SAKARYA UNIVERSITY JOURNAL OF COMPUTER AND INFORMATION SCIENCES
http://saucis.sakarya.edu.tr/
e-ISSN: 2636-8129
Publisher: Sakarya University
Vol. 7, No. 2, 2024, 138-155
DOI: 10.35377/saucis...1436915

RESEARCH ARTICLE

The Importance of Rhythm Activity in Epilepsy EEG Signal Classification (An Educational Article)

Negin Melek[1]

[1] Department of Computer Engineering, Faculty of Engineering, Giresun University, Giresun, Türkiye

Corresponding author:
Negin Melek, Department of Computer Engineering, Giresun University, Giresun, Türkiye
negin.melek@giresun.edu.tr

Article History:
Received: 14.02.2024
Accepted: 30.05.2024
Published Online: 23.08.2024

ABSTRACT

Electroencephalography (EEG), used to record the random electrical activity in brain, is a known medical test. In this test, a graphical waveform is obtained by measuring the electrical activity of the cells. In the medical world, the relationship between epilepsy and EEG can be understood by examining changes in brain activity during or between epileptic seizures. EEG is a useful tool in the early treatment and diagnosis of epilepsy. Whether seizures, generally known as abnormal electrical discharges in brain cells, are of epileptic origin, comes to light through EEG. The main goal of our study was to demonstrate the EEG rhythm effectiveness for the diagnosis of epilepsy in EEG data obtained from the epilepsy center of Bonn Freiburg University Hospital. Time domain feature extraction of EEG band classification results was examined in detail against the classification results of frequency domain feature extraction of EEG rhythms in healthy subjects and subjects with epilepsy. By extracting effective features from EEG data in both time and frequency domains, the k nearest neighbor (KNN) algorithm was used for the time and frequency domain. It cannot be overlooked that among the four methods used for performance evaluation in the designed model, the classification success of frequency domain features was more successful than that of time domain features. Using the KNN algorithm, healthy individuals and epilepsy patients with seizures were classified with 100% success.

Keywords: EEG, Epilepsy, Rhythm, Feature extraction, Classification

1. Introduction

Electroencephalography (EEG) resembles a melody capturing a graceful dance of brain waves, a window that reveals the mysterious rhythm of the mental world. EEG contains useful information about the random nature and dynamics of brain waves and plays a major role in disease diagnosis. From a general perspective, the scientific application areas of EEG can be divided into three categories: entertainment, engineering, and medical [1]. Figure 1 shows the percentage distribution of EEG application areas according to the research conducted in 2020. The entertainment application field includes examples such as brain-eye combinatorial controls, brain-based control and robotic games, car racing, social interaction, and virtual marketing [2]. Most EEG-based studies in engineering can be divided into two subgroups: biometrics and brain-computer interface (BCI) [3]. The process of classifying individuals based on their behavioral and physiological characteristics is defined as biometrics. One of the extraordinary features offered by EEG signals is their typical biometric identification for each individual [4]. Thus, thanks to this uniqueness, we can say that each person has a unique neural signature [5]. BCI, another important application area of EEG in engineering, is a technology that enables direct communication between the brain and an external device such as a computer. Focusing on BCI technology, great efforts continue to be made to increase EEG signal quality and communication speed, as well as to improve model accuracy. Research in this field represents a broad spectrum, creating major revolutions from healthcare to entertainment [6].

From a medical and healthcare perspective, detailed analysis of brain signal activities provides great convenience in predicting diseases and abnormalities in neurology. In this context, EEG is an important tool for understanding brain functions, learning, and memory processes.

Quantitative EEG (QEEG) was used to diagnose and predict Parkinson's disease dementia in 2011 [7]. This study proved that θ band relative power analyses are a determining factor in the incidence of dementia.

Another study showed that the independent component analysis (ICA) preprocessing method increases the performance in developing an Alzheimer's disease detection system with automatic brain waves [8].

