

MACHINE LEARNING-BASED STROKE DISEASE DIAGNOSIS USING ELECTROENCEPHALOGRAM (EEG) SIGNALS

AKTHAM F. SAWAN¹, MOHAMMED AWAD^{2,*},
RADWAN QASRAWI^{1,3}, MOHAMMAD SOWAN⁴

¹Department of Computer Science, Al Quds University,
Jerusalem, Palestine

²Department of Computer Systems Engineering, Arab American
University, Jenin, Palestine

³Department of Computer Engineering, Istinye University,
34010, Istanbul, Turkey

⁴Department of Emergency Medicine, Al Bashir Hospital,
Amman, Jordan

*Corresponding Author: Mohammed.awad@aaup.edu

Abstract

Stroke is currently ranked as the third leading cause of death worldwide. While computed tomography (CT) and magnetic resonance imaging (MRI) are commonly used for stroke diagnosis, they have their limitations. CT scans can be time-consuming, taking up to 8 hours to complete diagnosis, while MRI procedures can be lengthy, often making it impractical for most stroke patients. This has led to the necessity of exploring new methods for stroke detection, particularly utilizing EEG signals. In this paper, we propose a cloud computing-based machine learning (ML) system that leverages MUSE2 to diagnose stroke patients by analysing EEG signals. Our dataset, collected from Al Bashir Hospital between 2021 and 2022, consists of a randomly selected sample of 31 stroke patients and 31 healthy individuals. To pre-process the collected dataset, we employ Fourier and wavelet transformations. The processed EEG signals are then transmitted over the Internet to the ML model for stroke diagnosis. Real-time results are delivered to authorized personnel via SMS. During our research, various classifiers were evaluated, and a modified XGboost classifier emerged as the most effective choice. It outperformed other ML classifiers with an impressive accuracy of 96.87%.

Keywords: Cloud, EEG, Machine learning, MUSE2, Stroke, Wearable devices.

1. Introduction

Every year, approximately sixteen million individuals worldwide suffer from strokes, with six million of them losing their lives and another five million left permanently disabled [1]. Improving the classification of strokes during their early stages would significantly enhance patient outcomes and their overall quality of life, even in cases where current treatments cannot fully prevent the stroke's consequences. Swift recognition and treatment of ischemic stroke offer a higher chance of survival and complete recovery. Therefore, identifying the condition before reaching the hospital, whether at the patient's residence or in transit via ambulance, can be the determining factor between life and permanent disability [2].

Currently, computed tomography (CT) and magnetic resonance imaging (MRI) are employed to detect haemorrhagic or ischemic strokes due to their ability to provide detailed anatomical and pathological information about the brain. CT scans have proven reliable in diagnosing strokes within 6-8 hours. However, MRI imaging offers higher precision and the potential to detect a stroke within just 30 minutes. Unfortunately, MRI is not suitable for all patients and is typically available only in major medical centres [3-5]. It is crucial to accurately diagnose the type of stroke (brain infarction, cerebral haemorrhage) and determine the extent of brain damage within three hours of the stroke's onset [6].

Clinical studies have demonstrated that the rate of stroke recurrence within a year ranges from 10 to 15 percent [7], although this varies depending on the type of stroke and individual risk factors. Therefore, it is crucial to develop tools for early prediction in stroke patients and those with a history of stroke. Over the past decade, non-invasive structural imaging techniques such as EEG have transitioned from research and medicine to the commercial sector. By interpreting and correlating the electrical activity recorded by EEG with corresponding mental processes, researchers can gain valuable insights into how the brain functions.

In the realm of medical studies, clinical research stands out as the field that extensively utilizes clinical neuroimaging instruments, which demand high precision and advanced capabilities [8]. Among these instruments is MUSE2, a portable device capable of monitoring EEG waves at a sampling frequency of 256 [9]. Consequently, the primary objective of this study is to develop a faster and more user-friendly solution that can be implemented in real-time, either by the patient at home or during transport in an ambulance en route to the hospital. The classification of strokes using wearable EEG sensors, coupled with the application of ML algorithms, requires further investigation. This approach is expected to yield more accurate results compared to previous methods.

Businesses and various sectors can leverage the outcomes of predictive data processing facilitated by ML techniques [10]. The abundance of data available today has paved the way for the establishment of data and market structures, as well as statistical models that aid in making informed decisions [10]. Algorithms based on ML techniques are then applied to these extensive datasets to determine which model best captures the relationship between the provided descriptive features and the target feature [10].

In recent years, advancements in technology have introduced innovative methodologies for understanding the functioning of the central nervous system, including Electroencephalography (EEG). This technology enables the collection

of vital brain signals and has proven to be an invaluable instrument for studying brain activity. By serving as a brain-machine learning tool, EEG allows for the extraction of valuable data on the internal structural dynamics of the brain. The output waveform obtained represents the underlying behavior of brain regions beneath the cortex, with symmetrical representation between the right and left hemispheres [11].

Typically, EEG data is represented in terms of rhythmic patterns, classified into four recurring categories: alpha, theta, delta, and beta. The alpha wave, associated with relaxation and tranquillity, was the primary focus of this study [12]. Recent studies have demonstrated the utility of raw EEG data alone in the real-time diagnosis of seizure patients [13]. To facilitate this process, a cloud computing-based system has been developed, allowing the patient's EEG signal to be captured using the portable MUSE2 device and transmitted over the Internet for prompt diagnosis. To expedite the process and enhance accessibility, the patient's data is processed through a prebuilt model, and the results are promptly sent as text messages to authorise individuals within minutes.

In this research, the XGBoost classifier was modified as detailed in paper [14] to aid in the stroke classification process. Since EEG signals were the primary source of data for stroke diagnosis, the XGBoost classifier needed to be customized to better understand and interpret them.

To improve its performance and efficacy in analysing EEG signals, the XGBoost classifier was modified in a number of key ways. These adjustments probably included adjusting the parameters of the algorithm and incorporating feature engineering methods tailored to the analysis of EEG data [14].

The study aimed to enhance the precision and consistency of EEG-based stroke classification. To accomplish this, we used the modified version of the XGBoost classifier that takes into account the specific features and patterns found in the EEG data. Because of this, stroke cases could be diagnosed with greater accuracy.

The dataset obtained from Al Bashir Hospital was used with a variety of classifiers, including Modified XGBoost, Naive Bayes (NB), , Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), Decision Tree Bayes (DT), Random Forest (RF), and Logistic Regression (LR).

Overall, the utilization of the modified XGboost classifier in this research demonstrates the researchers' efforts to tailor ML algorithms specifically for EEG-based stroke classification, potentially leading to more accurate and efficient diagnostic outcomes.

