for the efficiency of production activity, the basis for choosing means and methods of management, and a guarantee of success in competition [7, 11].

2. Results of the research and discussion

Having established the competitiveness of the enterprise as the most general characteristic of the development of the enterprise, we will evaluate the indicator according to the methodology presented by O. Kuzmin, O. Melnyk and O. Romanko [4] where the main emphasis is placed on highlighting competitiveness through indicators of financial and economic efficiency, indicators of production efficiency and indicators of commercial efficiency. The peculiarity of the method is taking into account the share of defective products that are manufactured and the presence of which can distort the further production process. In order to identify the share of defects in the manufacture of mechanical engineering products, it is important to take into account the complexity of the manufactured products.

The work of O. Danchenko, O. Zanora, V. Borkun [3] presents results where the share of defects in the manufacture of mechanical engineering products of various precision was clearly established. It has been proven that in the process of precise processing there is a defect, the costs of which are 2% of the total cost of processing blanks in the case of processing according to the 8th quality and 17% of the cost – according to the 7th quality. With a further increase in the accuracy of processing to the 6th quality, the cost of defects reaches 32% of the cost of processing blanks. Therefore, let's assume the following levels of shortage for the industries: agricultural, nuclear, electrical engineering, heavy and transport, construction, road and utility engineering, engineering for light and food industry, animal husbandry and fodder production – 2%; machine-tool, tool, chemical, petroleum and power engineering – 17%; instrumentation, production of automation and control equipment (32%). We will use 17% of defects of the total cost of processing blanks for researched enterprises [2, 14].

The effective tool for modeling the development of an enterprise and its activities is the formation of an integral indicator of the enterprise's competitiveness. To form the integral level of the enterprise's competitiveness, there is a need to translate indicators into relative values according to the developed scale presented in the work [5]:

$$X_1 = \begin{cases} \text{from 0 to 0.5 – if the indicator is worse than the baseline;} \\ \text{0.5 – if the indicator is completely identical to the base one;} \\ \text{from 0.5 to 1.0 – if the indicator exceeds the baseline.} \end{cases}$$

It is worth noting that the average industry or market indicators should be taken as the base. According to the noted expert scale, the above mentioned indicators are converted into point estimates, and weighted values are also assigned to them, which makes it possible to obtain integral levels of various components of the efficiency of the enterprise's functioning. Based on the generalization of the values of these levels using the geometric mean, the integral level of the enterprise's competitivness is determined (E_p):

$$E_c = \sqrt[3]{E_f \cdot E_p \cdot E_{com}} \quad (1)$$

The interpretation of the obtained results is carried out similarly: if E_p exceeds the value of 0.5 and approaches to 1, then we can talk about a high level of competitivness of the enterprise's functioning and exceeding the average industry or market average level according to this parameter, if $E_p = 0.5$, then this indicates a complete compliance with competitiveness standards formed in the industry, if E_p is less than the limit of 0.5, then this demonstrates significantly worse parameters of the enterprise's competitiveness compared to competitors.

Based on the specified parameters, we consider it expedient to assign $X_i = 0.2$, if the investigated indicator is lower than the industry average, and $X_i = 0.7$, if the investigated indicator exceeds the industry average, and $X_i = 0.5$ if the indicator is completely identical to the base one. Therefore, based on the analysis of the theoretical requirements for the financial indicators of the company's activity and statistical indicators, a system of average industry indicators was formed (table 1).

Table 1. Average industry performance indicators of mechanical engineering enterprises in Ukraine in 2020

A group of indicators	Partial indicators	Value
Indicators of financial and economic efficiency (E_f)	Coefficient of profitability of capital in mechanical engineering	-8..% [8]
	Total liquidity ratio	1.5…2.5, but not less than 1
	Autonomy ratio	more than 0.5
	Financial leverage ratio	1
	Asset turnover rati	>=1.0
Indicators of production efficiency (E_p)	Fund return (F)	0.65
	Material return (Mv)	2.73*
	Productivity (Pr)	823.24*
	Profitability of production (Rv)	-5.15
	The share of defects in commodity products	0.17
Indicators of commercial efficiency (E_{com})	Operating activities profit/Net sales revenue	-0.00005
	The number of orders completed on time/Total number of orders	≈1
	Warehouse occupancy level	≈1
	The term of repayment of payables	Reduction
	Repayment period of receivables	Reduction

3. Mathematical model of of the competitiveness of Ukrainian mechanical engineering enterprises

Using a methodical approach, a number of performance indicators of well-known mechanical engineering enterprises in Ukraine in 2017–2020 was formed (Tab. 2 i 3, Fig. 1).

