Journal of Education for Pure Science- University of Thi-Qar
Vol.13, No.2 (June., 2023)
Website: jceps.utq.edu.iq *Email: jceps@eps.utq.edu.iq*

DOI: http://doi.org/10.32792/utq.jceps.10.01.01

Epileptic Seizure Detection Using Feature Importance and ML Classifiers

Alaa Lateef Noor Al-hajjar[1,a] Ali Kadhum M. Al-Qurabat[1,a,*]

[a]alaa.noor.gsci105@student.uobabylon.edu.iq [b]alik.m.alqurabat@uobabylon.edu.iq

[1]Dept. of Computer Science, College of Science for Women, University of Babylon, Babylon, Iraq

Received 23/3/2023, Accepted 16/4/2023, Published 11/6/2023

Abstract

Epilepsy seizure (ES) monitoring and detection are only two examples of the many problems that may be addressed by combining the Internet of Medical Things (IoMT) with machine learning (ML) techniques and cloud computing services. Epilepsy, a potentially fatal neurological disorder, is a worldwide problem that poses a significant threat to human health. There is an urgent need for a reliable way of identifying epileptic seizures in their early stages to save thousands of epileptic patients every year. With the use of IoMT, several medical treatments, such as epileptic monitoring, diagnosis, and other procedures, may be performed remotely, hence lowering healthcare costs and enhancing service quality. EEG datasets have made use of feature importance-based data reduction to address the problem of a high number of data points and improve the delivery of service to the end user. In this article, we use the feature importance method by applying two popular machine learning techniques extra tree classifier (ETC) and the extreme gradient boosting classifier (XGBoost). Finally, the performance of a number of tests is evaluated using experimental data from Bonn University. Also achieved is a comparison of the two approaches used. The collected findings demonstrate the efficacy of the XGBoost technique and its greater accuracy in comparison to the ETC strategy.

Keywords: Internet of Things, Internet of Medical Things, Machine learning, Epileptic seizures.

1. Introduction

There has been a lot of interest in recent years in the Internet of Medical Things (IoMT), a subset of the IoT used in the healthcare industry. When it comes to improving patient care, the healthcare business is quite pragmatic, and IoMT provides a broad variety of possibilities. Different networks may allow for the communication of a variety of sophisticated medical sensors and devices, providing access to vital information about patients' health. The improved diagnosis and treatment procedure that results from a deeper familiarity with the symptoms will allow for earlier detection of illness and faster recovery, as well as allow for remote patient monitoring. One use of IoMT in the medical field is the diagnosis of epileptic seizures [1].

Epilepsy is a common neurological illness that has a profound effect on the human brain by triggering seizures that may occur at any time. There have been a slew of recent petitions for various reasons. The disruption of normal electrical activity inside the brain is the primary cause. Possible causes include anomalies, hypoglycemia, and oxygen deprivation during labor and delivery [2] [3]. Epilepsy affects around 50 million individuals worldwide, and a further 100 million have had seizures at some point in their lives [4]. Recurrent seizures, brought on by an abnormality in brain electrical activity, are the defining feature of this illness. Shaking limbs and even passing out are common effects. Important clinical symptoms, such as strange sensations, emotions, aberrant behavior, memory loss, etc., are exacerbated by epilepsy because the condition disrupts the normal neuronal activity pattern of the human brain. People with epilepsy sometimes have life-altering accidents while doing seemingly harmless activities like driving a vehicle, swimming in a pool, crossing a road, etc. A loss of consciousness may be deadly for someone with epilepsy. They are doomed to a lesser standard of living and dependence on others for the rest of their lives. Seizures that are difficult to identify the need to be monitored closely to prevent any potential harm from occurring. Researchers are keen on seizure prediction systems [5] as a means of avoiding the associated challenges. In Figure 1, we see the recording of the seizure.

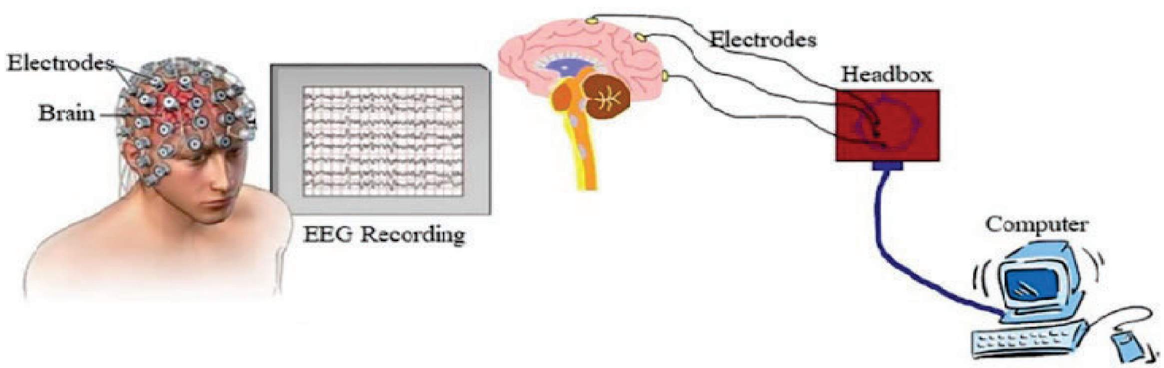

Figure 1. A general methodology of epileptic seizure recording.

As may be seen in Figure 1, electroencephalography (EEG) recordings are often utilized to identify epileptic episodes. The diagnosis of neurological disorders like epilepsy relies heavily on the analysis of EEG waves. Human brain activity may be recorded by electric impulses via EEG sensors. EEG is used to detect abnormalities in the brain's electrical activity and classify it accordingly. Electroencephalogram (EEG) records brain activity via the use of electrodes placed on the scalp or elsewhere on the body [6].

Automatic seizure detection requires identifying and assessing the unique characteristics of EEG signals before, during, and after an event. There are a few distinguishing characteristics that may be used to characterize seizure behavior. Either the signals' time- and space-invariant behavior or their dynamic features, including chaoticity and non-linearity, might be represented by these [7].

The interpretation of electroencephalography (EEG) data is complex, making aids to this process essential. While feature extraction is a common procedure in many research projects, few resources exist to assist with the process [8].

When developing a system to identify and classify seizures, it is crucial to use characteristics that accurately characterize EEG signals' behavior. There are a wide variety of suggested features and processing algorithms; some of them include time-domain [8, 9], frequency-domain [10], and temporal frequency analysis [11].