This paper makes several contributions. Firstly, it develops a cloud computing-based system that allows stroke diagnosis in a few minutes using the stroke dataset collected from Al Bashir Hospital in Jordan within two years. Secondly, it utilizes a modified XGboost classifier for stroke classification. Lastly, it designs the framework of a decision support system in the cloud computing environment, providing the ability to diagnose strokes quickly and efficiently from any location at any time.

The remainder of this paper is organized as follows: Section 2 presents a comprehensive literature review on EEG and ML for stroke classification. Section 3 delves into the cloud computing setting and describes the proposed EEG-based

stroke diagnosis system based on ML. Section 4 provides a detailed discussion of the experiments conducted, while Section 5 concludes the paper and provides recommendations for future studies.

2. Related Works

Recent studies have expanded the scope of EEG. The rapid detection of seizures is addressed in paper [14], where the authors propose a Modified XGboost Classifier model. This model is specifically designed to expedite the process of identifying seizures through a classification approach. The model incorporates a focal loss function to minimize training and testing inaccuracies, thus improving the accuracy of epilepsy predictions. The evaluation involves using sample data from the CHB-MIT SCALP Electroencephalography (EEG) dataset, obtained from multiple patients. The goal is to achieve effective differentiation between seizures and non-seizures by employing an efficient classification process. The obtained performance results are compared to various state-of-the-art techniques, focusing on average sensitivity and average specificity. The performance evaluation includes metrics such as classification accuracy, sensitivity, and specificity for each patient.

Lee et al. [15] and Yang et al. [16] studied epilepsy, revealing abnormal brain waves in patients with various conditions such as stroke, schizophrenia, and depression. In the identification of brain dysfunction, absolute power value (based on relative power value) and the frequency domain have proven to be more reliable indicators than the raw EEG spectrum [2]. Absolute power value measures the ratio of total amplitude across the frequency band, providing a measure that is independent of electrical resistance, skull thickness, or other non-brain wave electrical activity. Interpreting EEG signals requires expertise in noise filtering, psychoanalysis, neuroscience, and understanding the distinction between absolute and relative power [17].

The MUSE2 portable device has been extensively studied for EEG data processing in stroke detection. For instance, Wilkinson et al. [18] demonstrated the feasibility of pre-hospital stroke diagnosis using a portable EEG unit. They recorded brainwaves from 25 participants, including 16 with acute ischemic stroke, and correlated the results with accuracy controls that included stroke mimics. EEG, a physiological signal used for detecting and analysing brain waves, has been employed in studies investigating stroke patients [19-23]. Stroke is characterized by abnormal and slow signals in the delta wave and decreased normal and rapid activity in the alpha wave, as confirmed by Choi et al. [21].

The results of these tests demonstrate the efficacy of using ratios of delta power to alpha power and theta wave power to beta wave power, as well as the combining of these two parameters, for the detection and prediction of stroke. Yu et al. [22] used EEG frequency analysis and topographic maps to show that 27 out of 30 patients with a moderate stroke had an increase in big delta waves and a decrease in alpha wave activity. Varelas, and Hacein-Bey [23] verified that a patient with epilepsy had a high-amplitude theta wave as opposed to a delta wave following a stroke. The theta wave, beta, alpha, and high gamma increased rapidly in the right hemisphere, as reported by Ip et al. [20], who also confirmed that the activity and stability of theta and delta waves in stroke patients were affected by measurements of brain waves in the cerebral cortex.

Gottlibe et al. [24] explored the classification of patients into "control" and "stroke" groups using a single, brief EEG recording. They collected data from individuals who had recently experienced an ischemic stroke, while volunteers with intermediate scores were included as the control group. By evaluating the spectral energy similarity between the two brain hemispheres using the pdBSI (Updated Brain Symmetry Index), they assessed the feasibility of classifying patients. In another study by Djamal et al. [25], data from 25 individuals, including 16 with severe ischemic stroke, were recorded using the portable EEG device MUSE2. Significant improvement ($p < 0.01$) was observed in patients with major ischemic strokes. The authors employed MUSE2 and an enhanced ML method to analyse EEG data and achieve accurate stroke detection.

In their study, Sawan et al. [26] utilized the same dataset as described in [25]. However, they employed different machine learning classifiers and methodologies. Specifically, they explored eight machine learning techniques for their analysis, and the XGBoost classifier demonstrated superior performance compared to other classifiers. It achieved an impressive accuracy rate of 83.89% in identifying strokes. These findings showcased a significant improvement of 7.89% in accuracy compared to the previous study [25].

In summary, the key highlight of this research is the ability to obtain stroke classification results in real-time. However, there are limitations associated with the use of deep learning algorithms, as they can be time-consuming. On the other hand, traditional machine learning algorithms are not specifically designed for the classification of EEG signals, resulting in less accurate outcomes. Nevertheless, the reviewed related studies have confirmed the effectiveness of the modified XGboost classifier developed in [14] for EEG signal classification. It has demonstrated promising results and is considered suitable for EEG data analysis.

By employing the modified XGboost classifier, we successfully addressed the limitations associated with both deep learning and traditional machine learning algorithms. Furthermore, according to the reviewed literature, EEG classification has primarily been utilized in hospital-based applications, requiring patients to be physically present before analysis can be conducted. However, our research leveraged cloud computing technology in conjunction with the modified XGboost classifier, enabling the transmission of EEG classification results from any location with internet coverage. This innovative approach yielded highly accurate results within a matter of minutes, a novel contribution that, to the best of our knowledge, has not been explored within this domain before.

3. Background

3.1. MUSE2 EEG

The MUSE2 EEG device is equipped with electrodes and a pressure sensor to record brain waves. Electrodes are placed on specific locations on the scalp, including TP9, TP10, AF8, and AF7 as can be seen in Fig. 1, which are identified using the internationally recognized 10-20 system. These electrodes capture the electrical activity of the brain, and the MUSE2 device measures this activity at a sampling rate of 256 Hz. EEG data can be collected and wirelessly transmitted via Wi-Fi for real-time analysis when used in conjunction with the MUSE2 display program on mobile devices [9].

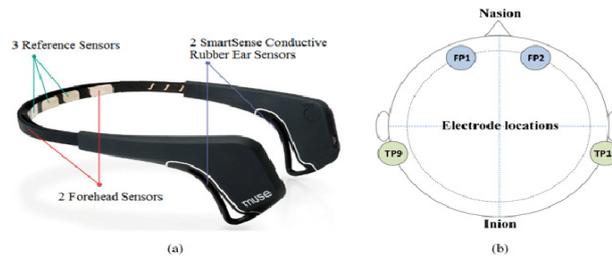


Fig. 1. (a) Device (b) Electrode positions [27].

