

Diagnosing Epilepsy from EEG Using Machine Learning and Welch Spectral Analysis

Esmira Abdullayeva¹, Humar Kahramanlı Örnek^{2*}

¹ Graduate School of Natural and Applied Sciences, Selcuk University, Konya 42250, Turkey

² Department of Computer Engineering, Selcuk University, Konya 42250, Turkey

Corresponding Author Email: hkahramanli@selcuk.edu.tr

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.410237>

ABSTRACT

Received: 21 May 2023

Revised: 29 August 2023

Accepted: 18 October 2023

Available online: 30 April 2024

Keywords:

Electroencephalogram (EEG), epilepsy, Welch method, Support Vector Machine (SVM), Naive Bayes (NB), Random Forest (RF), Levenberg-Marquardt (LM), Long Short Term Memory (LSTM) neural network

Epilepsy is a neurological disorder that is characterized by recurring seizures. Seizures are electrical disturbances in the brain that develop suddenly and uncontrollably. They can cause various symptoms, depending on what part of the brain is affected. The cause of epilepsy is often unknown, but it can be caused by brain injury, brain infections, genetics, or other medical conditions. EEG analysis is a very important aspect of the diagnosis and treatment of epilepsy. It includes the interpretation of electrical activity patterns recorded from the electrodes. In this study, the machine learning methods and deep learning methods have been examined for epilepsy diagnosis. Random Forest (RF), Naive Bayes (NB) algorithm, Support Vector Machine (SVM), Levenberg-Marquardt (LM), and Long Short Term Memory (LSTM) were used for classification, while the Welch method has been used for feature extraction. The Bonn EEG dataset has been used for application. As a result, the RF method showed the best accuracy as 99.87%. RF achieved 99.84% precision, 99.9% sensitivity, 99.87% F1-Score, and 99.87 AUC. LSTM achieved the second accuracy degree as 99.39%. LSTM achieved 99.52% precision, 99.29% sensitivity, 99.39% F1-Score, and 99.40 AUC. LM, SVM, and NB achieved 98.82%, 97.90%, and 97.66% classification accuracies respectively. LM achieved 97.85% precision, 99.97% sensitivity, 98.87% F1-Score, and 98.92 AUC. SVM achieved 96.10% precision, 100% sensitivity, 97.99% F1-Score, and 98.10 AUC. NB achieved 98.80% precision, 96.42% sensitivity, 97.27% F1-Score, and 97.61 AUC.

1. INTRODUCTION

Electroencephalography (EEG) is a method that measures potentials reflecting brain wave activity. EEG, a widely utilized measurement tool, offers reliable insights into brain functions, abnormalities, and neurophysiological dynamics because it is low-cost, portable, and has a high temporal resolution. EEG is frequently used by doctors to study brain function and diagnose neurological disorders. Researchers examine brain functions such as vision, memory, motor imagery, intelligence, perception, emotion, and recognition through EEG recordings. In addition, EEG is one of the important tools for the diagnosis of neurological diseases such as sleep disorders, epilepsy, dementia, brain tumors, head trauma, and for monitoring the depth of anesthesia during surgery. It is also worth noting that abnormalities, and behavioral disorders (eg., autism), are also useful in the treatment of learning problems, attention disorders, and language delays.

An epileptic seizure refers to a temporary manifestation of signs and/or symptoms caused by abnormal and excessive neuronal activity in the brain, occurring in a synchronous manner [1]. Seizures are defined as sudden changes in the electrical functioning of the brain and cause changes in behaviors such as jerky movements, loss of consciousness, temporary loss of breath, and memory loss [2]. Epilepsy, one

of the most common neurological diseases experienced worldwide, is a brain disorder characterized by a persistent predisposition to produce epileptic seizures and the neurobiological, psychological, cognitive, and social consequences of this condition. To define epilepsy it needs to occur at least one epileptic seizure [1]. Epilepsy patients suffer from the psychological, physical, and social consequences of this disease [3]. Illness can sometimes cause patients to need one even in their daily lives. Due to the variable and unpredictable nature of the disease, it is almost impossible to take precautions against it. The World Health Organization (WHO) estimates that about fifty million people worldwide are affected by epilepsy. Also according to the WHO, 70% of epilepsy sufferers might avoid seizures if their condition was adequately identified and treated. In recent years Machine learning helps doctors to deal with this issue. Machine learning algorithms are widely used to diagnose epileptic seizures. In this regard, some research demonstrating the use of machine learning in EEG analysis was presented. Guler et al. [4] used obtained statistics from the Lyapunov exponent are used as features. The statistics used were: the mean of the absolute values of the Lyapunov exponents, the largest of the absolute values of the Lyapunov exponents, the mean power of the Lyapunov exponents, the standard deviation of the absolute values of the Lyapunov exponents. Recurrent Neural Networks (RNN) and Levenberg-Marquardt Backpropagation