In this study, we evaluate the ETC and XGBoost models' ability to pick relevant features from a dataset and compare the results to those obtained without this step. To produce predictions and achieve high accuracy, machine learning relies on a large amount of data, features, and variables. More so than when developing the prediction model itself, feature selection is crucial. Using the dataset without first pre-processing it will only lead to inaccurate predictions.

This study is structured as follows: Sections 2 presents the complete description of the data set used in the experimental part along with the proposed methodology. Section 3 comprises the detailed discussion about the experimental procedure. Results are discussed in section 4 along with the conclusion and future work is stated in section 5.

2. Related Works

There are a number of publications using publicly accessible epilepsy seizure classification benchmark datasets.

Using the Empirical Mode Decomposition (EMD), Alam et al. [13] create a method for detecting seizures and epilepsy. To begin, appropriate characteristics are extracted using higher-order statical moments like variance, kurtosis, and skewness. After gathering these attributes, an ANN is used to diagnose epilepsy and seizures. This method is quicker than time-frequency-based approaches, yet only three features are needed to achieve an 80% classification accuracy.

The time-delay approach is investigated by Niknazar et al. [14] as part of a system identification tool for ESD. The Error-Correction Output Codes (ECOC) classifier uses features derived from recurrence quantitative analysis (RQA). Since it is not necessary to have any specific knowledge of the signal (its duration, noise, etc.), the RQA is widely used. Accordingly, the proposed ESD approach does not need any transformations or prior models. It also has the added benefit of being able to process signals of varying intensities, frequencies, and other characteristics.

In [15], an ESD technique that combines two separate approaches is shown. In order to streamline the data and extract useful characteristics, a Dual-Tree Complex Wavelet Transformation (DTCWT) is used first. Subsequently, Complex-Valued Neural Networks (CVANN) are fed the characteristics in order to complete the classification process. Using wavelet transforms, this study explores the signals at many scales.

For EEG categorization, [16] creates a data-driven approach using Multi-Layer Perceptron Neural Networks. When decomposing EEG signals into frequency subbands, a Discrete Wavelet Transform (DWT) is used initially. The K-means method is then used to classify the sub-band wavelet coefficients. After that, we calculate the probability distributions using the wavelet coefficient distributions. At long last, we feed these distributions into the MLPNN model as inputs. In this study, the k-means algorithm-based clustering method is preferred over the standard statistical methods, and is permitted to operate on the wavelet coefficients to get better results.

For epilepsy diagnosis using EEG data, Tiwari et al. [17] use a Local Binary Pattern (LBP) approach. First, a Difference of Gaussian (DOG) pyramid is used to identify certain filtered signals. (DoG). After that, a Support Vector Machine (SVM) is used to categorize the data. The therapeutic value of this diagnostic technique lies in the computational simplicity of LBP features, the ability to achieve high detection accuracy with a smaller percentage of the EEG data, and the system's applicability for online epileptic detection with lower computing load.

Acharya et al. [18] use a number of entropy-based extracted characteristics to differentiate between baseline, pre-ictal, and ictal states. We begin by exploring four distinct forms of Entropy: Approximate Entropy (ApEn), Sample Entropy (SampEn), Phase Entropy 1 (S1), and Phase Entropy 2 (S2). The characteristics are then used as input for seven different classifiers, the best of which is the Fuzzy Sugeno Classifier (FSC).

Tzallas et al. [12] create an automated ESD approach that divides seizures into three distinct groups. The time-frequency analysis is used to extract many characteristics that are then put into an artificial neural network. (ANN). The best classification performance is then shown utilizing 40 features by the classifier.

Relatedly, [19] evaluates the significance of features in classification models for the phenotype of colorectal cancer cases in Indonesia. In addition, [20] investigates the significance of variables for emotion classification and emotional speech synthesis. These traits may also be used as covariates in future genetic association studies of colorectal cancer. Feature significance analysis for an industrial recommendation system is also performed in [21], with encouraging results. In this study, we demonstrate the importance of carefully selecting characteristics from a dataset of EEG readings.

3. Material and Proposed System

Here, we provide the process that was followed to detect seizures in a database of recorded EEG disturbances. Epileptic seizure detection system schematic shown in Figure 2.

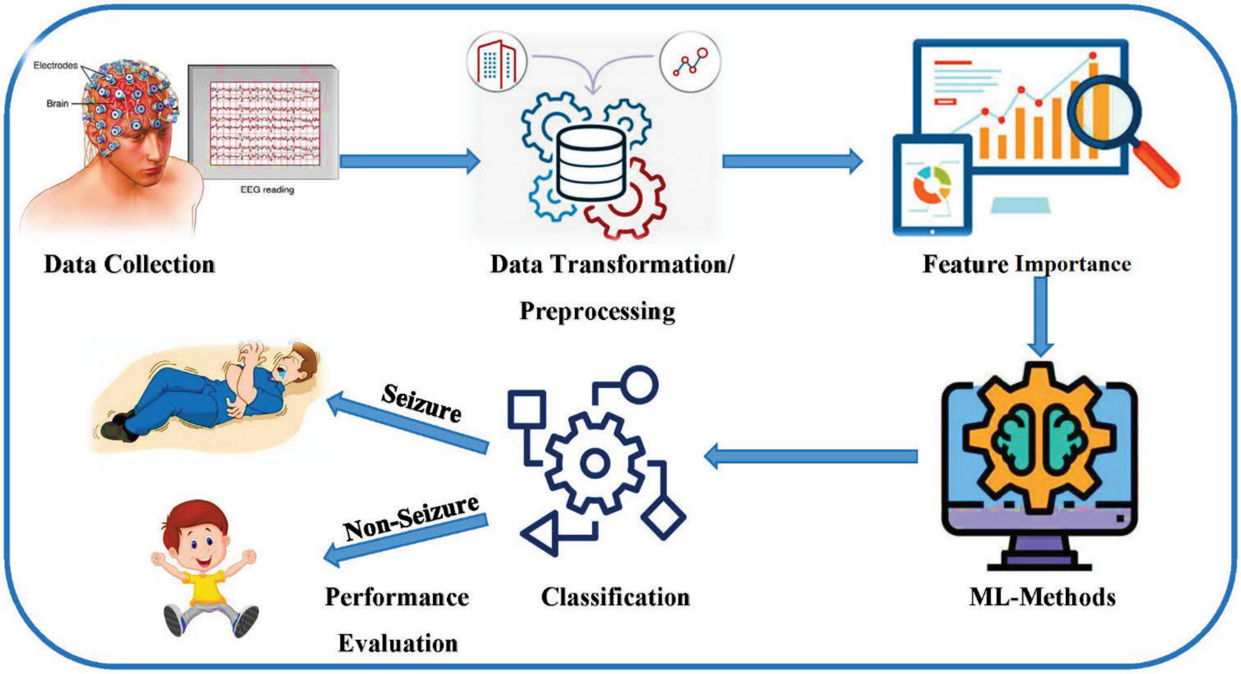

Figure 2. An outline of how EEG data and ML techniques may be used to detect seizures in a patient suffering from epilepsy.