3.2. Machine learning classifiers

The background information on the classifiers that were utilized in this study can be found in a condensed form in the following section.

3.2.1. EXtreme Gradient Boosting (XGBoost) and the Modified XGBoost

Friedman invented gradient boosting, and XGBoost implemented it efficiently and scalably. It uses linear models, trees, and solver learning. Regression, scoring, and classification are among its objective functions. XGBoost's expandability lets users easily customize their goals [28]. A focal loss function was used in a modified XGBoost model [14]. This change reduced the performance gap between training and test sets, improving model accuracy and reliability. The predictive performance of the model improved by integrating the focal loss function with the original XGBoost classifier.

3.2.2. Random forest

The RF algorithm is an ensemble classifier that constructs a variety of decision trees by making a random selection of the training variables and data. Because of its trustworthy classifications, it has a strong reputation in the field of remote sensing [29]. RF improves accuracy and robustness by utilizing multiple decision trees, which is one of the reasons why it is such a popular choice for use in remote sensing applications.

3.2.3. Decision tree

To construct a DT, the feature space is partitioned iteratively based on the training set. The aim of a DT model is to generate a set of decision rules that effectively divide the feature space, resulting in an efficient and robust hierarchical classification model. The process begins with the entire training set at the root node. In each step, the algorithm identifies the feature that provides the best data split, employing certain criteria like maximizing information gain or minimizing impurity. This splitting process continues recursively, creating branches and sub-nodes, until a stopping criterion is satisfied, such as reaching a maximum depth or having a minimum number of samples in a node [30].

3.2.4. Support-vector machines

SVM refer to a category of machine learning algorithms utilized for solving classification and regression problems. SVMs are particularly effective when it

comes to binary classification tasks, where the objective is to categorize data points into two distinct groups. The fundamental concept behind SVMs involves transforming the input vectors into a high-dimensional feature space using a non-linear mapping. Within this transformed feature space, SVMs strive to identify an optimal decision boundary, known as the hyperplane, that maximally separates the two classes. This decision boundary is constructed in a manner that maximizes the margin, which represents the distance between the decision boundary and the closest data points from each class. These data points, situated closest to the decision boundary, are referred to as support vectors [31].

3.2.5. Naive Bayes classifiers

NB classifiers are particularly well-suited for large-scale problems because the number of parameters they require scales linearly with the number of variables (features/predictors) involved in a learning task. This makes them highly applicable in such scenarios. Unlike many other classifiers that employ iterative approximation, maximum likelihood classifiers can be trained in linear time using a closed-form expression evaluation, which is more time-efficient and avoids unnecessary computational overhead [32].

3.2.6. Logistic regression

LR is a regression analysis technique where the log-odds of an event are modelled as a linear combination of one or more independent variables in a logistic model, also known as a logit model [33]. This method is used to estimate the parameters (coefficients) of the logistic model in regression analysis. In binomial logistic regression, there is a single dependent variable with two possible values (0 or 1), represented by an indicator variable. The independent variables can be binary (0 or 1) or continuous (any real value). The label "1" indicates the possibility of the value falling between 0 and 1, and the logistic function is utilized to convert logarithmic odds into probabilities. The term "logit" is derived from "logistic unit," as it represents the unit of measurement on the log-odds scale [33].

3.2.7. Linear discriminant analysis

LDA also known as Normal Discriminant Analysis (NDA) or Discriminant Function Analysis (DFA), is a statistical method employed to discover a linear combination of features that effectively distinguishes between multiple groups or categories of things or events. LDA can be utilized in two primary ways: as a linear classifier or, more commonly, as a dimensionality reducer to facilitate subsequent classification. LDA is closely related to other statistical techniques such as ANOVA and regression analysis, all of which aim to model the dependent variable as a linear function of independent variables [34]. While discriminant analysis employs continuous independent variables and a categorical dependent variable, analysis of variance (ANOVA) employs categorical independent variables and a continuous dependent variable [35].

3.3. Fast Fourier transform

Fast Fourier Transform (FFT) calculates a sequence or signal's discrete Fourier transform (DFT) quickly. The DFT divides a time-domain signal into frequency components [36].

The FFT uses Fourier transform symmetry and periodicity to simplify computation. The algorithm recursively solves DFT subproblems until it reaches the base case.

Signal processing, image processing, data analysis, and more use the FFT algorithm. It has revolutionized these domains by enabling rapid and accurate frequency analysis, facilitating noise filtration, spectral analysis, compression, pattern recognition, and more [36].

3.4. Wavelet transform

The Wavelet Transform analyses time- and frequency-domain signals and data. It breaks a signal into wavelets, small functions with time and frequency localization. This makes it more detailed and flexible than the Fourier Transform [37].

The Wavelet Transform allows multi-resolution analysis of high- and low-frequency signals. This transformative approach provides valuable information about the amplitude and location of signal frequency components, enabling efficient signal processing and analysis [37].

4. Methodology

The depiction of the proposed method can be observed in Fig. 2. Exhaustive explanations of each sub step are provided below, offering a comprehensive understanding of the approach.

4.1. Data collection

The experiments exclusively focused on EEG data as the primary source of information. EEG activity data of patients were collected from the ICU and care units of Al Bashir Hospital during the years 2021 and 2022. Specifically, individuals who had experienced a stroke within the preceding 72 hours were selected as participants. In 2021, the study involved 17 stroke patients and 32 healthy controls, while in 2022, 16 stroke patients and 21 controls participated. Two measurements from stroke patients were excluded due to excessive noise and lack of clarity. To ensure a fair comparison, 31 stroke patients and 31 healthy individuals were randomly selected from the available pool.

Prior to commencing any experiments, participants or their legal guardians were provided with detailed information about the procedures and requested to provide informed consent. The participants' scalp and earlobes were cleaned using alcohol and swabs, and the MUSE2 device was also cleaned before and after each session using alcohol wipes. Following a brief break, two separate three-minute EEG recordings were made (one with the patient's eyes open and one with them closed). A fixation cross was placed in the centre of the patient's field of vision so that he could rest with his eyes open and focus on it.

The final dataset contains 32 men, with an average age of 58.18, a standard deviation of 12.75, the oldest being 79, and the youngest being 31. The mean age of the 30 women in the sample was 55 years old, the standard deviation was 12.77 years, the age range represented was 37-81, and the median age was 55. The expert physicians at Al Bashir Hospital agreed on the outcomes for all patients.

4.2. Data cleaning and features extraction

Data preparation for algorithmic analysis is a complex procedure that necessitates close attention to detail, especially when dealing with signal errors. In the first step of data processing, data cleaning is performed using FFT and Wavelet Transform. The FFT was used to examine the signal's frequency spectrum. This required frequency-domain transformations of signals and the application of filters to remove unwanted interference and noise. The Wavelet Transform was also used to separate the signal from the noise by removing the high-frequency noise components. Reconstructing the original signals after removing or weakening these components significantly reduced the background noise level.