Neural Networks were used as classifiers. The results showed that RNN was more successful. Polat and Güneş [5] presented a study on the diagnosis of epilepsy. The study has two stages: feature extraction using FFT and classification. They used a decision tree (DT) as a classifier. The validation of the system was conducted using 10-fold cross-validation, classification accuracy, specificity and sensitivity values. They achieved 98.72% classification accuracy. Wang et al. [6]'s study consists of three stages. In the first step, features are extracted by applying wavelet transform. In the second stage, classification was carried out. For this, the k-Nearest Neighbors (k-NN) method is used. For validation purposes, k-fold cross-validation was applied. Finally, classification successes were calculated. The results showed that 2, 5, and 10 times cross-validation achieved successful results. Ahammad et al. [7] presented methods for automatically detecting the event and onset of epileptic seizures in their study. A linear classifier was employed for classifying normal and epileptic EEG signals. Three types of EEG signals were classified: EEG signals recorded from a healthy volunteer with eyes open, epilepsy patients in the epileptogenic zone during a seizure-free period, and patients during epileptic seizures. The classifier's performance was evaluated based on sensitivity, specificity, and accuracy. The overall accuracy achieved was 84.2%. Acharya et al. [8] used Z-score normalization, standard deviation, and zero means for data preprocessing. CNN, a deep learning algorithm, was used for classification. The CNN used consists of 13 layers. These layers are ordered as a convolution and a pooling layer, and the last layer is the fully connected layer. ReLU and softmax are used as activation functions. It has been shown that the results obtained as a result of 10-time cross-validation have an important place in the literature. 88.67% accuracy achieved as accuracy. Zhang et al. [9] studied data collected at Boston Children's Hospital in their study. This dataset consisted of EEG recordings collected from 23 children. EEG recordings were made according to the 10-20 system and sampled at 256 Hz. First, the original EEG recordings were decomposed into five sub-frequency bands that fit the Gaussian distribution. Then it was classified with the VGGNet network. SVM, MLP, ELM, and Long Short-Term Memory (LSTM) are used to compare performance. As a result, it has been shown that the proposed method can classify with 95.12% success. In Zeng et al.'s study [10] the signals were first divided into 0.58-second windows, and then the gray recurrence graph of the segments was obtained. These graphs were labeled and given to the DenseNet network. They stated that the developed system can predict 100% success. Tuncer [11]'s study includes different steps from the others. The Nonlinear Hamsi-Pat method was used for feature extraction in the study. The Hamsi-Pat method uses the S-box feature of the Hamsi hash function. Feature extraction adjustable Q-factor wavelet transforms and iterative neighbor component analysis (INCA), k-NN is used for classification. It has been shown that 99.20% of classification success is achieved. Al-jumali et al. [12] investigated various feature extraction and classification methods for seizure classification. They used data from Temple University Hospital Seizure Corpus for this aim. In the first stage of the study, they used FFT. For classification, they used SVM, NB, DT, and K-Nearest Neighbor (KNN) classifier. They stated that SVM achieved the best result. Janga and Edara [13] proposed an integrated framework for epilepsy detection, which involves exploring EEG signals using a combination of Multi-class SVM and the Improved Chaotic Firefly algorithm. For feature

extraction, they used Discrete Wavelet Transform (DWT). The proposed method achieved 99.63% classification accuracy. Ahmed et al. [14] proposed a machine learning-based ensemble learning technique in this study to predict epileptic seizures. In the first stage, they used Power line noise reduction. As a classifier, they used DT, SVM, Artificial Neural Network (ANN), and CNN. They used the PhysioNet dataset for the application aim. They achieved 91% classification accuracy. It can be observed from literature the many researches preferred Bonn EEG dataset for application aim [4-8, 10-11, 13].

In this study, Welch method has been examined as a feature selection and the machine learning methods have been examined for epilepsy diagnosis. Naive Bayes (NB) algorithm, SVM, RF, LM, and LSTM were used for classification. The results show that Welch method is successful as preprocessing. The rest of this paper is organized as follows. In the second chapter background theory are presented. In the third chapter, the results of the experiments are presented. In the final chapter, this paper is concluded.