3.1. Data Collection

The dataset of brain signals must be acquired as a preliminary stage. In this study, we examine the electronic epilepsy database (EEG) created and distributed by the Bonn University epilepsy center.

The whole EEG database consists of five sets labeled A, B, C, D, and E. There are 100 individual EEG signals from five categories, each lasting 23.6 seconds. The 10-20 electrode placement technique is often used for EEG recording. The identical 128-channel amplifier setup was used to record everyone's EEGs, and a mean reference was used for all of them. The analog signals were converted to digital at a rate of 173.61 samples per second and a bit depth of 12. This filter has a frequency range of 0.53 Hz to 85 Hz. There are 4097 samples in total, with an average of 173.61 samples per second across all channels. To test the efficacy of our approach, we have included all five datasets in this study [22].

The details of this database is available in Andrezejak et al. [23].

3.2. Data Transformation

After data gathering, transforming the data signal into a 2-D table representation is an important next step. This facilitates analysis and provides useful information, such as the identification of epileptic seizures. Therefore, we cannot rely on this raw data to offer useful information. The processing challenge has been approached using a variety of feature significance approaches. In this phase, the dataset is presented as supervised, meaning that it includes examples of classes for the class attribute [8].

3.3. Data Preprocessing

The data must be preprocessed prior to engaging in feature significance or classification actions. Since EEG data are often collected in noisy surroundings, it is challenging to design algorithms for epileptic categorization. EEG recording devices might be affected by their surroundings. Muscle and eye movements might also contribute to the commotion. The signal after being subjected to a noise filter should serve as the input signal for epileptic detection. Furthermore, the acquired input EEG signals include duplicated data with undesired noise and distortions. Eliminating these doubts is essential before continuing with post-processing on the data. When EEG signals are preprocessed, only the data that is directly relevant to the signal will remain. Electrode movement is a major source of artifacts and environmental noise in real-world EEG recordings [24]. As a consequence, the reliability of the classifications made based on the EEG signals is diminished. Two feature importance models are used to analyze the provided data.

3.4. Data Reduction Based on Feature Importance Scores

In this study, we focus on a specific subset of feature extraction called "feature importance," which employs methods that assign a numerical value to each of the input characteristics of a model. These values simply reflect the relative "importance" of the features. A higher score indicates that the attribute in question will have a greater impact on the predictive model. Choosing which characteristics to include in our model is crucial since correlated and non-redundant features may help our model perform better [25].

In this study, we show how feature importance ratings may be employed for this purpose. To achieve this, we use feature importance scores, which use a model to reduce a dataset to a subset containing just the desired characteristics. A model that has already been trained, maybe using the complete dataset, can be used with this method. To choose which characteristics to employ, it might apply a threshold. This threshold ensures that the same features are picked for both the training and testing data sets.

In addition to helping us train our model more quickly, it also simplifies the model, making it simpler to comprehend, and increasing our performance on the metrics of accuracy, precision, and recall. There are four major factors that make data minimization crucial. Reduce the complexity of the model by eliminating unnecessary parameters first. Then, we'll look at ways to shorten training sessions, improve accuracy while maintaining scalability, and circumvent the curse of dimensionality. There might be a lot of variables and features in a dataset, and they are what ultimately decide the data's usefulness and relevance in the area of data processing and analysis [26].

When determining the significance of a function's variables and features, the model's use information is crucial. The benefit of a model-based method is that it may be able to include the correlation structure between the predictors into the significance calculation, and this benefit is more tightly linked to the model performance. In a nutshell, significance is determined by a quantitative analysis. For every category, there will be a unique set of factors that each predictor considers crucial. Then, we scale all the major metrics up to a maximum of 100.

In this study, we use the whole datasets for both training and testing our models. The model is then encased in an instance whose contents are determined by the feature importance's computed from the training dataset. This allows us to choose features from the training dataset, train a model with only those features, and then assess it on the test set using the same feature selection criteria.

Various thresholds for picking features based on feature importance might be experimented with for fun. In particular, we may test each subset of features in order of significance, beginning with all features and ending with a subset with the most significant feature, by considering the feature importance of each input variable.

Classification data analysis is an area where several models shine. We'll look at the accuracy of two different classifier methods using a variety of characteristics and then choose the most effective one.

3.5. Classification

The identification of epileptic episodes is addressed by proposing two distinct ML models: XGBoost and ETC for epilepsy recognition.

3.5.1. Extra Tree Classifier

An example of ensemble learning, Extra Trees Classifier takes a "forest" of independent decision trees and uses their combined classification accuracy to draw conclusions. The main conceptual difference between this and a Random Forest Classifier is in how the decision trees in the forest are built.

The training sample is used to create one Decision Tree for the Extra Trees Forest. Each decision tree is then given a random subset of k features from the feature-set at each test node, and using some mathematical criterion, must choose the most informative feature to use in the subsequent data split (typically the Gini Index). By picking out characteristics at random, we can generate many independent decision trees.

To apply the aforementioned forest structure for feature selection, we compute, for each feature, a normalized total reduction in the mathematical criteria used for the choice of feature of split (the Gini Index, if the Gini Index is used in the construction of the forest). How relevant a trait is may be measured using the Gini Importance. An individual selects the top k attributes based on their relative importance as measured by the Gini Index [27].