Table 2. Integral indicator of competitiveness of mechanical engineering enterprises in Ukraine

Enterprise	Period			
	2017	2018	2019	2020
PJSC "Kharkiv Tractor Plant"	2.38	2.19	2.35	2.35
PJSC "Zaporizhzhia Automobile Plant"	2.06	1.86	2.03	2.33
PJSC "Kryukov Railway Car Building Works"	2.87	3.19	3.01	2.64
State Enterprise "Production Association Yuzhny Machine-Building plant named after A.M. Makarov"	1.71	1.69	2.69	1.69
PrJSC "Barskii Machine-Building Plant"	2.87	2.84	2.64	2.64
PrJSC "Kalynivsky Machine Building Plant"	2.31	2.13	2.43	2.13
PJSC "Nasosenergomash Sumy"	2.65	2.18	2.56	1.96

Table 3. Mathematical models of dependence of competitiveness of machine-building enterprises on profit mechanical engineering

Enterprises	Mathematical models	R^2
PJSC "Kharkiv Tractor Plant"	$y = 8E\text{-}17x^3 - 2E\text{-}11x^2 + + 4E\text{-}07x + 2.3926$	1
PJSC "Zaporizhzhia Automobile Plant"	$y = 3E\text{-}19x^3 - 1E\text{-}12x^2 + + 5E\text{-}07x + 2.2924$	1
PJSC "Kryukov Railway Car Building Works"	$y = -1E\text{-}17x^3 + 1E\text{-}11x^2 + - 2E\text{-}06x + 2.7512$	1
State Enterprise "Production Association Yuzhny Machine-Building plant named after A.M. Makarov"	$y = 3E\text{-}18x^3 + 5E\text{-}12x^2 + + 3E\text{-}06x + 2.247$	1
PrJSC "Barskii Machine-Building Plant"	$y = 4E\text{-}05x + 2.6016$	0.9959
PrJSC "Kalynivsky Machine Building Plant"	$y = 0.0612\ln(x) + 1.7247$	0.3001
PJSC "Nasosenergomash Sumy"	$y = 8E\text{-}12x^2 + 3E\text{-}06x + 2.1715$	0.998

The modeling of the competitiveness of the studied enterprises proved the existence of polynomial ($y = b_0 + b_1x + b_2x^2$), linear ($y = b_0 + b_1x$) and logarithmic relationships (y = aln(x) + b), between the competitiveness of mechanical engineering enterprises and their profit. All researched enterprises, with the exception of PrJSC "Kalynivsky Machine Building Plant" and PrJSC "Barskii Machine-Building Plant" were characterized by polynomial regression – that is, the relationship between the independent variable x (profit) and the dependent variable (competitiveness) y, which is modeled as a polynomial of the n^{th} degree from x. That is, the detected polynomial regression corresponds to a non-linear relationship between the value of x and the corresponding conditional mean value of y. At the same time, we note that the polynomial regression models were selected using the least squares method, which minimizes the variance of the unbiased estimates of the indicators [3].

Therefore, the competitiveness of the Ukrainian mechanical engineering enterprises is described by 3 three types of dependencies:

1. Polynomial – the vast majority of enterprises;
2. Logarithmic – PrJSC "Kalynivsky Machine Building Plant";
3. PrJSC "Barskii Machine-Building Plant".

