Promoting the Public Health by Medical Image Analysis and Disease Diagnosis

Debarghya Biswas[1], Ankita Tiwari[2]
[1]*Assistant Professor, Department of CS & IT, Kalinga University, Raipur, India*
[2]*Research Scholar, Department of CS & IT, Kalinga University, Raipur, India*

KEYWORDS
Healthcare, Data Analytics, Blood Disorder Disease, Machine Learning, Classification, Feature Selection

ABSTRACT
Public health has improved as a result of data analysis tools since they make disease identification early and simple. In the absence of data analysis methods, medical professionals' judgement is the only means of identifying the illness. This could be problematic in places without access to skilled medical professionals, which could ultimately result in the illness patient's death. Data analysis techniques are currently being applied in numerous medical disease detection situations in order to address the aforementioned challenges. The methods for detecting blood disorders are presented in this article. This model detects blood disorder disease using a blood smear in addition to a clinical report. The sick patients are classified using an ML classifier after an ML model extracts information from blood smear images and combines them with clinical features.

1. Introduction

Data and images are important in the scientific and medical domains. Clinical data and photographs are critical to the analysis of medical domains. To comprehend the anatomical and functional mechanisms of any organ, it is imperative to view the high-resolution image. In addition to detecting the presence of a disease, medical imaging can also be used to explain the behaviour and function of numerous medical disorders [1]. Medical imaging uses certain data analysis techniques these days. Every data analysis method examines contaminated data and makes the appropriate adjustments. Many medical sectors use data analysis techniques like machine learning (ML) and deep learning (DL) to enhance patient outcomes. Machine learning is used in all scientific fields to examine data, recognise and learn patterns, and make judgements towards issue solutions without the assistance of humans.

Machine learning has demonstrated exceptional performance in a number of disciplines, including wireless sensor data, speech recognition, picture recognition, and traffic prediction [2]. These machine learning algorithms categorise medical diseases according to their symptoms, which are referred to as features of any disease and are used to discriminate between those with the condition and those who do not [3]. Learning challenging and discriminatory medical imaging features with hand-crafted and conventional machine-learning facilities remains an issue in traditional methodologies. As a result, these methods were unsuccessful when used for complicated diseases. In general, accuracy when using clinical data is not very great. Even in situations where accuracy is excellent, it can be challenging to gather data over an extended period of time. Machine learning and deep learning are used to classify diseases related to blood disorders [16]. This paper introduces ML and DL based lung disease detection techniques. Using characteristics, the proposed model correctly classifies this illness.

In this case, the introduction is examined in section 1 of the article while the pertinent literature is examined in section 2. Section 3 explains the goal of the work, Section 4 shows the results of the work, and Section 5 concludes up the project.

2. Literature Review

In essence, features-based machine learning and deep learning techniques may identify a wide range of medical conditions in people without the need for human intervention. Additionally, it offers a suitable degree of precision when categorising a medical illness process [4]. uses the k-nearest

clustering approach in [17] to classify white blood cells and red blood cells. Therefore, the aforementioned challenges might be more successfully overcome by an automated and repeatable process [5]. Thalassemia patients are categorised by Waranyu Wongseree and Nachol Chaiyaratana using genetic programming (GP) and multi-layer perceptrons (MLP) [6]. Neural networks and GP-based decision trees are used as classifiers in this. The 270 samples included in this problem come from both normal individuals and various thalassemia patient and trait kinds [11]. The performance of MLP with a single hidden layer and GP-based classification trees is identical to that demonstrated in [7]. On the other hand, GP-based classification trees perform worse than an MLP with two hidden layers. By modifying the data type to establish the suitable data type for each technique, Patcharaporn Paokanta et al. attempted to give a comparison between performances of several classification techniques using 127 β-thalassemia patients [18].