In this case, we will be using Information Gain as our criterion for action. The information's entropy is determined first. Take note that the entropy formula is:-

$$Entropy(S) = \sum_{i=1}^{c} - p_i(p_i) \qquad (1)$$

where c is the total number of distinct classes and p_i is the percentage of rows where the i output label was generated [28]. The formula for Information Gain is as follows:

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \qquad (2)$$

3.5.2. Extreme Gradient Boosting

One kind of ensemble learning is XGBoost. Results from a single machine learning model may not always be reliable. With ensemble learning, it's possible to systematically pool the foresight of numerous students. As a consequence, we have a unified model that synthesizes the results of several models. When utilizing gradient boosting, it is easy to extract relevance scores for each attribute after the boosted trees have been created. In most cases, the significance score offers a numerical representation of how significant each feature was during the development of the models boosted decision trees. In decision trees, the relative relevance of a characteristic increases as it is more relied upon to make pivotal choices.

To facilitate ranking and comparison, this significance is computed directly for each characteristic in the dataset. For a given decision tree, significance is determined by summing the improvements to the performance measure at each attribute split point, weighted by the node's share of the observations. The purity (Gini index) of the data used to determine the split points is one possible performance indicator. The model's decision trees are then averaged to determine an overall feature relevance [29].

For a single decision tree $\boldsymbol{T}$:

$$I_{l(T)}^2 = \sum_{t=1}^{J-1} \hat{\imath}_t^2 I(V(t) = l) \qquad (3)$$

Where the summation is over the nonterminal nodes $\boldsymbol{t}$ of the $\boldsymbol{J}$ -terminal node tree T, $\boldsymbol{v(t)}$ is the splitting variable associated with node $\boldsymbol{t}$, and $\boldsymbol{\hat{\imath}_t^2}$ is the corresponding empirical improvement in squared error as a result of the split, defined as $\boldsymbol{i^2(R_l, R_r)} = \frac{w_l w_r}{w_l + w_r}\left(\underline{y_l} - \underline{y_r}\right)^2$, where $\underline{y}_l$, $\underline{y}_r$ are the left and right daughter response means respectively, and $\boldsymbol{w_l, w_r}$ are the corresponding sums of the weights. The squared relative importance of variable $\boldsymbol{I_l}$ is the sum of such squared improvements over all internal nodes for which it was chosen as the splitting variable [29].

Simple averaging across the trees makes this important metric applicable to additive tree expansions.

$$I_l^2 = \frac{1}{M} \sum_{m=1}^{M} I_l^2(T_m) \qquad (4)$$

As a result of averaging's stabilizing effects, this metric proves to be more trustworthy than its single-tree version [30].

Algorithm 1 illustrates the complete steps of the proposed methods.

```
Algorithm 1: feature importance for data reduction
# input: Epileptic Seizure Recognition dataset
# output: reduced dataset
# load data
        dataset ← load − data('Epileptic Seizure Recognition')
# split data into X and y
X = dataset[:,0:178]
Y = dataset[:,178]

# change the y target column (make a binary classification)
        dic ← {5: 0, 4: 0, 3: 0, 2: 0, 1: 1}
        Y ← Y.map(dic)

# split data into train and test sets
        X_train, X_test, y_train, y_test ← split(X, Y)

# fit model on all training data using XGBClassifier()or ETClassifier()
        model ← Classifier()
        model.fit(X_train, y_train)

# make predictions for test data and evaluate
        y_pred ← model.predict(X_test)
        accuracy ← accuracy_score(y_test, y_pred)

# Fit model using each importance as a threshold
        thresholds ← sort(model.feature_importances)
        for thresh in thresholds:
    # select features using threshold
    selection ← SelectFromModel(model, threshold)
        select_X_train ← selection.transform(X_train)

    # train model
    selection_model ← Classifier()
        selection_model.fit(select_X_train, y_train)

    # eval model
    select_X_test ← selection.transform(X_test)
        y_pred ← selection_model.predict(select_X_test)
        accuracy = accuracy_score(y_test, y_pred)
        return selection
```

3.6. Evaluation Parameters

Methods are compared and contrasted based on how accurately they provide outcomes. Ten-fold cross-validation is widely used as a training method because in each fold, or a horizontal section of the dataset, one section is used as the testing dataset and the other nine are used as the training dataset. The classification performance is evaluated using a variety of performance measures, including accuracy, precision, recall, specificity, prevalence and false-positive rate. These are based on four possible classification outcomes—True-Positive (TP), True-Negative (TN), False-Positive (FP), and False-Negative (FN) as presented in Table 1.

Table 1. Classification outcomes

Acronym	Detection type	Real-world scenario
TP	True-positive	If a person experiences "seizures" and they are properly identified as "seizures"
TN	True-negative	If a person is in fact normal, and the classification algorithm also identified as a "non-seizure"
FP	False-positive	Wrong detection occurs when the classifier classifies a normal person as a case of seizures.
FN	False-negative	Inaccurate detection occurs when the classification algorithm classifies a person suffers to "seizures" as a regular person. In the field of health informatics research, this is a major issue.

The following are the definitions of the equations used to calculate the performance metrics:

$$\textbf{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \qquad \textbf{(5)}$$

The proportion of identified cracks to those that are correctly classified is used to calculate recall for a function. Equation 2 shows the rate of corrected cases that are selected:

$$\textbf{Recall} = \frac{TP}{TP+FN} \qquad \textbf{(6)}$$

The accuracy of positive predictions is defined similarly by precision, which is defined as the ratio of true-positives to all cases that are recognized as positive (TP+FP), as seen in Equation 3. The low rate of false positives indicates high precision.

$$\textbf{Precision} = \frac{TP}{TP+FP} \qquad \textbf{(7)}$$

The term "prevalence" is used to describe the frequency with which a certain condition occurs within a community. Therefore, it is a valuable indicator for administrators in determining the demand for services or treatment facilities since it provides an estimate of the population's total illness burden.

Prevalence dependence on the “inflow” and “outflow” of disease according to this formula

Prevalence ≈ (incidence rate) × (average duration of illness). **(8)**

When a disease is not present, a test's specificity measures how well it can rule it out.

$$\textbf{Specificity (Sp)} = \frac{TN}{TN+FP} \qquad \textbf{(9)}$$

Also known as the likelihood of a negative test result when the illness is not present.

4. Results

The specifics of the experiments used in this study are described here. Classification of EEG data to identify seizure activity and evaluate classifier performance is the initial stage. Many different types of classification models were tested in the experiments. How well our trained system can correctly predict classes is measured by its accuracy.