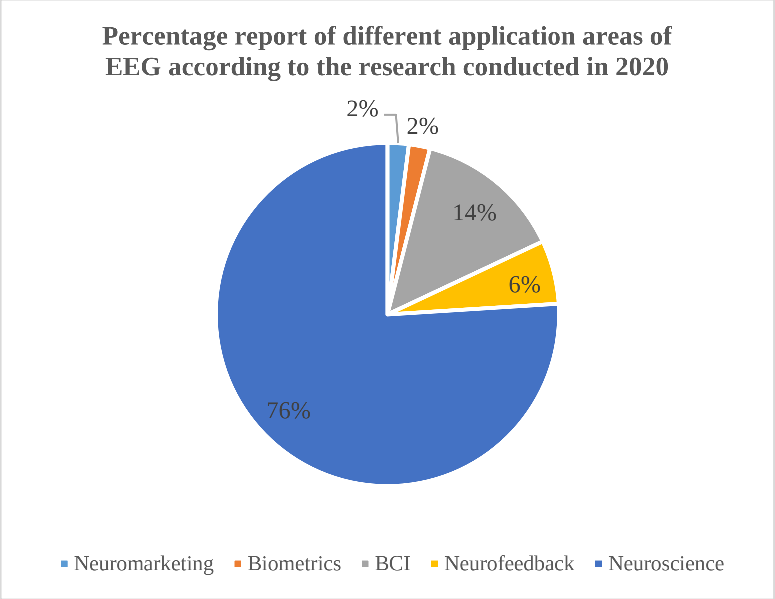


Figure 1. Percentage Distribution of EEG Application Areas According to the Research Conducted in 2020 [2]

EEG is not a direct method for treating language disorders and dyslexia, but it is useful for improving reading and writing skills by discovering abnormalities in the individual's neurological processes [9]. In addition, EEG is used as an auxiliary tool in treating language disorders, along with neurofeedback and auditory therapy methods.

Attention-deficit/hyperactivity disorder (ADHD) [10], a common psychiatric disorder of our age, is frequently observed in children. As a result of the valuable information contained in EEG bands, theta/beta ratio (TBR) is a frequently used measure in ADHD studies [11].

EEG plays an important role in detecting seizures due to abnormal electrical activities of neurons. These unwanted activities can be discovered and characterized using EEG waves [12]. There are many types of seizures with different electrical and wave patterns. EEG categorizes seizure types according to the characteristics of these patterns [13]. This application area of EEG is of great importance in medicine because it has serious benefits in the early and rapid management of individuals with seizures. It also enables healthcare experts to make informed and more precise decisions about early treatment and diagnosis of patients with seizures [14].

Early diagnosis and conscious support can increase the standards and quality of life of individuals with symptoms of autism spectrum disorder. In this context, the debate on whether the findings of research on brain waves and autism spectrum disorder are related to each other is still ongoing [15]. It has been proven that EEG signals taken from preschool children show signs of high-grade autism spectrum disorder. On the other hand, it was concluded that these symptoms are more pronounced while asleep than when awake [16].

In addition, many scientific studies have touched upon other different common and important application areas of EEG, such as insomnia detection [17] and treatment method determination, brain damage analysis and detection [18], consciousness assessment [19], and stress and anxiety disorder treatment [20], [21].

From a neuroscience perspective, real-time access to the random activity and behavior of brain waves provides instantaneous valuable information about the nature of the EEG. Thanks to this feature, the treatment of many cognitive disorders and the diagnosis of abnormal brain behavior disorders have become understandable. The ability of this type of EEG to send instant feedback in response to brain functions is considered a great advantage in early medical diagnosis and BCI studies [22]. Continuous observation of brain activities in epilepsy and monitoring of the abnormal patterns of the brain that occur during seizures have attracted many researchers in real-time EEG studies [23]. Thanks to this reality, early detection of seizures and detailed analysis of their characteristics have become possible.

Epilepsy disorder, caused by excessive electrical movements of neurons, is the fourth known neurological disorder worldwide [24]. Regardless of age, the result of these involuntary excessive electrical oscillations can manifest itself as recurrent seizures. The topics of interest in the evaluation of a patient with recurrent seizures can be expressed as follows: Is the source of the seizures neurological? What type of seizure is it? Are these seizures a sign of epilepsy? With the correct answers to these questions, the patient's treatment process is determined. In addition to determining treatment, it is important to provide regular follow-up and support for individuals with suspected epilepsy. The presentation that summarizes the care process of epilepsy patients to improve their quality of life is shown in Figure 2.

Figure 2. The Care Process of Epilepsy Patients in Order to Improve the Quality of Life [25]

2. Related Studies

Seizures in epilepsy directly affect the patient's quality of life. These seizures, which vary in duration and severity from person to person, are manifested by variable behavioral patterns [24]. These behaviors are generally classified as loss of consciousness, staring, immobility or involuntary motor symptoms, muscle stiffness, etc.