2. MATERIAL AND METHODS

2.1 Material

In this study, the Bonn EEG dataset [15] was used in the application stage. Each of the five subgroups, A through E, in the dataset, has 100 single-channel EEG segments that each last 23.6 seconds. Groups A and B consisted of sections from surface EEG recordings performed on five volunteers. All volunteers are healthy. The segments in A are taken with eyes open while the segments in B are with their eyes closed. A 10-20 electrode placement scheme was used. The segments in C are taken from five epileptic patients, during seizure-free intervals, while segments in D are from the hippocampal formation of the opposite brain hemisphere of these patients. EEG segments in S contain seizure activity. EEG segments in C, D, and E are from depth electrodes implanted symmetrically into the hippocampal formation. 128-channel amplifier system was used to record all EEG signals. The signals were digitized at 173.6 Hz with a 12-bit resolution. Consequently, the sample length of each segment is $173.6 \times 23.6 \approx 4097$ and the corresponding bandwidth is 86.8 Hz.

2.2 Used methods

In this investigation, the Welch method was used for data preprocessing and five machine learning methods: NB algorithm, RF, SVM, LM, and LSTM have used to build binary classification models.

2.2.1 Welch algorithm

The Welch algorithm is a nonparametric method used to estimate the Power Spectral Density. The Welch algorithm is utilized to obtain a smoother frequency spectrum compared to the raw FFT output. In this algorithm, the signal is divided into windows of equal size. The size of the window impacts the clarity of the result by filtering out frequencies with periods longer than the window size [16]. After the data is divided into overlapping segments, a time domain window is applied to the individual data segments. The Welch method is performed by dividing the time signal into successive overlapping segments, constructing the periodogram for each block, and averaging,

which reduces the variance of individual power measurements [17].

2.2.2 SVM

SVM are supervised learning models has become one of the most popular machine learning techniques over time [18]. SVM is used in machine learning to examine data for regression and classification. The SVM produces a model which categorizes samples into one of two categories based on a collection of training samples, making it a non-probabilistic binary linear classifier. In the context of classification, new data points are mapped into the same feature space, and their class labels are predicted based on which side of the hyperplane they fall on. Due to the usage of the kernel method, SVMs can perform effectively even when the data is very high dimensional or has a large number of features. When there is a significant imbalance between the classes—that is, when one class contains significantly more samples than the other—SVMs can still be useful. Another benefit of SVMs is that, especially when regularization is used, they may be less prone to overfitting than some other types of models. SVMs can be employed with a variety of kernel functions, including linear, Radial Basis, and polynomial kernels. The performance of the model can be significantly impacted by kernel selection. The selection of the kernel function and model parameters plays a crucial role in determining the performance of SVR [19]. The radial basis kernel function was used in this study.

2.2.3 Naive Bayes classifier

The NB is one of the most often used algorithms for classification issues due to its simplicity, efficiency, and robustness [20]. NB algorithm, which is developed based on the Bayesian theorem, is a straightforward probabilistic classifier that places each object in a class under the assumption that variables are independent of one another. The training process of the NB classifier is to estimate the class prior probability of samples and the conditional probability for each feature based on the training set. The advantage of using NB is its ability to make accurate predictions with limited training data. The posterior distribution of an instance is proportional to the prior distribution and likelihood, according to Bayes' theorem. The formula for NB is:

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

where, C_k refers to a certain class and X represents metric values.

2.2.4 Random forests

RF is a frequently preferred ensemble learning algorithm that can be used for regression and classification tasks. As the name suggests, this algorithm creates a forest with several trees [21]. It trains multiple decision trees on random subsets of data and then averages their predictions. One of the key advantages of RF is that they tend to be very accurate and robust, even when compared to more complex models. They also have the advantage of being relatively fast to train and easy to use, making them a good choice for many applications. Another advantage of RF is that it can handle high-dimensional and categorical data well, and it can automatically detect and handle missing values. For classification problems, the output of the random forest is determined by the majority class chosen by the ensemble of trees.

2.2.5 LM algorithm

The Levenberg-Marquardt algorithm (LM) is a second-order training method for feedforward neural networks [22]. It is a type of quasi-Newton algorithm, which means that it uses an approximation of the Hessian matrix to compute the search direction at each iteration. The Levenberg-Marquardt algorithm is often used in machine learning to train neural networks and other types of models. It is generally considered to be more efficient and robust than gradient descent, particularly in cases where the cost function is very ill-conditioned. However, one potential drawback of the Levenberg-Marquardt algorithm is that it can be sensitive to the choice of initial conditions, so it's important to choose good initial values for the parameters being optimized.