Let's visualize the importance of EEG signals using ETC as an example, as shown in Figure 3. To have even better plot, we sort the features based on importance value. A feature's score or ratio will be shown on the X-axis. The y-axis indicates the presence of features. The most noticeable trait that has a large influence to produce epilepsy illness is represented by the highest node on the feature significance graph.

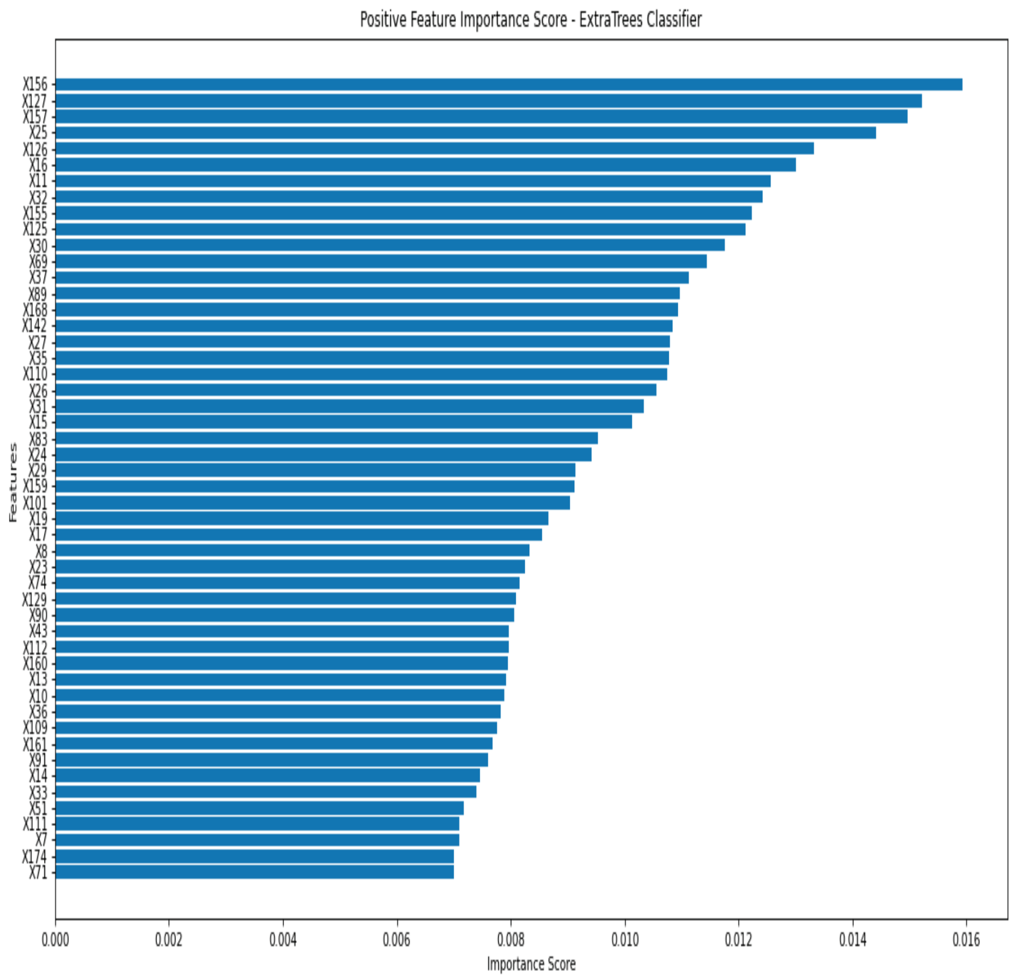

Figure 3. Feature importance of Extra Tree Classifier.

So here x156 is the most important feature as shown in Figure 3.

Table 2 shows the performance evaluation for extra tree classifier, while Table 3 shows all results about XGboost.

Table 2. Results for ETC classifier.

Performance metrics	Score
Accuracy	0.998
Recall	0.992
Precision	0.995
Specificity	0.999
Prevalence	0.199

Table 3. Results for XGBoost classifier.

Performance metrics	Score
Accuracy	1.000
Recall	1.000
Precision	1.000
Specificity	1.000
Prevalence	0.199

From Table 3, it is clearly shown that the accuracy is 100%, which gives more accurate results for the XGboost classifier than the ETC.

The quicker processing speed of XGboost explains why it achieves better outcomes than the ETC. It has various qualities, including speed, parallel computing, etc., and is likely three times as fast as the ETC classifier.

After applying the machine learning models and extracting the most important features from the data, we convert the importance values of those features into percentages in order to determine the percentage of the most importance features that we feed back to the models with taking into account the balance between high accuracy and reducing the cost of processing and transmitting this data via the IoMT by deleting the least importance features. If we deal with all the data without reducing the data, assuming that we need 64 bits to store one feature, this means we need 11392 bit to store 178 features (number of the features of used dataset) and $5.72 \times 10^{-4} Joule$ is the average of energy consumption in normal case (without data reduction).

Figure 4 shows the number of remaining features, depending on the importance of these features, as calculated using the aforementioned two methods. We can see from the figure that the XGBoost method deletes more features than the ETC method. For example, when deleting features whose importance is 10%, we note that the remaining features are 52 and 167, according to the GXBoost and ETC methods, respectively.

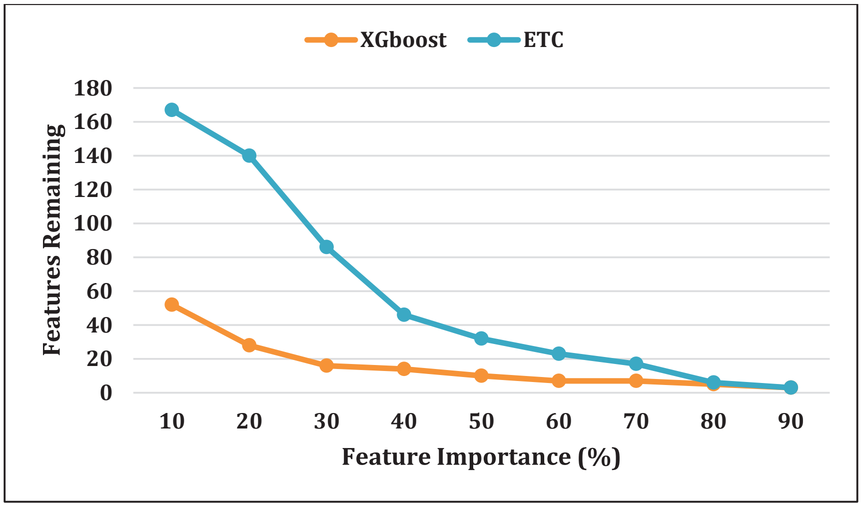

Figure 4. Features remaining using FIM and XGBoost, ETC.