Based on the World Health Organization (WHO) report in 2019, the stark reality is that the majority of epilepsy patients live in poor countries and do not have access to low-cost anti-seizure drugs. This report can report the good news that a quarter of epilepsy patients will be saved from annoying and untimely seizures by using low-cost drugs.

Computer-based models, which provide great convenience in the medical world, can diagnose many disorders by systematically analyzing physiological signals [26]. EEG, a physiological signal, is a magical tool revealing different brain disorders. EEG still maintains its important position as a diagnostic tool in diagnosing epilepsy disorders caused by excessive electrical activity of the brain [27].

In a review study conducted in 2015 [28], EEG was proven to be useful in diagnosing epilepsy syndrome, determining whether the attack belongs to the origin of epilepsy, and predicting attack recurrence. In this situation, any abnormal activity in brain cells leaves a bright trace of information worth investigating in EEG signals.

Important factors in diagnosing epileptic patients include determining the type of epileptic seizures and investigating the region in the brain causing epilepsy syndrome, thanks to the collaboration of EEG, neuroimaging, cheap drugs, and genetic tests [29].

Epilepsy diagnosis studies based on machine learning are increasing rapidly every day. Numerous studies have been conducted to diagnose seizures using different machine-learning techniques by analyzing EEG data [30]. An effective method was presented for detecting epileptic seizures in different EEG signal combinations using four common classifiers. In that study, the weighted complex networks model appeared useful in diagnosing epilepsy syndrome despite its high noise tolerance.

With the advancement of technology, automatic epilepsy seizure detection models based on machine learning have replaced traditional and time-consuming diagnostic methods. In a detailed study conducted in 2018 [31], the support vector machine algorithm was selected for the classification step using two feature extraction techniques. The procedure of this successful study was recommended for use in epileptic and healthy signal classification.

A different approach based on machine learning using EEG signals for epileptic seizure diagnosis was presented by Amin et al. [32]. In this study, automatic seizure diagnosis was achieved using wavelet analysis and arithmetic coding. Since the visual evaluation of EEG signal analysis, widely used in epilepsy diagnosis, is error-prone, erratic, and sensitive to subjective variability, automatic machine learning-based methods have been the focus of the cited study. This computer-aided diagnosis (CAD) technique consists of three steps and can be used easily in clinical studies by providing extraordinary performance in classifying epileptic and healthy individuals.

In CAD techniques utilized in epilepsy studies, effective features in the frequency, time, or time-frequency domain are extracted from the frequently preferred EEG signals [33]. This effective step directly affects the performance of classification algorithms. Classification algorithms are then employed to diagnose seizures or to distinguish epileptic signals from healthy signals. With the aim of shedding light on medical and clinical studies, the subject of rapid and successful seizure diagnosis and epileptic/healthy signal classification continues to be current and attracts attention in the scientific literature.

Diagnosis of epileptic seizures was possible in EEG signals with four classification methods using a genetic algorithm [33]. In this study, the success of classification algorithms was evaluated in terms of accuracy, specificity, and sensitivity. Among the algorithms used, artificial neural networks showed a more successful result than other algorithms, with 97.82%.

In another machine learning-based study for seizure diagnosis, a different approach was presented, concentrating on discrete wavelet transform and using a feature selection technique. In this new approach, using linear discriminant analysis (LDA), principal component analysis (PCA), and statistical features, epileptic and healthy EEG signals were classified with k nearest neighbor (KNN) and naive Bayes methods [31]. This model, applied in the epilepsy database of the University of Bonn, has achieved great success for the KNN classification algorithm.

Time-frequency statistical feature selection was used in another machine learning-based EEG epileptic classification model. In the model, independent component analysis (ICA) [34]was performed to effectively extract various processes from EEG. For the subjective-independent analysis of dynamic and highly non-stationary EEGs, a detailed analysis was carried out using ANOVA-based feature selection and fuzzy classification methods. The accuracy of the study was determined to be 96.48% in terms of seizure diagnosis [35].