Signals were transformed and cleaned to extract many features. These extracted features include: The Delta-Theta Ratio (DAR) is calculated by dividing the total power at the delta frequency range (1-3 Hz) by the power at alpha (8-13 Hz) in an EEG signal. Formula 1 shows the relative power or activity in these frequency bands.

$$DAR = \frac{(Delta)}{(Alpha)} \quad (1)$$

The Delta-Theta-Alpha-Beta Ratio (DTABR) is derived from EEG signals. It is calculated by adding the EEG signal's delta (1-3 Hz) and theta (4-6 Hz) voltages. The sum of these voltages is divided by the sum of the voltages associated with the alpha frequency range (8-13 Hz) and the beta frequency range (14-20 Hz). The DTABR can be calculated using formulas 2.

$$BATR = \frac{(Delta + Theta)}{(Alpha + Beta)} \quad (2)$$

The pdBSI (Pair-Derived Brain Symmetry Index) measures spectral power density symmetry between EEG electrode pairings. The pdBSI formula is:

$$\sum_{j=1}^M \sum_{i=1}^n \left| \frac{(R_{ij} - L_{ij})}{(R_{ij} + L_{ij})} \right| \quad (3)$$

In this formula, The variables R_{ij} and L_{ij} represent the spectral power density of the signals at each electrode pairing ($i=1, 2, \dots, M$) and frequency ($j=1, 2, \dots, N$) in this formula. The i values range from 1 to M , and the j values range from 1 to N . The difference in power densities of the signals coming from the right (R) and left (L) electrodes is what the numerator of the formula is used to calculate. On the other hand, the power densities are added together when using the denominator to perform the calculation. The primary categories, features, and parameters that have been extracted from the dataset are outlined in detail in Table 1, which offers a comprehensive overview of these elements.

4.3. Processing

The frequency properties, such as beta, alpha, theta, delta, and gamma, were gathered and compiled during this stage of the process. The gathered information was then uploaded to Dropbox for further processing. The first step in the preprocessing stage involved removing artifacts from the signals using the Wavelet Transform and FFT. This process, as depicted in Fig. 2, was carried out using Python within the Jupyter Notebook environment on Google Colab.

In order to achieve the best possible results and the highest level of accuracy, ML methods were employed in the classification stage. The accuracy of the modified XGBoost classifier was validated by comparing it with various other ML classifiers.

Table 1. Features from signals.

Main categories for features	
Age	Gyroscope- Root Mean Square (RMS)-X
Gender	Gyroscope-RMS-Y
Delta-Theta Ratio (DAR) hemisphere	Gyroscope-RMS-Z
Delta-Theta-Alpha-Beta Tatio (DTABR)	Gyroscope-Standard Deviation(STD)-X
Pair-Derived Brain Symmetry Index (pdBSI)	Gyroscope-STD-Y
Relative-Beta power	Gyroscope-STD-Z
Relative-Alpha power	Gyroscope-RMS-X
Relative-Theta power	Accelerometer-RMS-Y
Relative-Delta power	Accelerometer-RMS-Z
High-Frequency-pdBSI	Accelerometer-STD-X
Low-Frequency-pdBSI	Accelerometer-STD-Y
PdBSI-Front and Back	Accelerometer-STD-Z
Frequency of Delta	Frequency of Alpha
Frequency of Beta	Frequency of Theta

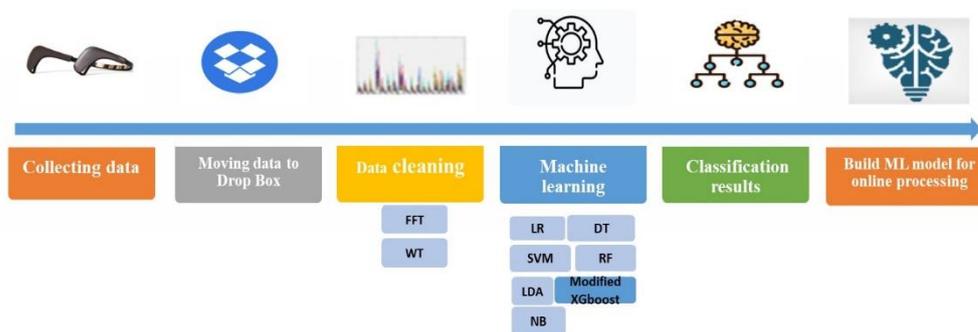


Fig. 2. Processing.

4.4. Data preparation and ML steps

The data preparation and ML steps are listed as below:

- Replace the null value with a zero value.
- Using the dummies approach, convert category variables to binary values.
- Normalize and scale data Normalization refers to the calculation of measured statistical characteristics in the range of 0 to 1. The data were subjected to normalization in order to guarantee that each characteristic is accorded the same amount of significance and lies within the acceptable range of values, which is between 0.0 and 1.0. The equation for normalization (Equation 4)

made use of the mean (μ) and standard deviation (σ) of attribute X, with a weighted value (α) that was set to 1/0.

$$DAR = \frac{(\Delta)}{(\alpha)} \quad (4)$$

- RFE selects features. It removes less important features from a dataset to find the most important ones. RFE reduces the "curse of dimensionality" in high-dimensional data. RFE improves model interpretability and performance by focusing on the most informative features.
- Five types of data partitioning are used, including 5-fold cross-validation, 10-fold cross-validation, 20-fold cross-validation, an 80/20 training-testing split, and a 67/33 training-testing split.
- Seven different machine learning classifiers are used on the dataset, and their effectiveness is analysed in terms of accuracy, recall, precision, and F-score.

The results of accuracy, recall, precision, and F-score based on the confusion matrix are presented in Table 2.

Table 2. Confusion matrix.

Confusion Matrix	Actual Stroke	Actual Normal
Prediction Stroke	TP	FP
Prediction Normal	FN	TN

- FP (False Positives): This occurs when a healthy person is mistakenly diagnosed as having a stroke.
- FN (False Negatives): This happens when someone who is experiencing a stroke appears normal and is not detected.
- TP (True Positive): This refers to correctly identifying an individual who is experiencing a stroke.
- TN (True Negative): This indicates successfully identifying a normal individual without a stroke.