2.2.6 Long Short-Term Memory

Recurrent Neural Network (RNN) is a type of neural network that is well-suited for modeling sequences of data. They are called "recurrent" because they make use of sequential information, by performing the same computation for every element in a sequence and using the output of that computation as input for the next element in the sequence. LSTM is a type of RNN that is well-suited for modeling sequential data. It was introduced in 1997 by Sepp Hochreiter and Jürgen Schmidhuber. One of the key features of LSTMs is that they can remember information for long periods, thanks to the use of gating mechanisms that control the flow of information into and out of the memory cells. This makes LSTMs well-suited for tasks such as language translation and language modeling, where the context of previous words is important in understanding the meaning of the current word.

LSTMs are composed of memory cells, which are responsible for storing information, and three different types of gates: input, forget and output gates [23]. The input gate controls which information from the current input is stored in the memory cell. The forget gate controls which information from the previous memory state is discarded or retained in the current memory state. The output gate controls which information from the current memory state is outputted as the prediction. The gates are all controlled by weights, which are learned during training. At each time step, the LSTM takes in an input and the previous memory state, and it outputs a prediction and the updated memory state.

2.2.7 Metrics

In this study, a confusion matrix and evaluation metrics were employed to measure the success of the classification. Therefore, a complexity matrix was created for each class based on the definition of actual values and predicted values in Table 1. Each column of the confusion matrix represents the instances in a predicted class while each row represents the instances in an actual class (or vice versa). In the context of binary classification, a true positive (TP) and a true negative (TN) are outcomes where the model correctly predicts the positive and the negative classes, respectively. In a medical test for a disease, a true positive would be when the test correctly identifies a patient with the disease). A false positive (FP) and a false negative (FN) are outcomes where the model incorrectly predicts the positive and the negative classes respectively. TP, TN, FP, and FN are used together to calculate evaluation metrics such as accuracy, sensitivity, precision, Area Under the Curve (AUC), and F1-Score.

Table 1. Confusion matrix definition

Reference	Predictions	
	0	1
0	TP	FN
1	FP	TN

A confusion matrix was used to measure the performance of each class, such as Accuracy, Sensitivity, Precision, F-Score, and AUC. The formulas of these measurement metrics are shown in Table 2.

Table 2. Classification evaluation metrics

Metric	Formula
Accuracy	$Acc = \frac{tp+tn}{tp+tn+fp+fn}$
Precision	$P = \frac{tp}{tp+fp}$
Sensitivity	$S = \frac{tp}{tp+fn}$
F1-Score	$F1 - score = \frac{2\left(\frac{tp}{tp+fn}\right)\left(\frac{tp}{tp+fp}\right)}{\left(\frac{tp}{tp+fn}\right) + \left(\frac{tp}{tp+fp}\right)}$
AUC	$AUC = \frac{1}{2} \left(\frac{tp}{tp+fn} + \frac{tp}{tp+fp} \right)$

3. RESULT AND DISCUSSION

The application part consists of 4 stages: segmentation, feature extraction, classification, and evaluation of results. The implementation steps of the study are given in Figure 1. First, each segment is 4096 long, divided into windows of 256 samples. The amount of overlap was determined as 128. As a result, 31 sub-segments were created from each segment. In the next step, the Welch method was applied to the sub-segments obtained. As a result of Welch, 129 features were extracted from each sub-segment. For application, A and E clusters were used. Each of these clusters consists of 100 data. Since 31 sub-segments were produced from each segment, a total of 6200 data were obtained from 200 segments. 3130 of these 6200 segments belong to the healthy class, and 3070 of them belong to the diseased class. The healthy class was assigned a label of 1 and the diseased class was assigned a label of 0. As a result, 3130 and 3070 segments were on class 1 and 0, respectively. 6200 data is divided into 10 clusters as 10-fold cross-validation will be applied. In the third stage, the classification process was carried out. For this, a total of 5 algorithms, namely SVM, NB, RF, LM, and LSTM, were used. For SVM polynomial, linear, and radial basis function kernels were tested. The radial basis kernel function produced the best result. For LM linear, sigmoid, and hyperbolic tangent activation functions were tested. The number of neurons in the hidden layer was set as 2-15. The network produced the best result with two neurons in the hidden layer. Sigmoid and hyperbolic tangent functions were used as activation functions in hidden and output layers respectively. In the fourth stage, the results obtained as a result of the classification were evaluated. Tables 3-7 show the confusion matrixes of classifiers. For LSTM relu, linear and softmax layers, and adam optimizer produced the best result.