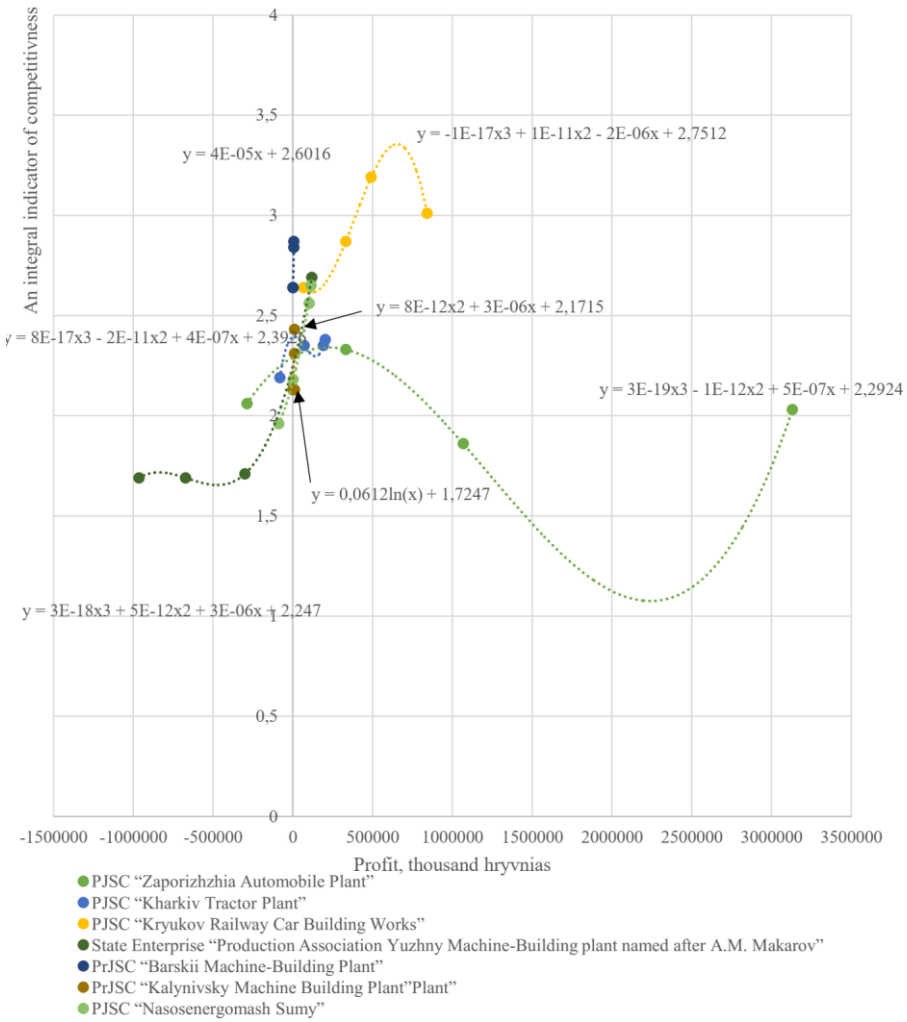


Fig. 1. Modeling the competitiveness of mechanical engineering enterprises in Ukraine

The degrees of the polynomial determine the number of extremes (determining factors of influence). Therefore, the polynomial of the third degree, which was recorded in the regression dependence of most enterprises, indicates the presence of two extremums, the use of which can allow a qualitative description of the development process of such mechanical engineering enterprises. The enterprises PJSC "Kharkiv Tractor Plant", PJSC "Zaporizhsky Automobile Plant", PJSC "Kryukov Railway Car Building Works", State Enterprise "Production Association Yuzhny Machine-Building plant named after A.M. Makarov" are characterized by two defining extremes, that is, the influence of two separate influencing factors. It is quite an interesting fact that the presence of the two most important factors of influence is characteristic of both large enterprises engaged in the production of a large number of heterogeneous products, and small enterprises operating in a narrow niche. The closeness of extremums on the graph indicates an increase in the influence of the time factor on the enterprise's activity, that is, an increase in its competitiveness, and, accordingly, the technical condition of these enterprises.

The logarithmic function revealed at the PrJSC "Kalynivsky Machine Building Plant" indicates that there is a large set of factors influencing competitiveness, the values of which change rapidly at first, and gradually stabilize over time, which indicates a significant adaptability of the enterprise to external conditions. Therefore, the company has the ability to control the main influencing factors and quickly react to the emergence of new circumstances.

The linear function in the development of PrJSC "Barskii Machine-Building Plant" reflects a direct relationship between the competitiveness of the enterprise and the level of its profit, which indicates the constancy of other factors influencing competitiveness, which can be ensured on the basis of high performance indicators of the enterprise in the previous period, which has already been proven in other studies [9].

Developing the prospects of the established enterprises, it is worth emphasizing that in 2021–2022 PJSC "Kryukov Railway Car Building Works", PJSC "Zaporizhzhia Automobile Plant", State Enterprise "Production Association Yuzhny Machine-Building plant named after A.M. Makarov" and PJSC "Kharkiv Tractor Plant" are located in regions where hostilities are ongoing. Therefore, as of the beginning of 2023, only 2 enterprises among researched have prospects for increasing competitiveness - PrJSC "Kalynivsky Machine Building Plant" and PrJSC "Barskii Machine-Building Plant". According to previous studies [8], the last two enterprises ensure the growth of competitiveness due to active foreign economic and research activities.

The totality of the identified trends in the development of mechanical engineering enterprises was evidenced on the basis of modeling the dependences between the absolute values of the company's profit and the relative values of the integral indicator of the competitivness. The extremely negative trends regarding the systemic impairment of the competitiveness of mechanical engineering enterprises in the past, which were additionally worsened with the beginning of the war, were evidenced. Therefore, the mechanical engineering industry, despite certain profits received by enterprises, is in a systemic, predictable crisis, and only individual enterprises that maintain their own line of economic behavior are successful, they increase competitiveness and have prospects for further economic growth.

4. Conclusions

The study of competitiveness was made based on the analysis of seven mechanical engineering enterprises wich engaged in different directions of production. As a result of the analysis, an integral indicator of competitivness of mechanical engineering enterprises in Ukraine was formed for each studied enterprise with dynamics over four years. The use of an integral indicator made it possible to simulate the value of the competitiveness of enterprises depending on the change in their profitability. The simulation results showed extremely negative trends regarding the systemic deterioration of the competitiveness of mechanical engineering enterprises.