Figures 5–10 show the AUC, precision, accuracy, recall, specificity, and energy consumption for different thresholds in terms of feature importance for the two classifiers, XGBoost and ETC. From the figures, it is observed that the (FIM+XGBoost) combination provides a better result compared to (FIM+ETC).

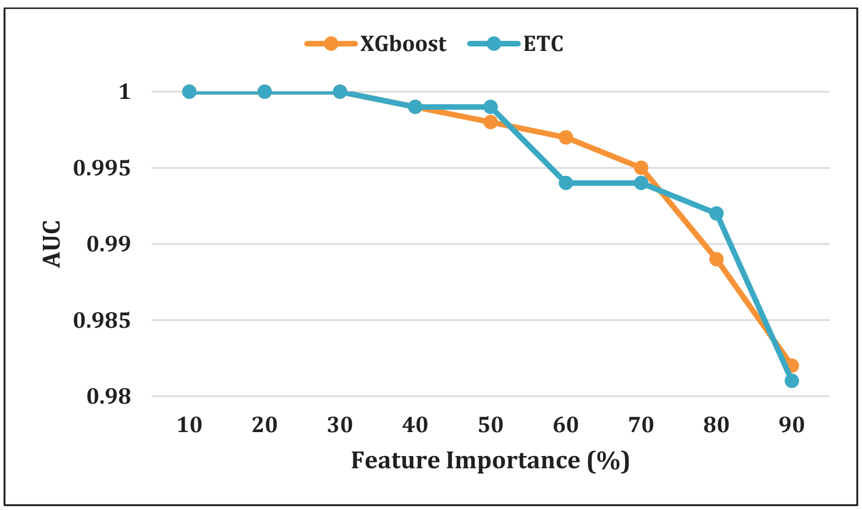

Figure 5. The AUC using FIM and XGBoost, ETC.

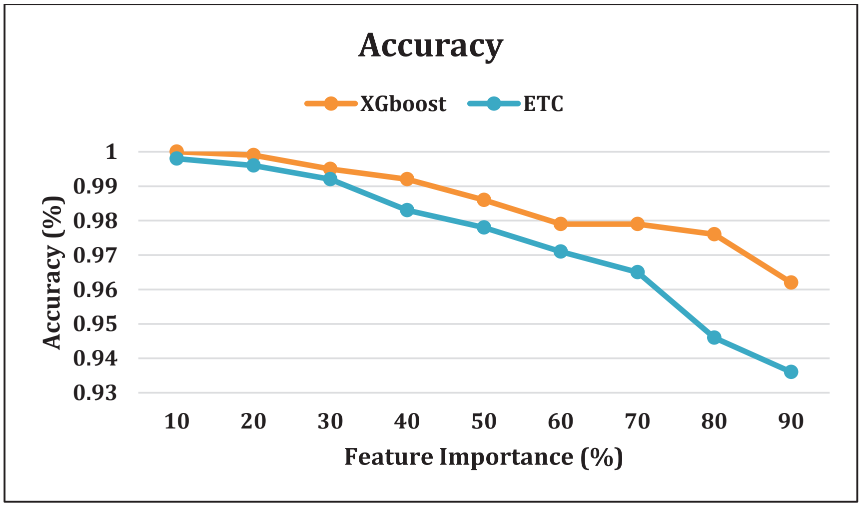

Figure 6. The accuracy using FIM and XGBoost, ETC.

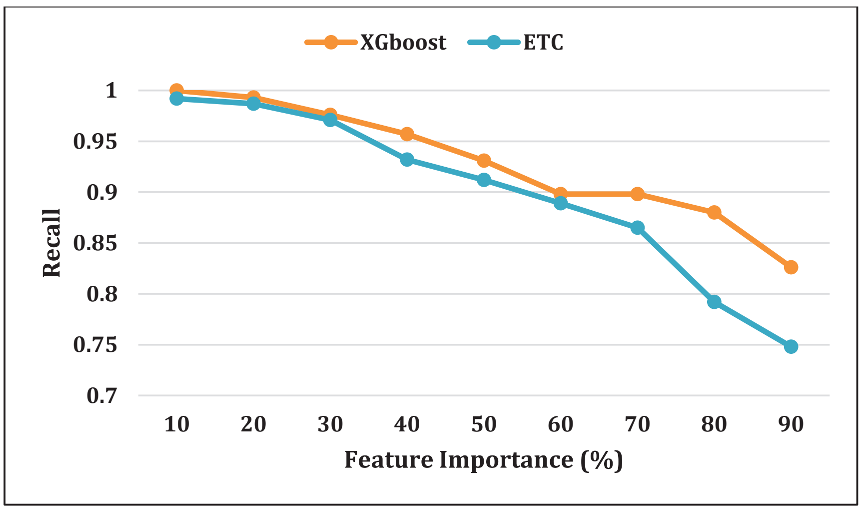

Figure 7. The recall using FIM and XGBoost, ETC.

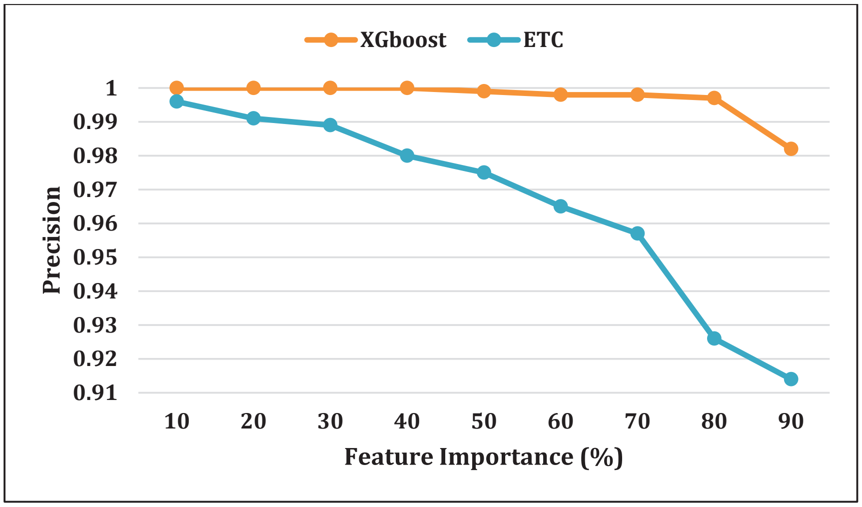

Figure 8. The precision using FIM and XGBoost, ETC.