In Table 8, the accuracy, sensitivity, precision, f1-score, AUC, and average result values of the SVM algorithm are given. Table 8 shows a balanced success rate when looking at the folds. 97.90% success was achieved as an average result. The classification accuracy of the SVM algorithm is 97.42% in the first fold, 96.61% in the second fold, 97.75% in the third

fold, 99.03% in the fourth fold, 99.84% in the fifth fold, 96.45% in the sixth fold, 95.32% in the seventh fold, 97.09% in the ninth fold, and 99.52% in the 10th fold. The highest classification accuracy was obtained in the 8th fold as 100%. Upon examining Table 8, it can be observed that sensitivity is 100% for all folds. It shows that FN is 0 for all folds.

Table 3. The confusion matrix of SVM classifier

	0	1
0	2940	130
1	0	3130

Table 4. The confusion matrix of RF classifier

	0	1
0	3066	4
1	6	3124

Table 5. The confusion matrix of NB classifier

	0	1
0	3031	39
1	112	3018

Table 6. The confusion matrix of LM classifier

	0	1
0	2998	72
1	1	3129

Table 7. The confusion matrix of LSTM classifier

	0	1
0	3047	15
1	23	3115

Table 8. Accuracy, precision, sensitivity, f1-score, AUC, and average result values of SVM algorithm

Folds	Accuracy	Precision	Sensitivity	F1-Score	AUC
1	97.42	95.13	100	97.50	97.57
2	96.61	93.72	100	96.76	96.86
3	97.75	95.72	100	97.81	97.86
4	99.03	98.11	100	99.05	99.06
5	99.84	99.69	100	99.84	99.85
6	96.45	93.44	100	96.61	96.72
7	95.32	91.53	100	95.58	95.77
8	100	100	100	100	100
9	97.09	94.57	100	97.21	97.29
10	99.52	99.05	100	99.52	99.53
Average Result	97.90	96.10	100	97.99	98.10

The classification accuracy of the RF algorithm was equal to 100% in the 2nd, 3rd, 4th, 5th, 6th, and 7th folds, as the highest results. Equal results as 99.84% were obtained in the 1st and 10th folds. The algorithm reached an accuracy rate of 99.52% in the 8th fold and 99.35% in the 9th fold. The accuracy, sensitivity, precision, f1-score, AUC, and average result values of the RF algorithm are shown in Table 9. Considering the outcomes in Table 9, it is seen that the same results are obtained in almost all of the clusters. On average, 99.87% accuracy was achieved. Upon examining Table 9, it can be observed that sensitivity and precision are 100% almost for all folds.

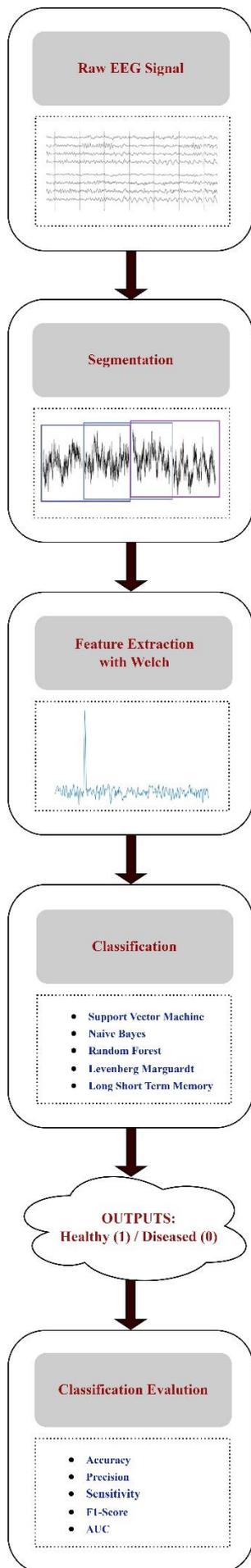


Figure 1. Implementation steps of the study

The classification accuracy of the NB algorithm was equal to 99.68% in the 1st and 6th folds. 99.52% in the 2nd fold, 98.71% in the 3rd fold, 99.22% in the 7th fold, and 96.94% in the 9th fold. Equal results were obtained as 99.84% in the 4th, 5th, and 10th folds. The lowest result was 83.37% in the 8th fold. In Table 10, the accuracy, precision, sensitivity, f1-score, AUC, and average result values of the NB algorithm are given. As seen in Table 10, although it produced almost equal results in 9 folds, a low result was achieved in one fold. The average result was 97.66% accuracy.