The following criteria are used to evaluate classifiers:

- Accuracy is a performance metric utilized to assess the effectiveness of classifiers. It quantifies the overall accuracy of a classifier in correctly classifying instances as either stroke or non-stroke cases. Accuracy is computed by dividing the total number of correctly classified cases (true positives and true negatives) by the total number of cases present in the dataset. A higher accuracy score indicates a more dependable classifier that accurately predicts both stroke and non-stroke cases.

$$\frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (5)$$

- Precision is a performance metric commonly employed to evaluate classifiers, particularly in binary classification tasks. It quantifies the proportion of relevant instances (true positives) correctly identified out of all the instances classified as positive by the classifier (true positives and false positives). Essentially, precision assesses how effectively the classifier identifies true positive cases while minimizing the inclusion of false positive cases. A higher

precision value signifies a classifier with a lower rate of false positives and greater precision in identifying relevant instances.

$$\frac{(TP)}{(TP+F)} \quad (6)$$

- Recall, also referred to as sensitivity or true positive rate, is a performance metric employed to assess classifiers in binary classification tasks. It quantifies the proportion of actual positive instances (true positives) correctly identified by the classifier out of all the positive instances present in the dataset (true positives and false negatives). In essence, recall measures how well the classifier captures all relevant positive instances. A classifier with a higher recall value accurately identifies positive instances and has a lower false negative rate.

$$\frac{(TP)}{(TP+FN)} \quad (7)$$

- Precision and recall are combined in the F1-score performance metric. A classifier's performance is balanced by the harmonic mean of precision and recall. Precision and recall are considered in the F1-score. When the dataset has a positive-negative imbalance, this metric is useful. A classifier with a high F1-score balances precision and recall, resulting in better performance

$$\frac{2(Precision*Recall)}{(Precision+Recall)} \quad (8)$$

5. Results

5.1. Experiments setup

This section describes the experimental system environment and parameter settings. Google Colab hosted a Python Jupyter notebook for all experiments. This ensured accurate comparisons. The notebook's 12.68 GB of RAM and 107.72 GB of disk space allowed SKlearn, Matplotlib, Pandas, and Numpy to implement a variety of machine learning algorithms. Notebooks used these libraries.

Table 3 lists parameter settings. Cross-validation (CV) with fivefold, tenfold, and twentyfold iterations assessed machine learning algorithm performance. Five, ten, and twenty data samples were used for training (80%, 90%, and 95%) and testing (20%, 10%, and 5%) respectively. This process was repeated five, ten, or twenty times, ensuring that each sample had the opportunity to appear in both the training and testing sets. By utilizing CV, we obtained less biased results compared to the simple method, which was also used for testing in this paper (80% training and 20% testing) and (67% training and 33% testing). repeated five times, These different types of splitting were conducted to ensure the achievement of the highest accuracy possible. It is important to note that the best results are highlighted using boldface formatting.

Various classic and advanced classification algorithms were evaluated to determine the most effective one for the given dataset. Among the algorithms tested were Modified XGBoost, NB, NB, SVM, DT, RF, LR, and LDA. The specific configurations and settings of these approaches can be found in Table 3.

5.2. Data exploration and classification

After collecting the original dataset for this study, it consisted of 34 features. However, through the process of feature extraction, the dataset was expanded to include a total of 66 features that are suitable for ML classification. Figure 3 visually represents the data and highlights significant trends in certain features between patients with stroke and normal individuals.

Table 3. List of the parameter's settings for the classifiers.

Classifier	Parameter	Value
Decision Tree	Min samples	2
	Tree depth	5
Naive Bayes	Var smoothing	1.00E-09
Support Vector Machine	Degree	3
	Kernel set	linear
Modified XGBoost	Max depth	4
	Objective	binary: logistic
	N Estimators	100
	Min child weight	4
	Learning rate	0.2
Random Forest	Max depth	70
	N Estimators	400
	Bootstrap	TRUE
	Min samples split	10
Logistic Regression	Solver	newton-cg
	Penalty	None
	max iter	1000
Linear Discriminant Analysis	Solver	svd
	Shrinkage	auto

The results of different classifiers were presented in the section, including the modified XGBoost classifier, as well as DT, LDR, LR, SVM, NB, and RF. The performance evaluation of each classifier was conducted using five different approaches for data partitioning. Accuracy and F1-score results are presented in Table 4, while precision and recall results are shown in Table 5. The best results for each classifier on the test data are indicated in bold. The findings depicted in Table 4 clearly demonstrate that the modified XGBoost classifier outperformed all other classifiers in terms of accuracy and F1-score. Specifically, when employing 10-fold cross-validation to split the dataset, the modified XGBoost achieved the highest accuracy of 96.87% (as illustrated in Fig. 4) and an F1-score of 96.87%. These results highlight the superior performance of the modified XGBoost classifier in accurately classifying the dataset.

When compared to the dataset that was divided into 80% training and 20% testing, the modified XGBoost classifier demonstrated a significant improvement of 0.30%. This is an important finding that deserves to be highlighted. This indicates that a more extensive training dataset contributed, at least in part, to the improved accuracy of the classifier.

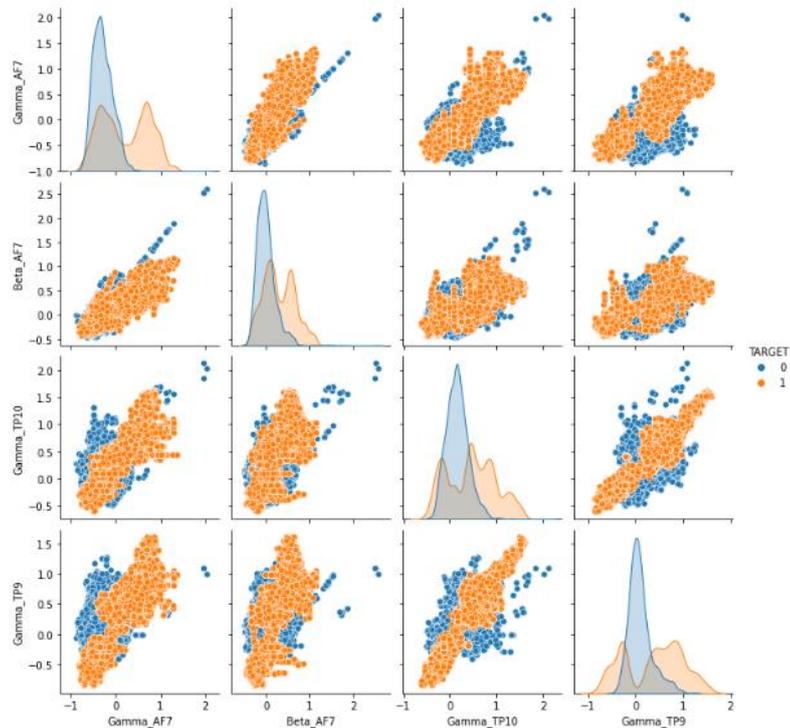


Fig. 3. Data bases on target.