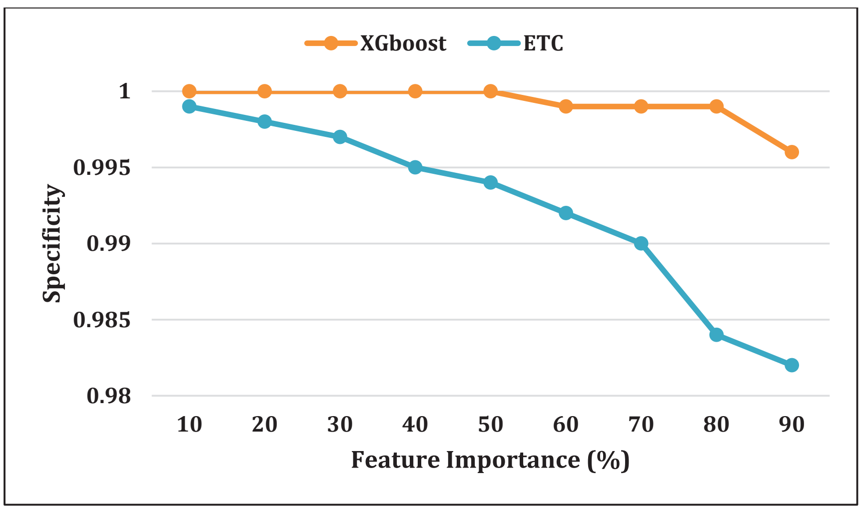

Figure 9. The specificity using FIM and XGBoost, ETC.

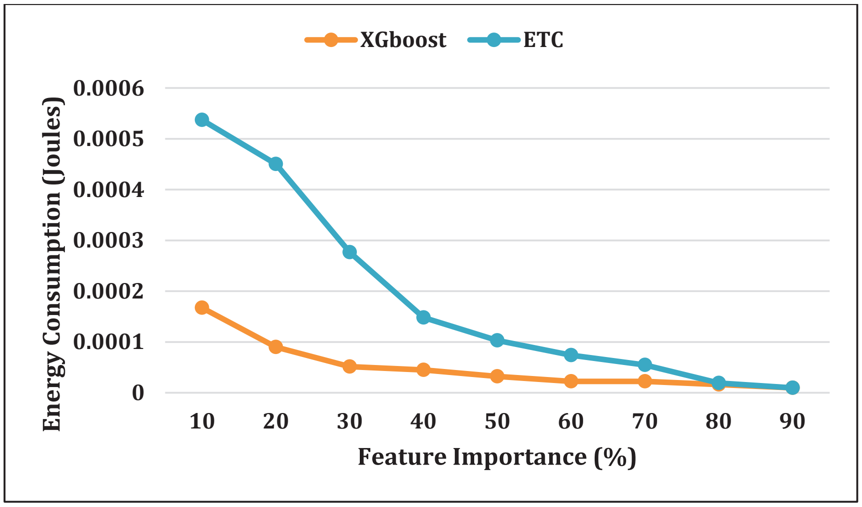

Figure 10. Energy consumption using FIM and XGBoost, ETC.

5. Conclusion and Future Works

In this paper, we propose a data reduction technique based on the feature importance method (FIM) to detect epileptic seizures from EEG recording signals based on IoMT by using two machine learning models: the extra tree classifier and the extreme gradient boosting classifier. In the proposed methodology, after applying ETC and XGBoost classifiers, the most important features are extracted to help in the analysis (classification) process by making it more easily and quickly in terms of computational speed and also increasing the degree of accuracy. In the second step, we made a comparative analysis of the results for each of FIM+ETC and FIM+XGBoost by applying performance evaluation metrics (AUC, accuracy, recall, precision, and specificity). We noticed from our results that the XGBoost classifier introduced excellent performances of the implemented methods much more than the ETC classifier, and the best results were obtained after applying the feature importance method, where the accuracy of the classifiers is clearly increased when we reduce less than 40% of the data, aside from clearly optimizing on energy consumption. But when we continue compressing the huge volume of EEG data, the accuracy of the classifiers decreases gradually, and the computation time is also reduced while the energy consumption keeps decreasing. Therefore, the balance between all these considerations is very important to maintain a high level of classification and at the same time reduce the waste of energy and the storage space of the history file of the epileptic patients, so it helps in the storage subject.

In future work, we will try to use other feature extraction strategies to improve the accuracy of predicting epileptic seizures. In addition, we are expanding this study to improve the running speed and decrease the high-dimensional feature space at the cost of a minimum loss of information. The suggested approach may be expanded to identify additional brain illnesses using EEG, such as Alzheimer's disease, autism, and dementia, and the fields of motor imagery EEG data and mental imagery task EEG data, just like it was used to detect epilepsy in real-time.

References

[1] M. K. Jabar and A. K. M. Al-Qurabat, "Human Activity Diagnosis System Based on the Internet of Things," in *Journal of Physics: Conference Series*, 2021, vol. 1879, no. 2, p. 22079.

[2] M. Iftikhar, S. A. Khan, and A. Hassan, "A survey of deep learning and traditional approaches for EEG signal processing and classification," in *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 2018, pp. 395–400.

[3] U. R. Acharya, Y. Hagiwara, and H. Adeli, "Automated seizure prediction," *Epilepsy Behav.*, vol. 88, pp. 251–261, 2018.

[4] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: a systematic review," *J. Neural Eng.*, vol. 16, no. 5, p. 51001, 2019.

[5] B. P. Prathaban, R. Balasubramanian, and R. Kalpana, "ForeSeiz: An IoMT based headband for Real-time epileptic seizure forecasting," *Expert Syst. Appl.*, vol. 188, p. 116083, 2022.

[6] P. Dhar, V. K. Garg, and M. A. Rahman, "Enhanced Feature Extraction-based CNN Approach for Epileptic Seizure Detection from EEG Signals," *J. Healthc. Eng.*, vol. 2022, 2022.