Table 9. Accuracy, sensitivity, precision, 1-score, AUC, and average result values of RF algorithm

Folds	Accuracy	Precision	Sensitivity	F1-Score	AUC
1	99.84	99.69	100	99.85	99.85
2	100	100	100	100	100
3	100	100	100	100	100
4	100	100	100	100	100
5	100	100	100	100	100
6	100	100	100	100	100
7	100	100	100	100	100
8	100	100	100	100	100
9	99.52	100	99.04	99.52	99.52
10	99.35	98.74	100	99.37	99.37
Average Result	99.87	99.84	99.9	99.87	99.87

Table 10. Accuracy, sensitivity, precision, f1-score, AUC, and average result values of NB algorithm

Folds	Accuracy	Precision	Sensitivity	F1-Score	AUC
1	99.68	99.36	100	99.68	99.68
2	99.52	99.05	100	99.52	99.53
3	98.71	97.51	100	98.74	98.76
4	99.84	99.69	100	99.84	99.85
5	99.84	100	99.68	99.84	99.84
6	99.68	99.37	100	99.68	99.69
7	99.22	98.71	97.77	98.24	98.24
8	83.37	100	67.09	80.30	80.55
9	96.94	94.27	100	97.05	97.14
10	99.84	100	99.69	99.84	99.85
Average Result	97.66	98.80	96.42	97.27	97.61

Table 11. Accuracy, sensitivity, precision, f1-score, AUC, and average result values of LM algorithm

Folds	Accuracy	Precision	Sensitivity	F1-Score	AUC
1	99.84	99.69	100	99.84	99.85
2	94.19	89.69	100	94.56	94.85
3	100	100	100	100	100
4	99.84	99.69	100	99.84	99.85
5	99.84	99.69	100	99.84	99.85
6	99.52	99.05	100	99.52	99.53
7	97.91	96.29	99.69	97.96	97.99
8	99.84	99.69	100	99.84	99.85
9	97.91	96.01	100	97.96	98.01
10	99.35	98.74	100	99.37	99.37
Average Result	98.82	97.85	99.97	98.87	98.92

As for the classification accuracy of the LM algorithm, 99.84% of results were obtained in the 1st, 4th, 5th, and 8th folds. Results were obtained as 94.19% in the 2nd fold, 99.52% in the 6th fold, and 99.35% in the 10th fold. 97.91% of results were obtained in the 7th and 9th folds. The highest classification accuracy was obtained in the 3rd fold at 100%. In Table 11, the accuracy, sensitivity, precision, f1-score, AUC, and average result values of the LM algorithm are given.

AUC, and average result values of the LM algorithm are given. The average result was 98.82% accuracy.

The classification accuracy of the LSTM network was equal to 99.84% in the 1st fold and the 6th fold. 98.06% in the 2nd fold, 99.19% in the 7th fold, 99.87% in the 8th fold, and 99.35% in the 9th fold. In clusters 3, 4, 5, and 10, the accuracy of the neural network was equal to 100%. The accuracy, sensitivity, precision, f1-score, AUC, and average result values of the LSTM network are shown in Table 12. The average result was 99.39%.

Table 12. Accuracy, precision, sensitivity, f1-score, AUC, and average result values of LSTM algorithm

Folds	Accuracy	Precision	Sensitivity	F1-Score	AUC
1	99.84	100	99.69	99.84	99.85
2	98.09	100	96.30	98.12	98.15
3	100	100	100	100	100
4	100	100	100	100	100
5	100	100	100	100	100
6	99.84	100	99.69	99.84	99.85
7	99.19	100	98.43	99.21	99.22
8	97.58	95.20	100	97.54	97.60
9	99.35	100	98.74	99.37	99.37
10	100	100	100	100	100
Average Result	99.39	99.52	99.29	99.39	99.40

The average results of the classifiers used are shown in Table 13. Figures 2 and 3 show the comparative accuracy and AUC results of algorithms.

Table 13. The average results of used algorithms

	Accuracy	Precision	Sensitivity	F1-Score	AUC
SVM	97.90	96.10	100	97.99	98.10
RF	99.87	99.84	99.9	99.87	99.87
NB	97.66	98.80	96.42	97.27	97.61
LM	98.82	97.85	99.97	98.87	98.92
LSTM	99.39	99.52	99.29	99.39	99.40

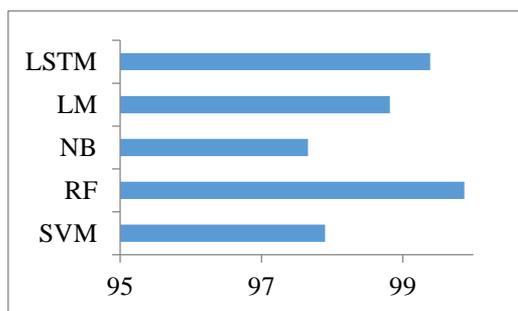


Figure 2. The average accuracy results of used algorithms

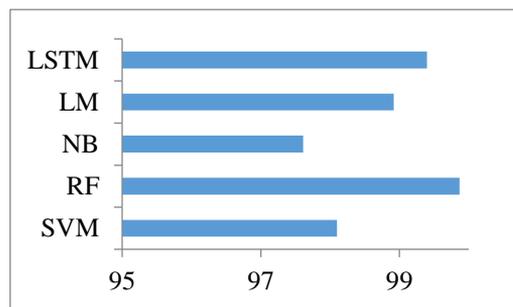


Figure 3. The average AUC results of used algorithms

The results show that the presented method achieves high success. Using different feature extraction methods may achieve the best result.