In his research, [9] employed principal component analysis (PCA), and the characteristics chosen were Hb, Ht, MCV, and RBC. Using 304 clinical records, they conducted a comparison study between k-NN, SVM, and MLP for thalassemia screening [10]. The study's findings indicate that the MLP classifier performs somewhat better than the SVM when there is a shortage of data. Utilising PCA Research for β-thalassemia knowledge discovery [19], this programme receives blood test results as input and reduces the data before classifying it. The MLP is the best, according to the results. Anand Upadhyay attempted to use artificial neural networks (ANN) to categorise individuals with β-thalassemia based on their quantitative blood test results [12]. The feed forward neural network in this study uses haemoglobin A (HbA), haemoglobin A2 (HbA2), and foetal haemoglobin (HbF) as inputs. For thalassemia screening, an unconstrained functional networks classifier was presented in [20], and its outcomes were compared to those of MLP and SVM. The outcomes demonstrate the dependability, adaptability, and superior performance of this unconstrained functional network classifier over the most popular ones now in use [22]. Certain hematologic characteristics are produced by hemochrome cytometric analysis, and Amendolia et al. [14] use these parameters to classify thalassemia using artificial neural networks. With 94% accuracy, several ANN combinations are used to classify thalassemia carriers and normal patients [13].

Over the years, numerous proposals for work in this topic have been made. Creating fresh samples, simulating data, augmenting data for tiny datasets, and transforming styles are a few examples. With the aid of sparse data, suggested strategies enhance or improve detection and classification accuracy [21]. When there are no doctors available to provide assistance, this model is adopted. In order to get around this, GA-based MLP machine learning applications are suggested for the medical domain, which can address the issue of blood condition diagnosis and classification [8].

3. Methodology

There are the following steps in the suggested model. For the purpose of the study, the entire number of birth instances is provided by the RBC image dataset that was obtained from the hospital, which is located in a remote location. Adaptive Diffusion Filter handles the preprocessing (ADF). Using the Grey Level co-occurrence of matrix method, feature extraction is carried out (GLCM). Machine learning heavily relies on feature selection, which is handled by genetic algorithms (GA). An approach to data analysis called machine learning automates the process of creating analytical models. Classification is the most well-known method for using machine learning algorithms [15]. The current study uses machine learning techniques to predict blood problem diseases.