Table 4. Accuracy and F1-score results.

Classifier	Train 80/20		Train 67/33		5-Fold CV		10-Fold CV		20-Fold CV	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
XGBoost	96.57	96.47	96.73	96.66	96.78	96.7	96.87	96.87	96.86	96.87
RF	92.61	92.16	92.44	92.03	92.63	92.38	92.52	92.25	92.51	92.23
SVM	87.05	86.77	86.96	86.72	86.74	86.61	86.53	86.45	86.52	86.48
NB	77.94	77.1	84.91	84.87	84.99	85.15	85.03	85.19	85.01	85.17
DT	88.64	88.41	90.14	90.45	89.18	89.4	89.06	89.31	88.97	89.11
LR	85.52	85.09	85.53	85.13	86.82	86.83	86.75	86.77	86.78	86.79
LDA	86.61	86.33	86.42	86.18	86.17	86.18	86.11	86.1	86.1	86.08

On the other hand, the NB classifier demonstrated the worst accuracy and F1-score, in particular when the dataset was divided so that 80% of it was used for training and 20% was used for testing. The accuracy of the NB classifier was measured at 77.94%, and its F1 score was measured at 77.1%. Based on these findings, it appears as though the NB classifier does not perform as well as other classifiers when it comes to accurately classifying the dataset.

The findings, taken as a whole, shed light on the outstanding performance of the modified XGBoost classifier, in particular when employing the 10-fold cross-validation method for the purpose of dataset splitting.

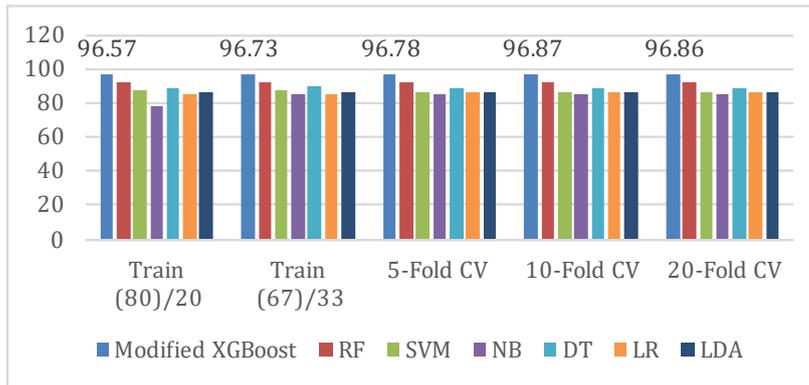


Fig. 4. The accuracy of the classifiers.

In Table 5, we see further evidence that the modified XGBoost classifier outperforms other classifiers in terms of precision and recall. The results clearly show that a split of 80% training data and 20% testing data yielded the highest precision (97.15%) for the modified XGBoost classifier. The modified XGBoost classifier also had the highest recall (97.12%) using 10-fold cross-validation. These results show that the improved XGBoost classifier distinguishes false positives and negatives well. The NB classifier had the lowest recall and precision rates with an 80% training and 20% testing data split, 77.47% and 76.73%, respectively. These results suggest that the NB classifier may underclassify stroke and non-stroke cases.

Table 5. Recall and precision results.

Classifier	Train 80/20		Train 67/33		5-Fold CV		10-Fold CV		20-Fold CV	
	Rec	Prc	Rec	Prc	Rec	Prc	Rec	Prc	Rec	Prc
XGBoost	95.81	97.15	96.53	96.8	97.03	96.55	97.12	96.62	97.11	96.64
RF	94.68	89.77	95.06	89.2	89.42	95.55	89.1	95.64	89.06	95.66
SVM	85.83	87.72	86.51	86.92	86.66	86.27	86.63	86.49	86.42	86.38
NB	77.47	76.73	83.35	86.45	86.11	84.2	86.22	84.19	86.22	84.16
DT	87.34	89.52	86	95.39	91.35	87.57	91.57	87.21	90.32	87.94
LR	84.86	85.32	85.65	84.62	86.95	86.73	86.93	86.63	86.93	86.68
LDA	85.32	87.37	85.86	86.51	86.23	86.14	86.12	86.11	86.06	86.14

In conclusion, the modified XGBoost classifier's precision and recall are excellent, as shown in Table 5. The modified XGBoost classifier outperforms the other classifiers, including the NB classifier, in dataset classification. The classifier outperforms other classifiers consistently. These results demonstrate the modified XGBoost classifier's ability to produce accurate and reliable results in the classification task.

The Modified XGBoost model's 10-fold cross-validation ROC curves are shown in Fig. 5. The curves demonstrate the model's classification accuracy.

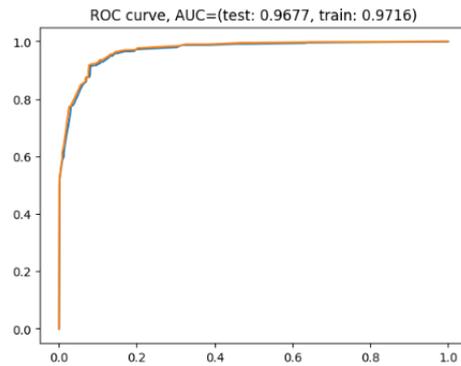


Fig. 5. ROC curves for the Modified XGBoost.

5.3. Comparing the results with other studies

We utilized the modified XGboost model on an international dataset that employed MUSE-2 for capturing EEG waves from stroke patients [38]. The results, as presented in Table 6, demonstrated a notable 9.2% increase in accuracy compared to previous findings. Additionally, when compared to the results reported in [26] on the same dataset, our approach achieved a further 1.31% increase in accuracy.

Table 6. Dataset from paper [38].

Dataset	Accuracy	Precision	Recall	F1-score
Results by using proposed model	0.8520	0.8616	0.8620	0.8617
Results in paper [38]	0.76			
Results in paper [26]	0.8389	0.8518	0.8473	0.8493

6. Discussion

The primary aim of this research is to utilize ML models for stroke classification based on EEG signals captured by the MUSE 2 portable device. We have developed an innovative cloud-based decision support system for stroke identification. The system's structure is illustrated in Fig. 6, and the key steps of the proposed model are as follows:

- Real-time data collection: EEG signals are collected using the MUSE 2 wearable device.
- Data gathering and preparation: Python in the Jupiter Notebook, accessed through Google Colab, is employed to retrieve data from Dropbox and perform the necessary cleaning and pre-processing steps to make it suitable for ML algorithms.
- Stroke outcome classification: Traditional and modified ML models classify stroke outcomes using the dataset.
- Categorization result and notification: An ML model classifies and sends an SMS to the specified phone number.