4. CONCLUSIONS

Epilepsy is a neurological disorder characterized by recurrent seizures caused by abnormal electrical activity in the brain. Epilepsy often begins in childhood or adolescence but can occur at any age. The diagnosis of epilepsy typically involves a neurological examination, the patient describing their symptoms and the frequency of seizures, and tests such as EEG to record the brain's electrical activity.

In recent years, there has been tremendous growth in the use of machine learning in medicine. In recent years, there has been tremendous growth in the use of machine learning in medicine. Artificial Intelligence especially machine learning widely uses in research and decision support in healthcare. In this study, Welch method has been examined as a feature selection for epilepsy diagnosis. For the classification part, several machine learning methods and deep learning methods have been examined. These classification methods are Naive Bayes algorithm, SVM, Random Forest, LM, and Long Short Term Memory. In the application stage, Bonn EEG dataset has been used. 200 single-channel EEG segments that each last 23.6 seconds were used. The application part consists of 4 stages: segmentation, feature extraction, classification, and evaluation of results. After segmentation, 6200 sub-segments have been formulated. Welch method has produced 129 features from each sub-segments. For classification, a total of 5 algorithms, namely SVM, NB, RF, LM, and LSTM, were used. As a result, it can be observed that all methods show successful results, while the best accuracy is 99.87% and the worst accuracy is 97.66%. The best result has been achieved by RF with 99.87% classification accuracy. LSTM, LM, SVM, and NB achieved 99.39%, 98.82%, 97.90%, and 97.66% classification accuracies respectively. The best precision was achieved by RF with 99.84%. LSTM, NB, LM, and SVM achieved 99.52%, 98.80%, 97.85%, and 96.10% precision respectively. SVM reaches the best sensitivity as 100%. RF has reached too close sensitivity to SVM as 99.9%. LM, LSTM, and NB achieved 99.97%, 99.29%, and 96.42% sensitivity respectively. F1-score and AUC results show that RF is best. As a result, it can be observed that RF and LSTM are very successful in diagnosing epilepsy. The result demonstrates that the presented approach holds great promise to be applied in clinical applications to epilepsy diagnosing.

REFERENCES

- [1] Fisher, R.S., Boas, W.E., Blume, W., Elger C., Genton, P., Lee, P., Engel, J. (2005). Epileptic seizures and epilepsy: Definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). *Epilepsia*, 46(4): 470-472. <https://doi.org/10.1111/j.0013-9580.2005.66104.x>
- [2] Siuly, S., Bajaj, V., Sengur, A., Zhang, Y. (2019). An advanced analysis system for identifying alcoholic brain state through EEG signals. *International Journal of Automation and Computing*, 16(6): 737-747. <https://doi.org/10.1007/s11633-019-1178-7>
- [3] Guo, L., Rivero, D., Dorado, J., Rabuñal, J.R., Pazos, A.