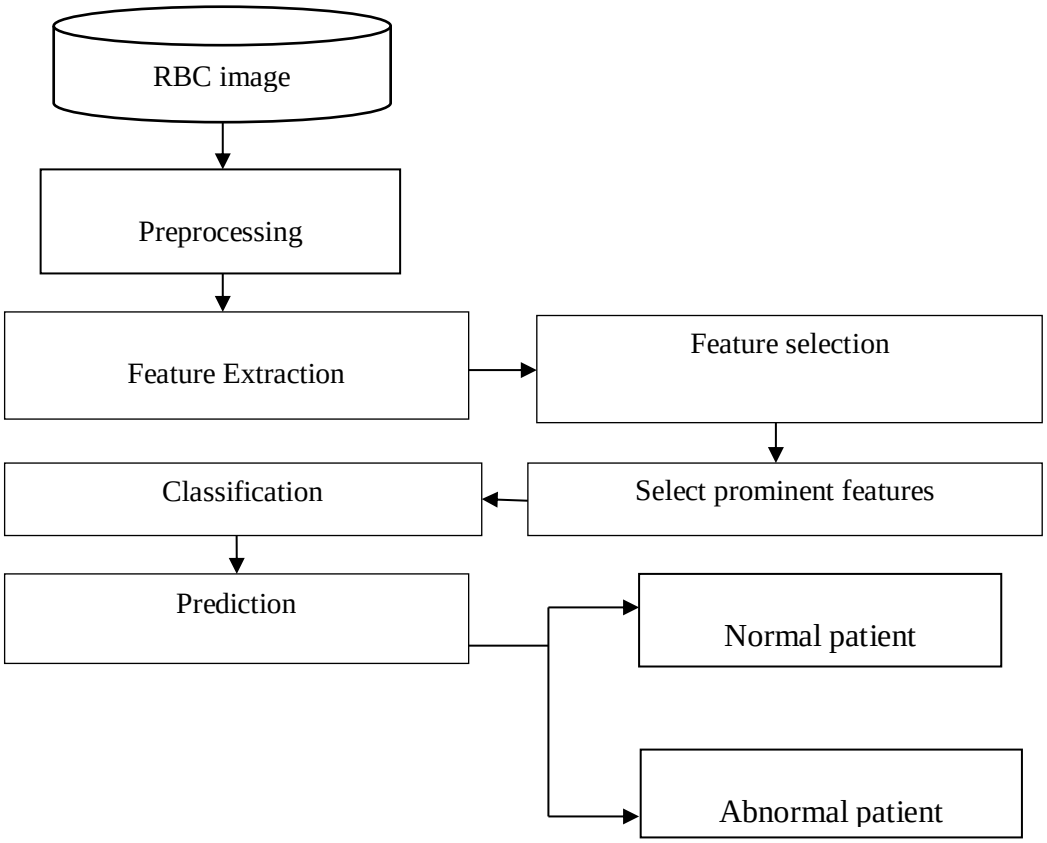

Figure 1: Framework of Proposed Method

Since every node in an MLP is connected to every other layer starting at the current layer, it is a fully connected neural network. Three or more layers make up this structure: input, output, and one or more hidden layers. Each layer provides input to the other layers. All nodes in the layers other than the input layer have non-linear activation functions. Many perceptrons, also known as hidden units, make up each hidden layer. A range of perceptrons make up a single hidden layer. The hidden layer's output is expressed as

$$f(x) = G(W^{T_x} + b) \quad (1)$$

MLP was created to give more than two classes a perceptron structure. We describe a set of learning rules for classifications based on neural networks. The input layer, hidden layer, and output layer are the three layers that make up the MLP architecture shown in Figure 5.6. Neural networks are a useful tool for working with complicated datasets, particularly those related to breast cancer, which have highly non-linear characteristics. This is due to their connected approach, which combines multiple layers. The training technique of these neural networks is backpropagation [11-12].

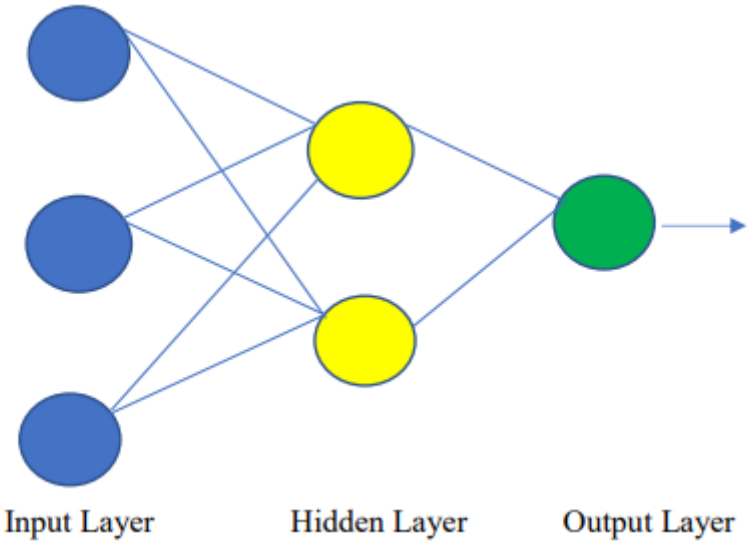

Figure 2: General Architecture of MLP

The MLP's computation phase starts with the hidden layers since the input layer doesn't require any computation; instead, it serves solely as a source of initial values for the hidden layer, which passes the data to the output layer, where the results are shown. Equations 1 and 2 illustrate the hidden layer processing:

$$I_j = \sum_i w_{ji} y_i + a_j \quad (2)$$

$$Y_j = F_i(I_j) \quad (3)$$

w_{ji} is the associated weights between input layer i and hidden layer j. a_j is bias between input layer and hidden layer. Y_i are the inputs to be provided at input layer. Y_j is output at hidden layer generated by the activation function. I_j is the sum of weights and bias. F_i is the activation function. Equation 3 and 4 shows the processing at output layer:

$$I_n = \sum_i w_{nj} y_n + b_n \quad (4)$$

$$Y_n = F_n(I_n) \quad (5)$$

w_{nj} is the associated weights between hidden layer j and output layer n. b_n acts as a bias between output and hidden layer. y_j is output produced by hidden layer. y_n is the result produced at output layer. In is sum of weights for output layer. F_n is activation function. It is to be noted from the above equations that input layer is represented as i, hidden layer is represented as j and output layer is represented as n.

4. Results and discussion

RBC blood cell pictures are used in the experiment to assess the methods' performance. These are separated into individual RBCs, and the number of RBCs found using different methods is compared to the number of RBCs that are truly present in the manually acquired image. Of the eighty-five photos, sixty-two are regarded as training data, and the remaining twenty-three as testing data. the foundational data for the entire dataset that was gathered from the pathologist. Table 1 displays the Accuracy and Processing Time for ANN, KNN, and GA based MLP. Thus, when compared to conventional classifiers, the suggested technique should have a high accuracy and minimal processing time.