Our stroke diagnosis system is novel because it uses these procedures. It uses the MUSE 2 device, cloud computing, and machine learning algorithms to gather real-time data, process it, and classify it to quickly identify and notify stroke cases.

Its affordability makes this system accessible to a wide audience. It offers a cost-effective solution for people of different financial means. The system's adaptability makes it easy to use anywhere.

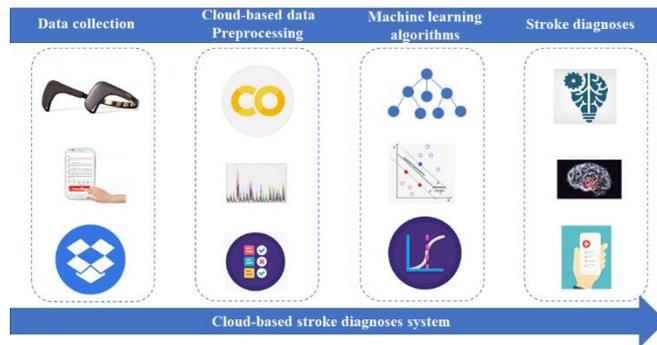


Fig .6. Module structure.

Internet and the MUSE2 device are needed to use the system, By having these two components, healthcare providers can initiate the diagnostic process and anticipate receiving the results on their mobile phones within a matter of minutes. This swift response time allows for timely interventions and decision-making concerning stroke-affected patients. In general, the system's cost-effectiveness, availability, and prompt result delivery establish it as an invaluable resource for stroke diagnosis, benefiting healthcare professionals and patients alike.

7. Conclusion and Future Work

Currently, the diagnosis of stroke presents challenges as it typically relies on CT scans, which can take up to 8 hours to confirm, or MRI scans, which are often impractical due to their lengthy thirty-minute procedure time. Consequently, there is a need for alternative approaches to analyse EEG waves for stroke diagnosis. Thus, we propose a cloud-based stroke diagnosis system that leverages the portable MUSE 2 device for this purpose.

In our research, we employed a modified XGboost along with other machine learning classifiers to classify the stroke dataset. We captured measurements from various positions using the four electrodes on the MUSE 2 portable device, enabling us to extract raw spectra and raw spectrum values. Additionally, we generated and evaluated 66 new attributes by extracting further features from the collected dataset.

To enhance efficiency and provide immediate access to results, diagnostic findings will be promptly delivered to authorized individuals via real-time short messages on their mobile phones. Furthermore, our study employed a modified XGboost in a 10-fold cross-validation, achieving a 96.87% prediction accuracy for stroke patients based on data gathered from Al Bashir Hospital.

These findings present compelling evidence that this system effectively delivers timely diagnoses for stroke patients, highlighting its potential impact.

However, there are certain considerations to bear in mind. While the cloud computing environment offers accessibility from any location, it is advisable to utilize high-speed internet for optimal performance. As this system has shown

promising results for stroke patients, our future research will focus on its practical application and integration into clinical trials for real-world implementation. Additionally, we plan to incorporate electrocardiogram (ECG), hypertension, and other vital signs readings into the system to enhance its diagnostic capabilities for stroke patients.

References

1. Wijeratne, T.; Menon, R.; Sales, C.; Karimi, L.; and Crewther, S. (2020). Carotid artery stenosis and inflammatory biomarkers: the role of inflammation-induced immunological responses affecting the vascular systems. *Annals of Translational Medicine*, 8(19), 1276.
2. Wilkinson, C.M.; Burrell, J.I.; Kuziek, J.W.; Thirunavukkarasu, S.; Buck, B.H.; and Mathewson, K.E. (2020). Application of the muse portable eeg system to aid in rapid diagnosis of stroke. *medRxiv*, 1-30.
3. Saba, L.; Loewe, C.; Weikert, T.; Williams, M.C.; Galea, N.; Budde, R.P.J.; Vliegthart, R.; Velthuis, B.K.; Francone, M.; Bremerich, J.; Natale, L.; Nikolaou, K.; Dacher, J.-N.; Peebles, C.; Caobelli, F.; Redheuil, A.; Dewey, M.; Kreitner, K.-F.; and Salgado, R. (2022). State-of-the-art CT and MR imaging and assessment of atherosclerotic carotid artery disease: the reporting—a consensus document by the european society of cardiovascular radiology (ESCR). *European Radiology*, 33(2), 1088-1101.
4. Puig, J.; Shankar, J.; Liebeskind, D.; Terceño, M.; Nael, K.; Demchuk, A.M.; Menon, B.; Dowlathshahi, D.; Leiva-Salinas, C.; Wintermark, M.; Thomalla, G.; Silva, Y.; Serena, J.; Pedraza, S.; and Essig, M. (2020). From “Time is Brain” to “Imaging is Brain”: A paradigm shift in the management of acute ischemic stroke. *Journal of Neuroimaging*, 30(5), 562-571.
5. Paprottka, K.J.; Kupfer, K.; Riederer, I.; Zimmer, C.; Beer, M.; Noël, P.B.; Baum, T.; Kirschke, J.S.; and Sollmann, N. (2021). Impact of dose reduction and iterative model reconstruction on multi-detector ct imaging of the brain in patients with suspected ischemic stroke. *Scientific Reports*, 11, 22271.
6. Dagonnier, M.; Donnan, G.A.; Davis, S.M.; Dewey, H.M.; and Howells, D. W. (2021). Acute stroke biomarkers are we there yet. *Frontiers in Neurology*, 12: 619-721.
7. Zhang, J.; Zhu, P.; Liu, B.; Yao, Q.; Yan, K.; Zheng, Q.; Li, Y.; Zhang, L.; Li, M.; Wang, J.; Zhu, C.; and Zhou, M. (2019). Time to recurrence after first-ever ischaemic stroke within 3 years and its risk factors in Chinese population: a prospective cohort study. *BMJ Open*, 9(12), e032087..
8. Soltanian-Zadeh, H.; Lau, A.; Elisevich, K.; and Soltanian-Zadeh, H. (2019). Multimodal analysis in biomedicine. In: *Big Data in Multimodal Medical Imaging*. Chapman and Hall/CRC, 193-203.
9. Ho, K.C.; Speier, W.; El-Saden, S.; and Arnold, C.W. (2017). Classifying acute ischemic stroke onset time using deep imaging features. *AMIA Annual Symposium Proceedings*, 2017: 892-901
10. Roy, S.; Bhattacharjee, S.; Mandal, B. K.; and Saha, M. (2022). A comparative study on different techniques for classification of brain waves from EEG signals. *American Journal of Science & Engineering*, 2(4), 1-7.