[7] L. Orosco, A. G. Correa, and E. Laciar, "Review: A survey of performance and techniques for automatic epilepsy detection," *Journal of Medical and Biological Engineering*, vol. 33, no. 6. pp. 526–537, 2013, doi: 10.5405/jmbe.1463.

[8] L. Cabanero-Gomez, R. Hervas, I. Gonzalez, and L. Rodriguez-Benitez, "eeglib: a Python module for EEG feature extraction," *SoftwareX*, vol. 15, p. 100745, 2021.

[9] R. Meier, H. Dittrich, A. Schulze-Bonhage, and A. Aertsen, "Detecting epileptic seizures in long-term human EEG: a new approach to automatic online and real-time detection and classification of polymorphic seizure patterns," *J. Clin. Neurophysiol.*, vol. 25, no. 3, pp. 119–131, 2008.

[10] G. R. Minasyan, J. B. Chatten, M. J. Chatten, and R. N. Harner, "Patient-specific early seizure detection from scalp EEG," *J. Clin. Neurophysiol. Off. Publ. Am. Electroencephalogr. Soc.*, vol. 27, no. 3, p. 163, 2010.

[11] A. G. Correa, E. Laciar, H. D. Patiño, and M. E. Valentinuzzi, "Artifact removal from EEG signals using adaptive filters in cascade," in *Journal of Physics: Conference Series*, 2007, vol. 90, no. 1, p. 12081.

[12] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Automatic seizure detection based on time-frequency analysis and artificial neural networks," *Comput. Intell. Neurosci.*, vol. 2007, 2007, doi: 10.1155/2007/80510.

[13] S. M. S. Alam and M. I. H. Bhuiyan, "Detection of seizure and epilepsy using higher order statistics in the EMD domain," *IEEE J. Biomed. Heal. Informatics*, vol. 17, no. 2, 2013, doi: 10.1109/JBHI.2012.2237409.

[14] M. Niknazar, S. R. Mousavi, B. Vosoughi Vahdat, and M. Sayyah, "A new framework based on recurrence quantification analysis for epileptic seizure detection," *IEEE J. Biomed. Heal.*

Informatics, vol. 17, no. 3, 2013, doi: 10.1109/JBHI.2013.2255132.

[15] M. Peker, B. Sen, and D. Delen, "A novel method for automated diagnosis of epilepsy using complex-valued classifiers," *IEEE J. Biomed. Heal. Informatics*, vol. 20, no. 1, 2016, doi: 10.1109/JBHI.2014.2387795.

[16] U. Orhan, M. Hekim, and M. Ozer, "EEG signals classification using the K-means clustering and a multilayer perceptron neural network model," *Expert Syst. Appl.*, vol. 38, no. 10, 2011, doi: 10.1016/j.eswa.2011.04.149.

[17] A. K. Tiwari, R. B. Pachori, V. Kanhangad, and B. K. Panigrahi, "Automated Diagnosis of Epilepsy Using Key-Point-Based Local Binary Pattern of EEG Signals," *IEEE J. Biomed. Heal. Informatics*, vol. 21, no. 4, 2017, doi: 10.1109/JBHI.2016.2589971.

[18] U. R. Acharya, F. Molinari, S. V. Sree, S. Chattopadhyay, K. H. Ng, and J. S. Suri, "Automated diagnosis of epileptic EEG using entropies," *Biomed. Signal Process. Control*, vol. 7, no. 4, 2012, doi: 10.1016/j.bspc.2011.07.007.

[19] T. W. Cenggoro, B. Mahesworo, A. Budiarto, J. Baurley, T. Suparyanto, and B. Pardamean, "Features importance in classification models for colorectal cancer cases phenotype in Indonesia," *Procedia Comput. Sci.*, vol. 157, pp. 313–320, 2019.

[20] J. Tao and Y. Kang, "Features importance analysis for emotional speech classification," in *Affective Computing and Intelligent Interaction: First International Conference, ACII 2005, Beijing, China, October 22-24, 2005. Proceedings 1*, 2005, pp. 449–457.

[21] V. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo, and M. Chica-Rivas, "Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines," *Ore Geol. Rev.*, vol. 71, pp. 804–818, 2015.

[22] R. G. Andrzejak, K. Schindler, and C. Rummel, "Nonrandomness, nonlinear dependence, and nonstationarity of electroencephalographic recordings from epilepsy patients," *Phys. Rev. E*, vol. 86, no. 4, p. 46206, 2012.

[23] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Phys. Rev. E*, vol. 64, no. 6, p. 61907, 2001.

[24] I. D. I. Saeedi and A. K. M. Al-Qurabat, "An energy-saving data aggregation method for wireless sensor networks based on the extraction of extrema points," in *Proceeding of The 1st International Conference on Advanced Research in Pure and Applied Science (Icarpas2021): Third Annual Conference of Al-Muthanna University/College of Science*, 2022, vol. 2398, no. 1, p. 050004, doi: 10.1063/5.0093971.

[25] R.-C. Chen, C. Dewi, S.-W. Huang, and R. E. Caraka, "Selecting critical features for data classification based on machine learning methods," *J. Big Data*, vol. 7, no. 1, p. 52, 2020.

[26] J. K. Jaiswal and R. Samikannu, "Application of random forest algorithm on feature subset selection and classification and regression," in *2017 world congress on computing and communication technologies (WCCCT)*, 2017, pp. 65–68.

[27] J. M. Montgomery and S. Olivella, "Tree-Based Models for Political Science Data," *Am. J. Pol. Sci.*, vol. 62, no. 3, pp. 729–744, 2018.

[28] J. Brownlee, “How to develop an extra trees ensemble with python,” *Mach. Learn. Mastery*, 2020.

[29] P. Carmona, F. Climent, and A. Momparler, “Predicting failure in the US banking sector: An extreme gradient boosting approach,” *Int. Rev. Econ. Financ.*, vol. 61, pp. 304–323, 2019.

[30] T. Chen *et al.*, “Xgboost: extreme gradient boosting,” *R Packag. version 0.4-2*, vol. 1, no. 4, pp. 1–4, 2015.

[31] R. Shafique, A. Mehmood, and G. S. Choi, “Cardiovascular disease prediction system using extra trees classifier,” 2019.