Table 1: Accuracy and Processing Time for GA based MLP

Features	Accuracy			Processing time (sec)		
	ANN	KNN	GA based MLP	ANN	KNN	GA based MLP
1	75	78	90	50	35	10
2	79	79	93	46	33	5

3	70	80	95	43	32	2
4	67	82	97	47	31	7
5	83	85	95	40	30	8

The accuracy of the ANN, KNN and GA based MLP can be shown in Figure 3. The proposed GA based MLP has high accuracy when compare with other methods.

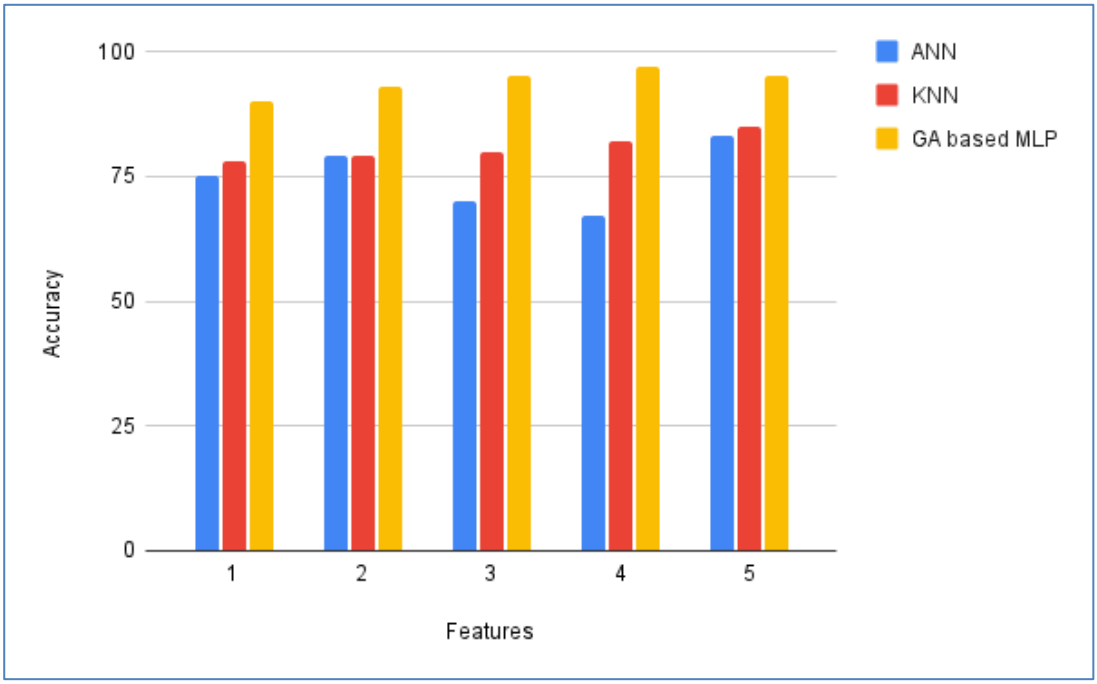

Figure 3 Accuracy for proposed GA based MLP

The Processing Time of the ANN, KNN and GA based MLP can be shown in Figure 3. The proposed GA based MLP has high Processing Time when compare with other methods