11. Abiri, R.; Borhani, S.; Sellers, E. W.; and Jiang, Y.; and Zhao, X. (2019). A comprehensive review of EEG-based brain-computer interface paradigms. *Journal of Neural Engineering*, 16(1): 011001.
12. Jusseume, K.; and Valova, I. (2022). Brain age prediction/classification through recurrent deep learning with electroencephalogram recordings of seizure subjects. *Sensors*, 22(21), 8112.
13. O'Shea, A.; Lightbody, G.; Boylan, G.; and Temko, A. (2020). Neonatal seizure detection from raw multi-channel EEG using a fully convolutional architecture. *Neural Networks*, 123, 12-25.
14. Raveendra, K.T.H.; Narayanappa, C.K.; Raghavendra, S.; and Poornima, G.R. (2022). A modified XG Boost classifier model for detection of seizures and nonseizures. *WSEAS Transactions on Biology and Biomedicine*, 19, 14-21.
15. Lee, M.; Ryu, J.; and Kim, D.-H. (2020). Automated epileptic seizure waveform detection method based on the feature of the mean slope of wavelet coefficient counts using a hidden Markov model and eeg signals. *ETRI Journal*, 42(2), 217-229.
16. Yang, B.; Huang, Y.; Li, Z.; and Hu, X. (2022). Management of post-stroke depression (PSD) by electroencephalography for effective rehabilitation. *Engineered Regeneration*, 4(1), 44-54.
17. Choi, Y.-A.; Park S.; Jun, J.-A.; Ho, C.M.B.; Pyo, C.-S.; Lee, H.; and Yu, J. (2021). Machine-learning-based elderly stroke monitoring system using electroencephalography vital signals. *Applied Sciences*, 11(4), 1761.
18. Wilkinson, C.M.; Burrell, J.I.; Kuziek, J.W. P.; Thirunavukkarasu, S.; Buck, B.H.; and Mathewson, K.E. (2020). Predicting stroke severity with a 3-min recording from the muse portable eeg system for rapid diagnosis of stroke. *Scientific Reports*, 10, 18465.
19. Cicioğlu, M.; and Çalhan, A. (2019). SDN-based wireless body area network ~ routing algorithm for healthcare architecture. *ETRI Journal*, 41(4), 452-464.
20. Ip, Z.; Rabiller, G.; He, J.W.; Yao, Z.; Akamatsu, Y.; Nishijima, Y.; Liu, J.; and Yazdan-Shahmorad, A. (2019). Cortical stroke affects activity and stability of theta/delta states in remote hippocampal regions. *Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany, 5225-5228.
21. Choi, Y.-A.; Park, S., Jun, J.-A.; Ho, C. M. B.; Pyo, C.-S.; Lee, H.; and Yu, J. (2021). Machine-learning-based elderly stroke monitoring system using electroencephalography vital signals. *Applied Sciences*, 11(4), 1761.
22. Yu, J.; Park, S.; Kwon, S.-H.; Ho, C. M. B.; Pyo, C.-S.; and Lee, H. (2020). AI-based stroke disease prediction system using real-time electromyography signals. *Applied Sciences*, 10(19), 6791.
23. Varelas, P.N.; and Hacein-Bey, L. (2017). Ischemic stroke, hyperperfusion syndrome, cerebral sinus thrombosis, and critical care seizures. *Seizures in Critical Care*, 155-186.
24. Gottlibe, M.; Rosen, O.; Weller, B.; Mahagney, A.; Omar, N.; Khuri, A.; Srugo, I.; and Genizi, J. (2020). Stroke identification using a portable eeg device - A pilot study. *Neurophysiologie Clinique*, 50(1), 21-25.

25. Djamal, E.C.; Ramadhan, R.I.; Mandasari, M.I.; and Djajasmita, D. (2020). Identification of post-stroke eeg signal using wavelet and convolutional neural networks. *Bulletin of Electrical Engineering and Informatics*, 9(5), 1890-1898.
26. Sawan, A.; Awad, M.; and Qasrawi, R. (2022). Machine learning-based approach for stroke classification using electroencephalogram (EEG) signals. *Proceedings of Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies BIODVICES*, 1, 111-117.
27. Liu, R.; Xu, M.; Zhang, Y.; Peli, E.; and Hwang, A.D. (2020). A pilot study on electroencephalogram-based evaluation of visually induced motion sickness. *Journal of Imaging Science and Technology*, 64(2), 205011-2050110.
28. Zhou, J.; Qiu, Y.; Khandelwal, M.; Zhu, S.; and Zhang, X. (2021). Developing a hybrid model of jaya algorithm-based extreme gradient boosting machine to estimate blast-induced ground vibrations. *International Journal of Rock Mechanics and Mining Sciences*, 145, 104-126.
29. Neto, M.P.; and Paulovich, F.V. (2020). Explainable matrix-visualization for global and local interpretability of random forest classification ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 27(2), 1427-1437.
30. Lan, T.; Hu, H.; Jiang, C., Yang, G.; and Zhao, Z. (2020). A comparative study of decision tree, random forest, and convolutional neural network 388 for spread-f identification. *Advances in Space Research*, 65(8), 2052-2061.
31. Kurani, A.; Doshi, P.; Vakharia, A.; and Shah M. (2023). A comprehensive comparative study of artificial neural network (ann) and support vector machines (SVM) on stock forecasting. *Annals of Data Science*, 10, 183-208.
32. Alizadeh, S.H.; Hediehloo, A.; and Harzevili, N.S. (2021). Multi independent latent component extension of naive bayes classifier. *Knowledge-Based Systems*, 213, 106646.
33. Ma, C.; Zhou, J.; and Yang, D. (2020) . Causation analysis of hazardous material road transportation accidents based on the ordered logit regression model. *International Journal of Environmental Research and Public Health*, 17(4), 12-59.
34. Amiri, V.; and Nakagawa, K. (2021). Using a linear discriminant analysis (LDA)-based nomenclature system and self-organizing maps (SOM) for spatiotemporal assessment of groundwater quality in a coastal aquifer. *Journal of Hydrology*, 603(Part C), 127082.
35. Stella, O. (2019). Discriminant analysis: An analysis of its predictship function. *Journal of Education and Practice*, 10(5), 50-57.
36. Brigham; and E.O. (1988). *The fast Fourier transform and its applications Discriminant analysis*. Prentice-Hall Book Company, Inc.
37. Percival; D.B.; and Walden, A.T. (2000). Wavelet methods for time series analysis. *Cambridge University Press*.
38. Wilkinson, C.M.; Burrell, J.I.; Kuziek, J.W.; Thirunavukkarasu, S.; Buck, B.H.; and; Mathewson, K.E. (2020). Predicting stroke severity with a 3-min recording from the muse portable eeg system for rapid diagnosis of stroke . *Scientific Reports*, 10(1), 18465, 1-11.