Health Care Surveillance Using Machine Learning and Data Analytics

Ashu Nayak[1], Kapesh Subhash Raghatate[2]

[1]*Assistant Professor, Department of CS & IT, Kalinga University, Raipur, India*
[2]*Research Scholar, Department of CS & IT, Kalinga University, Raipur, India.*

KEYWORDS

Health, Machine Learning, Data analytics, classification, RBFN.

ABSTRACT

Any country's healthy population are a true asset. Both developed and developing nations are spending enormous sums of money to strengthen their "healthcare systems" and the required "health infrastructure." Globally, only few countries have a proactive approach to healthcare. The pandemic that has been going on recently has taught countries hard lessons about how important it is to have strong healthcare systems. The formulation of effective health policies and initiatives depends heavily on studies on Public Health Surveillance (PHS) and the conclusions that follow. This task has become clearer with the introduction of modern computing techniques and technology. Humanity has always been saved by technology when it is applied correctly. In this context, the most promising machine learning techniques are applied in this work. An empirical investigation of RBFN classifiers is presented in this paper. A variety of performance criteria, including recall, f-score, accuracy, precision, and False Positive Rate (FPR), are used to evaluate the efficacy of the recommended procedures. The RBFN approach has the highest accuracy and the least amount of time complexity in identifying health diseases.

1. Introduction

Since 2019, the pandemic has produced a new normal where the age-old lesson "Health is Wealth" may be the ideal one to reconsider. Every study that addresses such a unique health scenario is important to researchers and vital to the welfare of humanity. A society's citizens, and consequently a region or country, find infectious and contagious diseases to be extremely concerning. A pandemic poses a far greater threat, with the majority of nations potentially affected. Pandemic illnesses are extremely infectious and contagious. The majority of the time, infection transmission and retransmission cause disruptions to daily living [1]. This transmission must occur between individuals, or between individuals and a cohort or cluster. Based on historical information, it is suggested that this pandemic occurs once every 100 years on average. Every time a pandemic strike, the effects on human evolution and history are unfathomable. In summary, pandemics are a signpost for humanity's quest for growth. The main obstacle to infection identification is that those in the asymptotic category (infected without symptoms) can remain infected for several days, making it extremely tough to track the spread of the virus in this category [2].

Emergencies in the health and economy happen at the same time. Millions of people contract the infection, become ill, and occasionally develop fatal illnesses. The society's health will be in unanticipated peril. After a while, one's physical and mental well-being are called into doubt. It is necessary to address the state of economic instability. Costs both direct and indirect go up, which raises living expenses. Handling the impact on social, health, and economic issues is nearly impossible [10]. There are limitations on travel and unexpected market closures, which could cause unanticipated disruptions in people's lives and social trauma that needs to be addressed collectively [3]. One most common scenario is high morbidity and mortality; this needs to be treated cautiously. Naturally, one more risk that arises during a pandemic is security on a regional, national, and international scale [7]. In recent times, machine learning techniques have been employed in the design and development of early warning systems. Machine learning techniques are utilised in this work to complete the goal of anticipating a pandemic. In this instance, section 1 of the paper examines the introduction, whereas section 2 examines the relevant literature. Section 3 provides an explanation of the planned work, Section 4 presents the work's outcomes, and Section 5 concludes the project.

2. Literature Review

Modern research relies heavily on machine learning (ML) to solve a wide range of challenging real-

world challenges [18]. The time series is a set of measurements taken at regular intervals. These intervals can be one year, for example, annual expenses, one month, for example, passenger traffic, one week, for example, sales, one day, for example, climate, one hour, for example, share trading, one minute, for example, call records, or one second, for example, web traffic. Data on infections and deaths by day has been available since January 30, 2021, and is still available today [5]. An epidemic model was created to determine the spread of infection from person to person using the classifications of susceptible (S), infected (I), and removed (R) individuals [12]. Later, an additional phase called exposed (E) was included to the model [6]. The result of a time series that projects future values based on a date interval is called forecasting [20]. As forecasting aids in medical management, it is imperative to predict infections and fatalities on a daily basis [17]. Following March 2020, a number of actions were done to contain and get rid of the infection. The actions include things like closing educational facilities, providing entertainment, transportation, preserving social distance, limiting the number of guests at community events, stopping public gatherings and religious activities, etc [15]. Many businesses were unable to satisfy their daily requirements and some were forced to close as a result of the closure [14]. The main goal at first is to stop the disease from spreading; socioeconomic aspects specific to each person are disregarded, and there is insufficient data to analyse and forecast the future [13]. Given the abundance of reliable data sources accessible at the moment, it is imperative to estimate the disease's spread in order to plan the population's economic condition. For the accompanying issues in the economy, health, and public life to be properly mitigated, accurate estimation and prognosis of contagious diseases is crucial. Only with the right methods and equipment are the successful suggestions achievable [8]. In order to provide a viable early warning system for infectious diseases using a machine learning approach, this research challenge was identified and formulated [4]. As a result, this study was extended to a typical pandemic, namely Covid 19 [19].