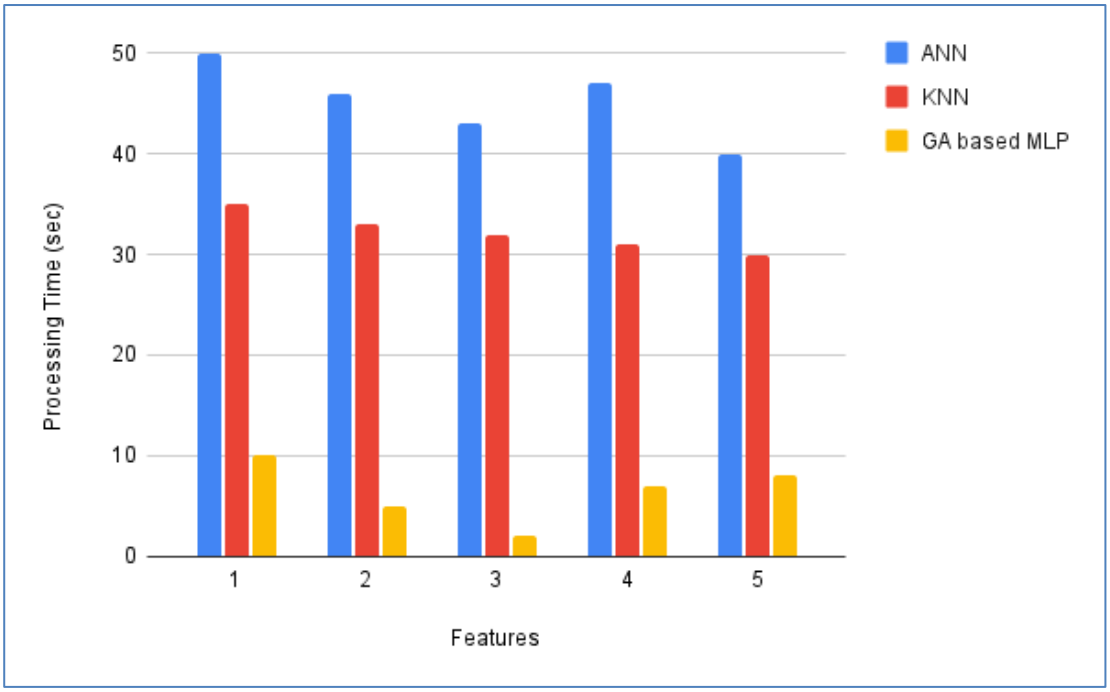

Figure 4 Processing Time for proposed GA based MLP

The test results are compared with the manual knowledge base for every kind of blood cell throughout

the testing phase. There are three feature sets that are used to quantify the classification efficiency: F1, F2, and F3. 62 and 85 sample data are regarded as training data, and the remaining 23 samples are regarded as testing data. The suggested GA-based MLP and MLP for testing efficiency are contrasted. The three feature selections are identified and designated as F1, F2, and F3 in the comparison provided in Table 2.

Table 2: Table 5.3 Testing efficiency Comparison

Features	**Activation function**	**Testing Efficiency**	
		MLP	**GA based MLP**
1	ReLU	87.5	97
2	Sigmoid	89.2	97.50
3	Tanh	90.5	98.15

When comparing the sets F1 and F2, F3 provided the highest efficiency of up to 70%. It is evident from the above table that the suggested GA-based MLP approach outperforms regular MLP in terms of results. As a result, the suggested GA-based MLP's overall testing efficiency produced higher results in F3.

5. Conclusion and future scope

Using an MLP classifier, a medical outcome support system called blood disorder disease diagnostic has been created for the diagnosis of iron deficiency. The system can identify between two types of red blood cells, one of which is malignant. This study investigates the identification of RBC for potential leukaemia diagnosis using GA-based MLP classifiers. In order to improve bone marrow cell discrimination, the proposed method makes use of discriminative shape, colour, and texture traits, which clearly convey information. Furthermore, for efficient classification, feature selection techniques based on GA and MLP are applied. After analysis of the data, 97% of the target was reached

Reference

[1] Kutlu, Hüseyin, Engin Avci, and Fatih Özyurt. "White blood cells detection and classification based on regional convolutional neural networks." *Medical hypotheses* 135 (2020): 109472.

[2] Ratley, Astha, Jasmine Minj, and Pooja Patre. "Leukemia disease detection and classification using machine learning approaches: a review." In *2020 First International Conference on Power, Control and Computing Technologies (ICPC2T)*, pp. 161-165. IEEE, 2020.

[3] Chatterjee, P., Siddiqui, S., Granata, G., Dey, P., & Abdul Kareem, R. S. (2024). Performance Analysis of Five U-Nets on Cervical Cancer Datasets. Indian Journal of Information Sources and Services, 14(1), 17–28.

[4] Abou El-Seoud, Samir, Muaad Siala, and Gerard McKee. "Detection and classification of white blood cells through deep learning techniques." (2020): 94-105.

[5] Bae, D., & Ha, J. (2021). Performance Metric for Differential Deep Learning Analysis. Journal of Internet Services and Information Security, 11(2), 22-33.