A dataset for epilepsy seizure classification was prepared at Izmir Katip Çelebi University in 2021. In this study, the Neurofax device recorded the EEG signals of 16 subjects using surface electrodes. To expand the research, this study utilized a dataset financed by the European Union and another dataset known as EPILEPSIAE. Thanks to four technical features, such as energy, correlation, power spectral distance, and statistical significance measures, and using the empirical mode decomposition (EMD) technique, a success rate of 96.8% was achieved for the naive Bayes classifier.

Based on the literature review, epilepsy crisis detection, seizure diagnosis, and signal classification continue to be the focus of many studies. Epileptic EEG signal classification based on machine learning is the main goal of numerous studies. In the current study, the widely used dataset supplied by the University of Bonn [36]was used for our machine learning model design. In healthy and epileptic EEG signal classification, the importance of band rhythm efficiency was emphasized by analyzing all EEG bands and the rhythms of the EEG band. In this detailed Epilepsy classification analysis, attention is paid to EEG rhythm analysis, which appears to be a way to increase model accuracy.

3. Materials and Method

3.1. Dataset

This study, which offers a different perspective for those not highly experienced in machine learning-based physiological signal analysis studies, uses the EEG dataset provided by Andrzejak et al. [37] from the University of Bonn. In this case, the author of the proposed study did not apply for ethical approval. In this dataset consisting of five subsets (Z (A set), O (B set), N (C set), F (D set), and S (E set)), EEG signals were sampled with a sampling frequency of 173.61 Hz. Each set consists of a single channel and contains 100 trials. There are 4097 samples in each of these trials, which last 23.6 seconds. These EEG signals were recorded with 12-bit resolution. As a result of visual inspection, artifacts resulting from eye and muscle movements were largely eliminated from EEG recordings. In this recording system, which has a bandwidth between 0.5 Hz and 85 Hz, a 40 Hz low-pass filter was applied to the raw EEG signal as a preprocessing step.

Table 1 summarizes this data to understand the dataset efficiently and comprehensively [36]. The electrode positions used for EEG recordings taken with surface electrodes in sets A and B are shown in Figure 3. When the results obtained from the scalp with extracranial electrodes in the EEG recording process are unsuccessful, intracranial electrodes are used by the surgeon to record brain activity by placing them on the patient's skull.

Sets A and B contain EEG data recorded from 5 healthy subjects in an awake and relaxed state with eyes open in set A and closed eyes in set B. Data from 5 patients who achieved complete control of their epilepsy after removal of the epileptic seizure focus are presented in sets C and D. Finally, E contains the only ICTAL (a seizure episode) activity observed in epileptiform regions. Figure 4 displays an example of an EEG signal related to each set.

Table 1. Dataset Details

Classes	Subject	State	Electrode Type	Number of Trials	Number of Samples
Set A	5 healthy subjects	Awake with open eyes	Extracranial	100	4097
Set B	The same 5 healthy subjects	Awake with closed eyes	Extracranial	100	4097
Set C	5 epileptic subjects	seizure free	Intracranial	100	4097
Set D	The same 5 epileptic subjects	seizure free	Intracranial	100	4097
Set E	The same 5 epileptic subjects	with seizure	Intracranial	100	4097