3. Methodology

This work's technique is grounded in empirical evidence and scientific data. The most promising and tested scientific techniques and algorithms are used in the analysis and interpretation of the experimental data. A thorough review of the literature has been done in order to comprehend the characteristics and nature of infectious diseases like influenza, SARS, MERS, and COVID-19. Based on the scientific data that is currently available, the models that are most effective in mitigating the challenges posed by these diseases have been investigated. Supervised, semi-supervised, and unsupervised training can be generically categorised in machine learning networks. The schematic of the thesis framework is shown in Figure 1. This algorithm additionally forecasts the spread of pandemics. This work encompasses multiple elements, including as changes in the environment, seasonal variations, and communicable diseases.

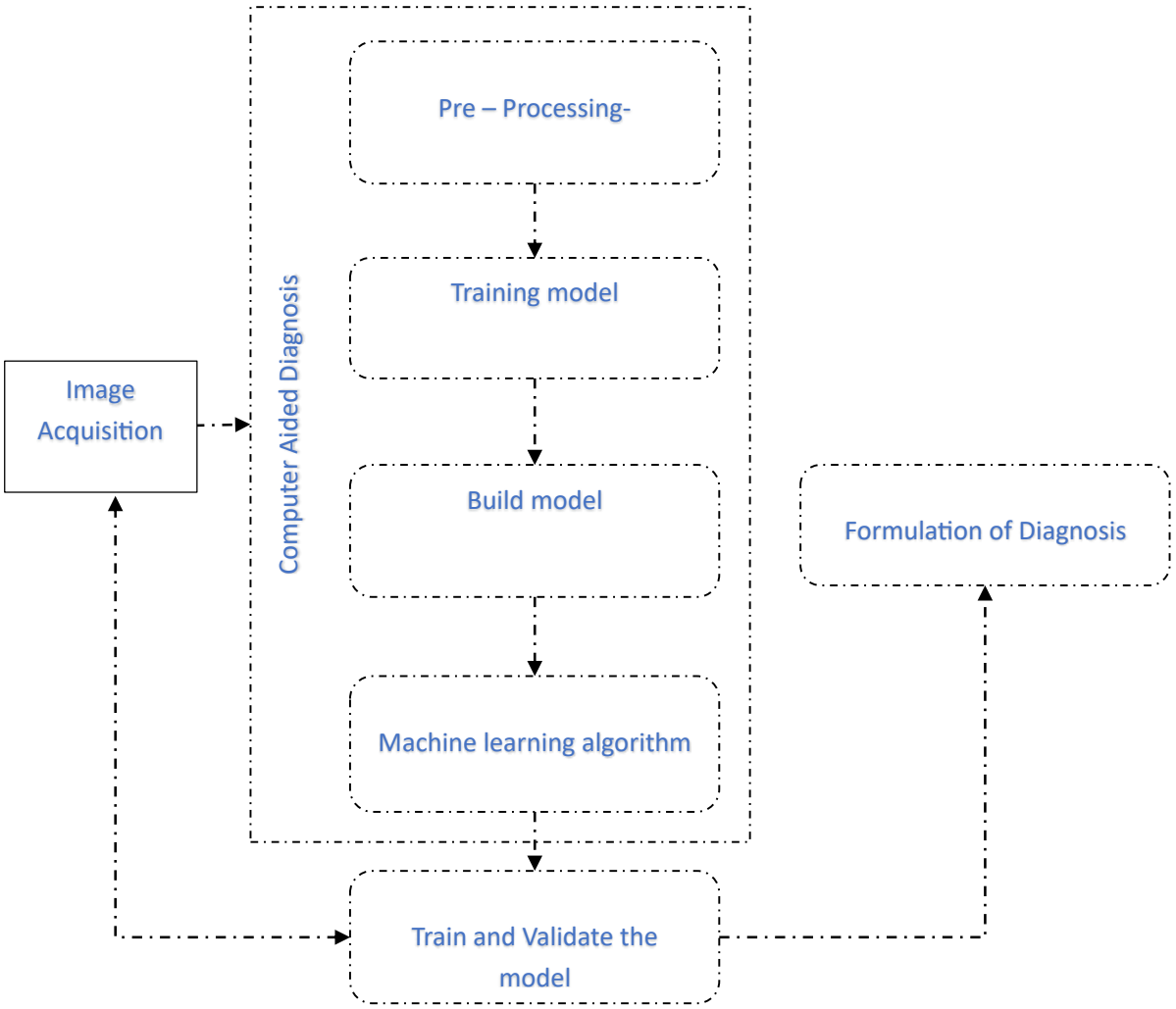

Figure 1. Schematic Diagram of proposed framework

Image acquisition

The target region's population density serves as the first input source, and data can be obtained online via India's NIC portal. The data collection offers all of the information based on the COVID-19 open repository dataset [9].

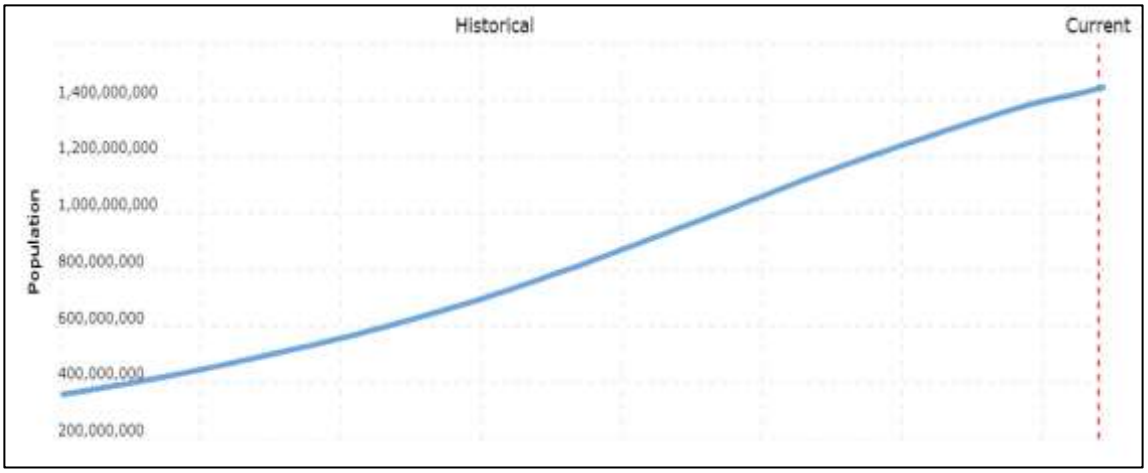

Figure 2. Projected Population Density in India Based on 2024- Census

It is clear from the aforementioned inputs that the annual rate of dengue cases varies, and that this variation is solely attributed to the factors of population, rainfall, and humidity.

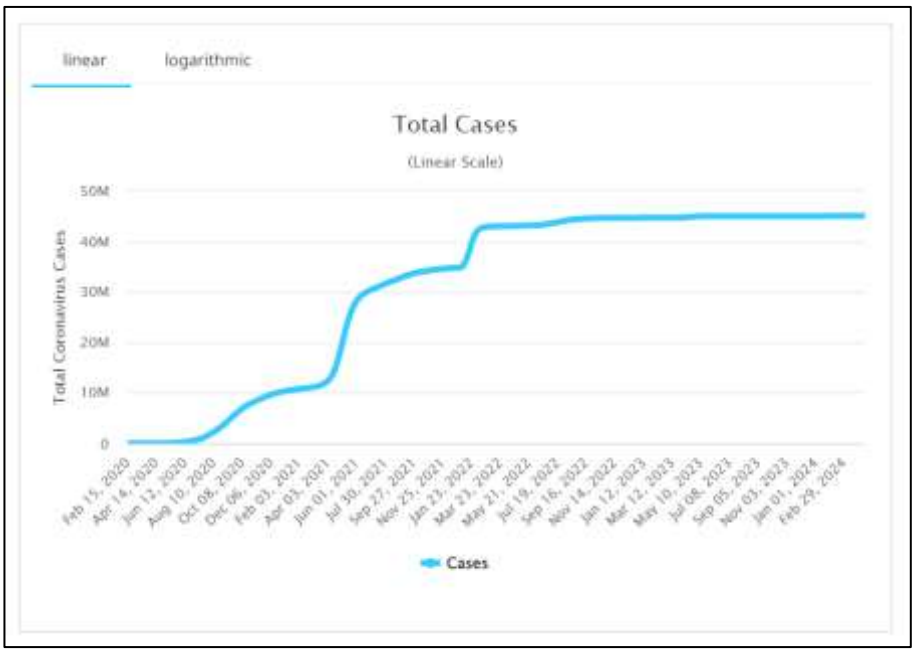

Figure 3. The Covid cases in India from the period from 2020 to 2024 across India