[6] Valarmathi, S., and R. Vijayabhanu. "A survey on diabetic retinopathy disease detection and classification using deep learning techniques." In *2021 Seventh international conference on bio signals, images, and instrumentation (ICBSII)*, pp. 1-4. IEEE, 2021.

[7] Praveen, Nalla, Narinder Singh Punn, Sanjay Kumar Sonbhadra, Sonali Agarwal, M. Syafrullah, and Krisna Adiyarta. "White blood cell subtype detection and classification." In *2021 8th international conference on electrical engineering, computer science and informatics (EECSI)*, pp. 203-207. IEEE, 2021.

[8] Sofiene, M., Souhaila, B., & Souhir, C. (2024). Machine Learning for Early Diabetes Detection and Diagnosis. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), 15(1), 216-230.

[9] Nithyaa, A. N., R. Premkumar, D. Kanchana, and Nalini A. Krishnan. "Automated detection and classification of blood diseases." In *Recent Advancements in System Modelling Applications: Proceedings of National Systems Conference 2012*, pp. 393-404. Springer India, 2013.

[10] Akalin, Fatma, and NEJAT YUMUŞAK. "Detection and classification of white blood cells with an improved deep learning-based approach." *Turkish Journal of Electrical Engineering and Computer Sciences* 30, no. 7 (2022): 2725-2739.

[11] Arora, G. (2024). Desing of VLSI Architecture for a flexible testbed of Artificial Neural Network for training and testing on FPGA. Journal of VLSI Circuits and Systems, 6(1), 30-35.

[12] Khan, Muhammad Umar, Sumair Aziz, Khushbakht Iqtidar, Galila Faisal Zaher, Shareefa Alghamdi, and Munazza Gull. "A two-stage classification model integrating feature fusion for coronary artery disease detection and classification." *Multimedia Tools and Applications* (2022): 1-30.

[13] Saravanan T., et.al Exploring domain auction and priority checking analysis, Middle - East Journal of Scientific Research, V-20, I-12, PP:2549-2553, 2014.

[14] Singh, Ishpreet, Narinder Pal Singh, Harnoor Singh, Saharsh Bawankar, and Alioune Ngom. "Blood cell types classification using CNN." In *Bioinformatics and Biomedical Engineering: 8th International Work-Conference, IWBBIO 2020, Granada, Spain, May 6–8, 2020, Proceedings 8*, pp. 727-738. Springer International Publishing, 2020.

[15] Varghese, Nisha. "Machine learning techniques for the classification of blood cells and prediction of diseases." *International Journal of Computer Science Engineering* 9, no. 1 (2020): 66-75.

[16] Ghaderzadeh, Mustafa, Farkhondeh Asadi, Azamossadat Hosseini, Davood Bashash, Hassan Abolghasemi, and Arash Roshanpour. "Machine learning in detection and classification of leukemia using smear blood images: a systematic review." *Scientific Programming* 2021, no. 1 (2021): 9933481.

[17] Bagasjvara, R. G., Ika Candradewi, Sri Hartati, and Agus Harjoko. "Automated detection and classification techniques of Acute leukemia using image processing: A review." In *2016 2nd International Conference on Science and Technology-Computer (ICST)*, pp. 35-43. IEEE, 2016.

[18] Bates, Imelda, and Barbara J. Bain. "Approach to the diagnosis and classification of blood diseases." *Dacie and Lewis Practical Haematology* (2012): 549.

[19] Poonia, Ramesh Chandra, Mukesh Kumar Gupta, Ibrahim Abunadi, Amani Abdulrahman Albraikan, Fahd N. Al-Wesabi, and Manar Ahmed Hamza. "Intelligent diagnostic prediction and classification models for detection of kidney disease." In *Healthcare*, vol. 10, no. 2, p. 371. MDPI, 2022.

[20] Yildirim, Muhammed, and Ahmet Çinar. "Classification of white blood cells by deep learning methods for diagnosing disease." *Rev. d'Intelligence Artif.* 33, no. 5 (2019): 335-340.

[21] Kutlu, Y., & Camgözlü, Y. (2021). Detection of coronavirus disease (COVID-19) from X-ray images using deep convolutional neural networks. Natural and Engineering Sciences, 6(1), 60-74.

[22] Jelena, T., & Srđan, K. (2023). Smart Mining: Joint Model for Parametrization of Coal Excavation Process Based on Artificial Neural Networks. Archives for Technical Sciences, 2(29), 11-22