- (2010). Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks. *Journal of Neuroscience Methods*, 191(1): 101-109. <https://doi.org/10.1016/j.jneumeth.2010.05.020>
- [4] Guler, N.F., Ubeyli, E.D., Guler, İ. (2005). Recurrent neural networks employing Lyapunov exponents for EEG signals classification. *Expert Systems with Applications*, 29: 506-514. <https://doi.org/10.1016/j.eswa.2005.04.011>
- [5] Polat, K., Gunes, S. (2007). Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. *Applied Mathematics and Computation*, 187(2): 1017-1026. <https://doi.org/10.1016/j.amc.2006.09.022>
- [6] Wang, D., Miao, D., Xie, C. (2011). Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection. *Expert Systems with Applications*, 38: 14314-14320. <https://doi.org/10.1016/j.eswa.2011.05.096>
- [7] Ahammad, N., Fathima, T., Joseph, P. (2014). Detection of epileptic seizure event and onset using EEG. *BioMed Research International*, 2014: 1-7. <https://doi.org/10.1155/2014/450573>
- [8] Acharya, U.R., Oh, S.L., Hagiwara, Y., Tan, J.H., Adeli, H. (2018). Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Computers in Biology and Medicine*, 100: 270-278. <https://doi.org/10.1016/j.compbio.2017.09.017>
- [9] Zhang, J., Wei, Z., Zou, J., Fu, H. (2020). Automatic epileptic EEG classification based on differential entropy and attention model. *Engineering Applications of Artificial Intelligence*, 96: 103975. <https://doi.org/10.1016/j.engappai.2020.103975>
- [10] Zeng, M., Zhang, X., Zhao, C., Lu, X., Meng, Q. (2021). GRP-DNet: A gray recurrence plot based densely connected convolutional network for classification of epileptiform EEG. *Journal of Neuroscience Methods*, 347: 108953. <https://doi.org/10.1016/j.jneumeth.2020.108953>
- [11] Tuncer, T. (2021). A new stable nonlinear textural feature extraction method based EEG signal classification method using substitution Box of the Hamsi hash function: Hamsi pattern. *Applied Acoustics*, 172: 107607. <https://doi.org/10.1016/j.apacoust.2020.107607>
- [12] Al-jumali, S., Duru, A.D., Ibrahim, A.A., Uçan, A.N. (2023). Investigation of epileptic seizure signatures classification in EEG using supervised machine learning algorithms. *Traitement du Signal*, 40(1): 43-54. <https://doi.org/10.18280/ts.400104>
- [13] Janga, V., Edara, S.R. (2021). Epilepsy and seizure detection using JLTM based ICFFA and multiclass SVM classifier. *Traitement du Signal*, 38(3): 883-893. <https://doi.org/10.18280/ts.380335>
- [14] Ahmed, M.I.B., Zaghdoud, R.A., Al-Abdulqader, M., Kurdi, M., Altamimi, R., Alshammari, A., Noaman, A., Ahmed, M.S., Alshamrani, R., Alkharraa, M., Rahman, A., Krishnasamy, G. (2023). Ensemble machine learning based identification of adult epilepsy. *Mathematical Modelling of Engineering Problems*, 10(1): 84-92. <https://doi.org/10.18280/mmep.100110>
- [15] Andrzejak, R.G., Lehnertz, K., Rieke, C., Mormann, F., David, P., Elger, C.E. (2001). Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state, *Physical Review E*, 64: 061907. <https://doi.org/10.1103/PhysRevE.64.061907>
- [16] Same, M.H., Gandubert, G., Gleeton, G., Ivanov, P., Landry, R. (2021). Simplified welch algorithm for spectrum monitoring. *Applied Sciences*, 11: 86. <https://doi.org/10.3390/app11010086>
- [17] Jwo, D.J., Chang, W.Y., Wu, I.H. (2021). Windowing techniques, the welch method for improvement of power spectrum estimation. *Computers, Materials & Continua*, 67(3): 3983-4003. <https://doi.org/10.32604/cmc.2021.014752>
- [18] Yilmaz, B.Y., Taspinar, Y.S., Koklu, M. (2022) Classification of malicious android applications using naive bayes and support vector machine algorithms. *International Journal of Intelligent Systems and Applications in Engineering*, 10(2): 269-274.
- [19] Tezel, G., Buyukyildiz, M. (2016). Monthly evaporation forecasting using artificial neural networks and support vector machines. *Theoretical and Applied Climatology*, 124: 69-80. <https://doi.org/10.1007/s00704-015-1392-3>
- [20] Arar, Ö.F., Ayan, K. (2017). A feature dependent Naive Bayes approach and its application to the software defect prediction problem. *Applied Soft Computing*, 59: 197-209. <https://doi.org/10.1016/j.asoc.2017.05.043>
- [21] Garai, D., Agrawal, H., Mishra, A.K., Kumar, S. (2018). Influence of initiation system on blast-induced ground vibration using random forest algorithm, artificial neural network, and scaled distance analysis. *Mathematical Modelling of Engineering Problems*, 5(4): 418-426. <https://doi.org/10.18280/mmep.050419>
- [22] Bilski, J., Kowalczyk, B., Marchlewska, A., Zurada, J.M. (2020). Local Levenberg-Marquardt algorithm for learning feedforward neural networks. *Journal of Artificial Intelligence and Soft Computing Research*, 10(4): 299-316. <https://doi.org/10.2478/jaiscr-2020-0020>
- [23] Lu, W., Li, J., Li, Y., Sun, A., Wang, J. (2020). A CNN-LSTM-based model to forecast stock prices. *Complexity*, 2020: 6622927. <https://doi.org/10.1155/2020/6622927>