Preprocessing

The term "data pre-processing mechanism" refers to the data modification that the algorithm accepts. A few procedures need to be used in order to convert some of the category features in the incoming data set to numerical form. The areas impacted by COVID-19 are divided into two categories: those where the disease does not occur and those where it does occur over time.

Feature selection

One of the main activities in data preprocessing is feature selection. Temperature, wind speed, humidity, dew, and population are the features taken into account in the suggested work in order to determine the impact of the virus. The Random Forest algorithms [16] are utilised to determine the comparative significance of the attributes.

Classification: Radial Basis Function Network

An N-dimensional space can be locally represented by radial-based function networks. The control zone, which is constrained by baseline functions, carries it out [11]. This baseline function's

requirements are computed by

$$\varphi_j(x) = \exp\left(\frac{\|x-\mu_j\|^2}{2\sigma^2{}_j}\right) \qquad (1)$$

Where μ_j reference vector and σ_j $is\ the$ circumstances of the influence field

Each RBFN unit that can be expressed mathematically as a function of a radial basis

$$\varphi_j(x) = \varphi(\|x - x_j\|) \qquad j = 1,2,\dots N \qquad (2)$$

Where N represents the dimension of the preparation model and $(\|x - x_j\|)$is the Euclidean norm of the vector $(x - x_j)$. The J[th] input data point x_j determines the RBF center, and the pattern vector x is added to the input layer. Gaussian function is used in the hidden layer of the network as the radial basis function in which each computing unit is located.

$$\varphi_j(x) = \varphi(x - x_j) = \exp\left(-\frac{1}{2\sigma_j^2}\|x - x_j\|^2\right) \qquad j = 1,2,\dots N \qquad (3)$$

Where, j is a measure of the width of the Gaussian j[th] function with $x_j$center. All the Gaussian hidden units are usually, but not always, allocated a specific width.

The RBF network structure's mathematical formation has the following mathematical form:

$$F(x) = \sum_{j=1}^{k} w_j\, \varphi(x, x_j) \qquad (4)$$

Where the input vector x is vector dimensional and every hidden unit is defined by the radial base function (x, x_j), where j=1, 2... K. The output function, which is expected to consist of a single element, is defined by the vector w of weight, whose dimensionality is also K.

4. Results and discussion

The goal of showcasing the real-time classification experiment was to identify Covid 19. An individual Windows 7 Virtual Machine (VM) was employed to execute the recommended RBFN. The training and testing sets' accuracy increases with epochs in our model. As the epoch lengthens, the loss of the training and testing sets diminishes. With an increasing number of training rounds, accuracy increases and loss decreases, but eventually it appears to be flat. We experimented every ten epochs, ranging from 10 to 50, to find an optimal epoch value.

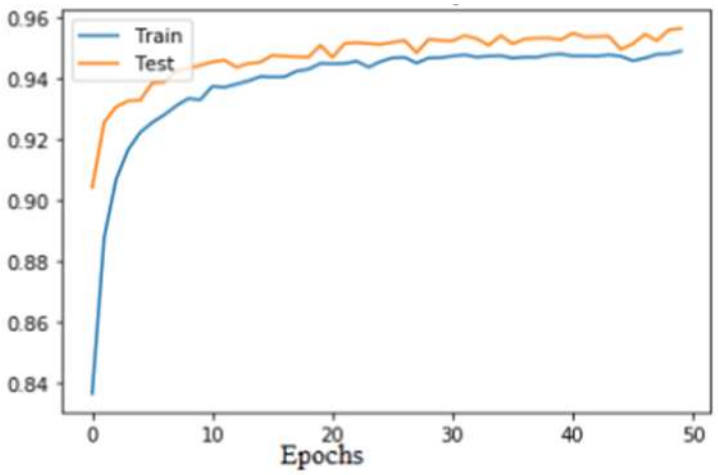

Figure 4. Accuracy of covid detection model

Table 4 shows the performance of the proposed model on the test dataset along with the accuracy, precision, recall, F-Score, false alarm rate, and misclassification rate for each class.