References

[1] Azarova A., Zhytkevych O.: Mathematical methods of identification of ukrainian enterprises competitiveness level by fuzzy logic using. Economic Annals-XXI 9–10 (2), 2013, 59–62.
[2] Azarova A.: Information Technologies and Neural Network Means for Building the Complex Goal Program "Improving the Management of Intellectual Capital". Lecture Notes on Data Engineering and Communications Technologies 77, 2022, 534–547.
[3] Danchenko O., Zanora O., Borkun V.: The analysis of the lack in production of parts at the enterprise. Management of the development of complex systems 5, 2011, 132–136.
[4] Kuzmin O., Melnyk O., Romanko O.: Enterprise competitiveness: planning and diagnosis. IFNTUNH, Ivano-Frankivsk 2011.
[5] Polishchuk L., Mamyrbayev O., Gromaszek K.: Mechatronic Systems II. Applications in Material Handling Processes and Robotics. Taylor & Francis Group – CRC Press, Boca Raton, London, New York, Leiden 2021.
[6] Profitability of operational activities of enterprises by types of industrial activity [http://www.ukrstat.gov.ua].
[7] Shyian A. A. et al.: Modeling communication between the public and the authorities while implementing innovative projects in the context of e-democracy and public administration. Science and Innovation 16(6), 2021, 18–27.
[8] Vitiuk A., Nesen L., Kuran O.: Leading engineering enterprises of the Vinnytsia region: analysis of investment attractiveness. Investments: practice and experience 4, 2022, 71–77.
[9] Vitiuk A.: Engineering enterprises in Ukraine: technical and economic features. Economy and state 9, 2022, 57–62.
[10] Vitiuk A.: Identification of the content and significance of the spatial development of the enterprise. Business Inform 1, 2020, 359–366.
[11] Voynarenko M. P. et al.: Managing the development of innovation business processes with automated information systems. Marketing and innovation management 4, 2017, 133–148 [http://doi.org/10.21272/MMI.2017.4-12].
[12] Wójcik W., Pavlov S., Kalimoldayev M.: Mechatronic Systems I. Applications in Transport, Logistics, Diagnostics and Control. Taylor & Francis Group – CRC Press, London, New York 2021.
[13] Yaremko S., Kuzmina O., Novitskyi R.: Using artificial intelligence technologies for forecasting business processes. Computer-Integrated Technologies: Education, Science, Production 23, 2021, 230–234 [http://doi.org/10.36910/6775-2524-0560-2021-43-38].
[14] Yarmolenko V. et al.: Practice Analysis of Effectiveness Components for the System Functioning Process: Energy Aspect. Lecture Notes on Data Engineering and Communications Technologies 77, 2022, 282–296.

Ph.D. Anna Vitiuk
e-mail: anna_vitiuk@ukr.net

Ph.D., associate professor, Vinnytsia National Technical University, Vinnytsia.
Scientific direction: spatial development problems of engineering enterprises, investment activity, investment management of engineering entities.

http://orcid.org/0000-0001-5211-4948

D.Sc. Leonid Polishchuk
e-mail: leo.polishchuk@vntu.edu.ua

D.Sc., professor, Vinnytsia National Technical University, Vinnytsia.
Scientific direction: dynamics of drive systems with devices and control systems with variable operating modes and diagnostics of metal structures of hoisting-and-transport and technological machines

http://orcid.org/0000-0002-5916-2413

D.Sc. Nataliia B. Savina
e-mail: n.b.savina@nuwm.edu.ua

D.Sc. in Economics, professor, Vice-Rector for Research and International Relations in National University of Water and Environmental Engineering. Scientific directions are evaluation and forecasting of the efficiency of investments in entrepreneurial projects, in logistics and economic systems of various levels, in human capital. Deals with Financial instruments of international scientific and technical cooperation, modeling of the marketing business-processes under the industrial enterprise's competitiveness reflexive control, and modeling of the dialogue "human-computer" for ergonomic support of e-learning.

http://orcid.org/0000-0001-8339-1219

Ph.D. Oksana O. Adler
e-mail: oksana_adler1983@ukr.net

Associate Professor of department of Enterprise Economics and Production Management, Vinnytsia National Technical University.
Research interests: advanced methods of diagnosis and increasing the competitiveness of modern enterprises in a market economy.

http://orcid.org/0000-0002-4673-366X

Ph.D. Gulzhan Kashaganova
e-mail: guljan_k70@mail.ru

Associate professor of the Department of Computer and Software Engineering, Faculty of Digital Technologies and Art, Turan University, Almaty, Kazakhstan.
Scientific interests: fiber-optic technologies, information systems, information protection.

http://orcid.org/0000-0001-8150-1621

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.

http://orcid.org/0000-0002-6744-4023