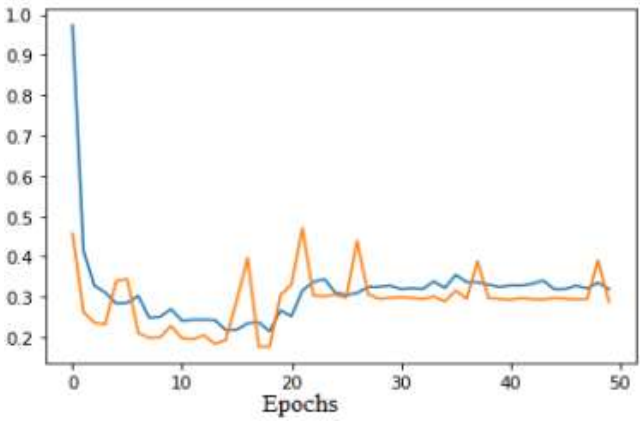

Figure 5. Loss of covid proposed model

Table 1. Performance metrics of disease prediction

Models	Accuracy (%)	False alarm rate	Misclassification rate	F1-Score	Precision (%)	Recall (%)
Normal	97.5	2.0	2.5	0.97	97	89.85
COVID 19	99.2	0.7	0.8	0.985	97.50	90.25
SARS	96.5	0.95	3.5	0.97	98.15	81.05
MERS	97	1.2	3	0.985	97.50	90.25
IN-FLU	98.5	1.0	1.5	0.97	98.15	81.05

The accuracy of our RBFN model was compared with both the newest machine learning model algorithms and the most established machine learning models. The experiment demonstrated this with excellent accuracy on the NIC dataset. Table 2 presents the suggested model alongside the traditional model.

Table 2. Performance metrics with conventional models

Models	Accuracy (%)	Time complexity
Conventional Algorithm		
SVM	90	5 min
MLP	90.7	2 min
XG-Boost	92	2 min
Ensemble method	93	1.9 min
Proposed- RBFN		
Normal	97.5	10 sec
COVID 19	99.2	5 sec
SARS	96.5	7 sec
MERS	97	2 sec

We conducted an experiment to demonstrate the feasibility and potential for effectively using machine learning capabilities for disease prediction while employing RBFN for health care disease surveillance through data analytics and machine learning.

5. Conclusion and future scope

The COVID-19 pandemic began in December 2019 and quickly expanded to every country in the world by 2020. Many people have experienced emotional breakdown, anxiety, stress, despair, and sleep difficulties as a result of this pandemic. A few situations that are affecting people's mental health are job insecurity, fear of losing one's job, inevitable lockdowns, and limitations on travelling outside. Individuals who have previously had mental health issues are more susceptible to these effects. Psychologists are expected to go above and above in their efforts to assist those who are marginalised and alone. This paper focuses on using basic RBFN on a population dataset to categorise people's opinions about the coronavirus as neutral, negative, or positive. It also compares the effects of accuracy. Compared to other models, the machine learning model provides a higher accuracy. To further illustrate the value of this technology, a list of helpful details about COVID mental health care that could be extracted from the cloud is provided.

Reference

[1] Hady, Anar A., Ali Ghubaish, Tara Salman, Devrim Unal, and Raj Jain. "Intrusion detection system for healthcare systems using medical and network data: A comparison study." *IEEE Access* 8 (2020): 106576-106584.

[2] Balyan, Amit Kumar, Sachin Ahuja, Sanjeev Kumar Sharma, and Umesh Kumar Lilhore. "Machine learning-based intrusion detection system for healthcare data." In *2022 IEEE VLSI Device Circuit and System (VLSI DCS)*, pp. 290-294. IEEE, 2022.

[3] Pande, Sagar, Aditya Khamparia, and Deepak Gupta. "An intrusion detection system for health-care system using machine and deep learning." *World Journal of Engineering* 19, no. 2 (2022): 166-174.

[4] Srinivasa Rao, M., Praveen Kumar, S., & Srinivasa Rao, K. (2023). Classification of Medical Plants Based on Hybridization of Machine Learning Algorithms. Indian Journal of Information Sources and Services, 13(2), 14–21.

[5] Thamilarasu, Geethapriya, Adedayo Odesile, and Andrew Hoang. "An intrusion detection system for internet of medical things." *IEEE Access* 8 (2020): 181560-181576.

[6] Öztürk, Tolgahan, Zeynep Turgut, Gökçe Akgün, and Cemal Köse. "Machine learning-based intrusion detection for SCADA systems in healthcare." *Network Modeling Analysis in Health Informatics and Bioinformatics* 11, no. 1 (2022): 47.

[7] Choudhary, A., Choudhary, G., Pareek, K., Kunndra, C., Luthra, J., & Dragoni, N. (2022). Emerging Cyber Security Challenges after COVID Pandemic: A Survey. Journal of Internet Services and Information Security, 12(2), 21-50.

[8] Saif, Sohail, Priya Das, Suparna Biswas, Manju Khari, and Vimal Shanmuganathan. "HIIDS: Hybrid intelligent intrusion detection system empowered with machine learning and metaheuristic algorithms for application in IoT based healthcare." *Microprocessors and Microsystems* (2022): 104622.

[9] Ashraf, Eman, Nihal FF Areed, Hanaa Salem, Ehab H. Abdelhay, and Ahmed Farouk. "Fidchain: Federated intrusion detection system for blockchain-enabled iot healthcare applications." In *Healthcare*, vol. 10, no. 6, p. 1110. MDPI, 2022.

[10] Malathi, K., Shruthi, S.N., Madhumitha, N., Sreelakshmi, S., Sathya, U., & Sangeetha, P.M. (2024). Medical Data Integration and Interoperability through Remote Monitoring of Healthcare Devices. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), 15(2), 60-72. https://doi.org/10.58346/JOWUA.2024.I2.005

[11] Jain, Anshul, Tanya Singh, and Satyendra Kumar Sharma. "Security as a solution: an intrusion detection system using a neural network for IoT enabled healthcare ecosystem." *Interdisciplinary Journal of Information, Knowledge, and Management* 16 (2021): 331-369.

[12] Juma, J., Mdodo, R.M., & Gichoya, D. (2023). Multiplier Design using Machine Learning Alogorithms for Energy Efficiency. Journal of VLSI Circuits and Systems, 5(1), 28-34.

[13] Subasi, Abdulhamit, Shahad Algebsani, Wafa Alghamdi, Emir Kremic, Jawaher Almaasrani, and Najwan Abdulaziz. "Intrusion detection in smart healthcare using bagging ensemble classifier." In *International Conference on Medical and Biological Engineering*, pp. 164-171. Cham: Springer International Publishing, 2021.

[14] Elankavi R., et.al Assessing Learning Behaviors Using Gaussian Hybrid Fuzzy Clustering (GHFC) in Special

Education Classrooms, Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, V-14, I-1, PP:118-125, 2023.

[15] Lee, Jae Dong, Hyo Soung Cha, Shailendra Rathore, and Jong Hyuk Park. "M-IDM: A Multi-Classification Based Intrusion Detection Model in Healthcare IoT." *Computers, Materials & Continua* 67, no. 2 (2021).

[16] Savanović, Nikola, Ana Toskovic, Aleksandar Petrovic, Miodrag Zivkovic, Robertas Damaševičius, Luka Jovanovic, Nebojsa Bacanin, and Bosko Nikolic. "Intrusion detection in healthcare 4.0 internet of things systems via metaheuristics optimized machine learning." *Sustainability* 15, no. 16 (2023): 12563.

[17] Sundararajan, T. V. P., and A. Shanmugam. "A novel intrusion detection system for wireless body area network in health care monitoring." *Journal of Computer Science* 6, no. 11 (2010): 1355.

[18] Akram, Faiza, Dongsheng Liu, Peibiao Zhao, Natalia Kryvinska, Sidra Abbas, and Muhammad Rizwan. "Trustworthy intrusion detection in e-healthcare systems." *Frontiers in public health* 9 (2021): 788347.

[19] Akbulut, G., & Yalniz, Ş. Ç. (2022). Impact of Covid-19 Pandemic on Public Aquariums in Turkey. Natural and Engineering Sciences, 7(3), 260-270.

[20] Vranješ, B., Vajkić, M., Figun, L., Adamović, D., & Jovanović, E. (2024). Analysis of Occupational Injuries in an Iron Ore Mine in Bosnia and Herzegovina in the Period from 2002 to 2021. Archives for Technical Sciences, 1(30), 33-44.