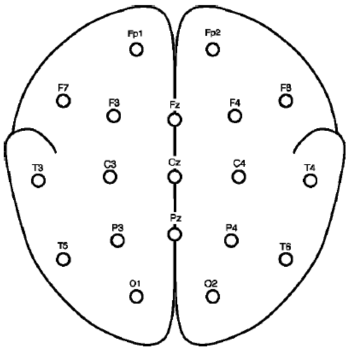


Figure 3. International 10-20 System Surface Electrode Positions

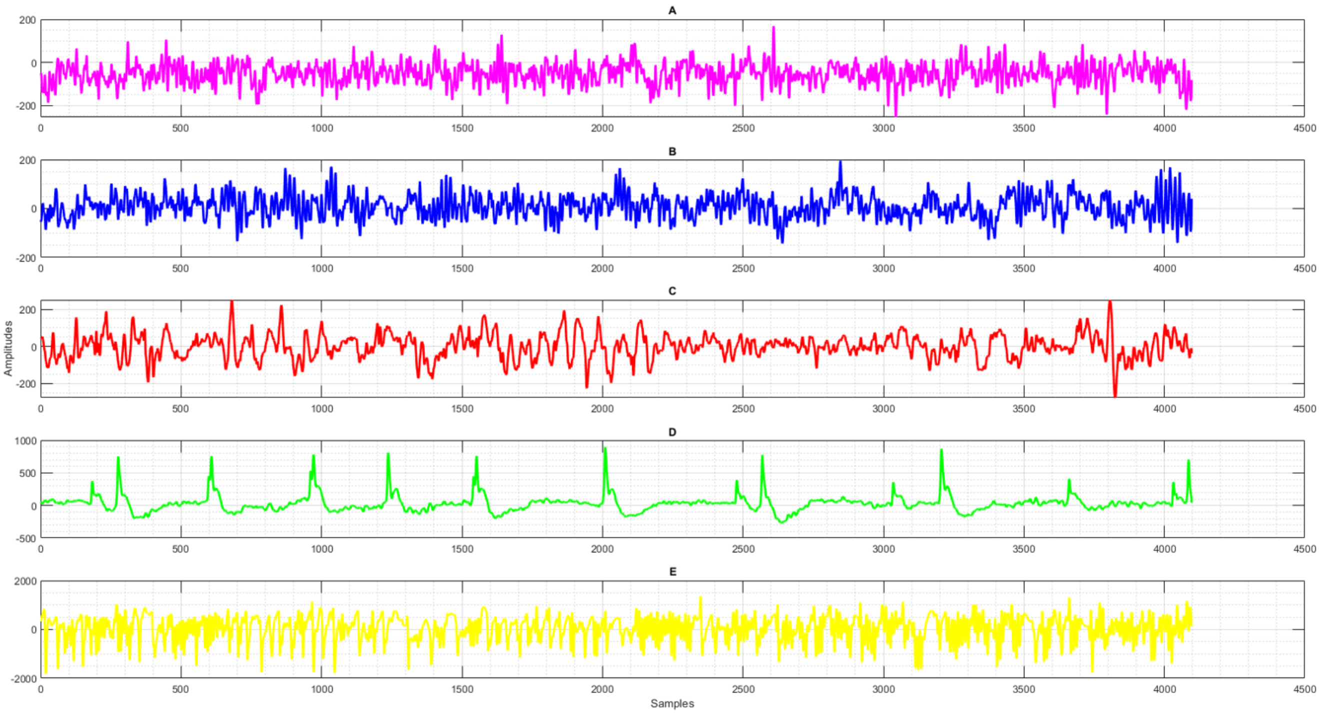


Figure 4. An Example of an EEG Signal Related to Each Set

3.2. Applying machine learning steps

This study aims to present comprehensive research to researchers who are pioneers in EEG signal processing by comparing the classification performance of time domain feature extraction [38] against the classification performance of frequency-domain feature extraction [39] in the diagnosis of epilepsy.

In the first stage, EEG records with a .txt extension are entered into MATLAB and prepared for analysis. Before analysis, these data are converted to (.mat) format. MATLAB R2023a was used for all analyses. As the initial stage of machine learning, pre-processing is performed on EEG data. In this study, noise reduction and separation of EEG data into rhythms were performed as preprocessing. In this step, noise reduction and common data deletion, which are important and useful features of preprocessing, will be realized. So that, by using the pre-processing method called filtering, unwanted components were removed and it became easier to extract meaningful information for successful classification results.

The process of measuring the parameters of data is called feature extraction. The purpose of this step is to facilitate the classifier process. A measurable parameter of the fact we observe is the feature. Valuable features properly fulfill their informative role by containing accurate and relevant information about the facts. Another point that represents a good feature is distinctiveness. That means having a different value in two or more classes of data. The acceptable feature is showing the maximum variance between classes while having similar values within the class. Finally, the condition that the features are independent of each other and provide new information about the signal directly affects the classifier's performance.

In the time domain feature extraction method, the rhythms of the EEG signal were obtained by applying the Fast Fourier transform (FFT). At this stage, we produced signals that carry each rhythm information of the EEG in the time domain. Rhythms were studied instead of focusing on the entire EEG band to prevent performance degradation because some common features are too many and the differences are few in the entire EEG. In this case, the extracted feature will be a good ambassador for the EEG signal. So, it would be more useful to separate the rhythms and extract the characteristics of each rhythm separately. Some rhythms may be good and some may be bad, but this does not matter because rhythms that carry useful information can be selected for classification using feature selection. This step is summarized for a trial from set A in Figure 5. Time domain feature extraction steps are presented in detail in Figure 6.

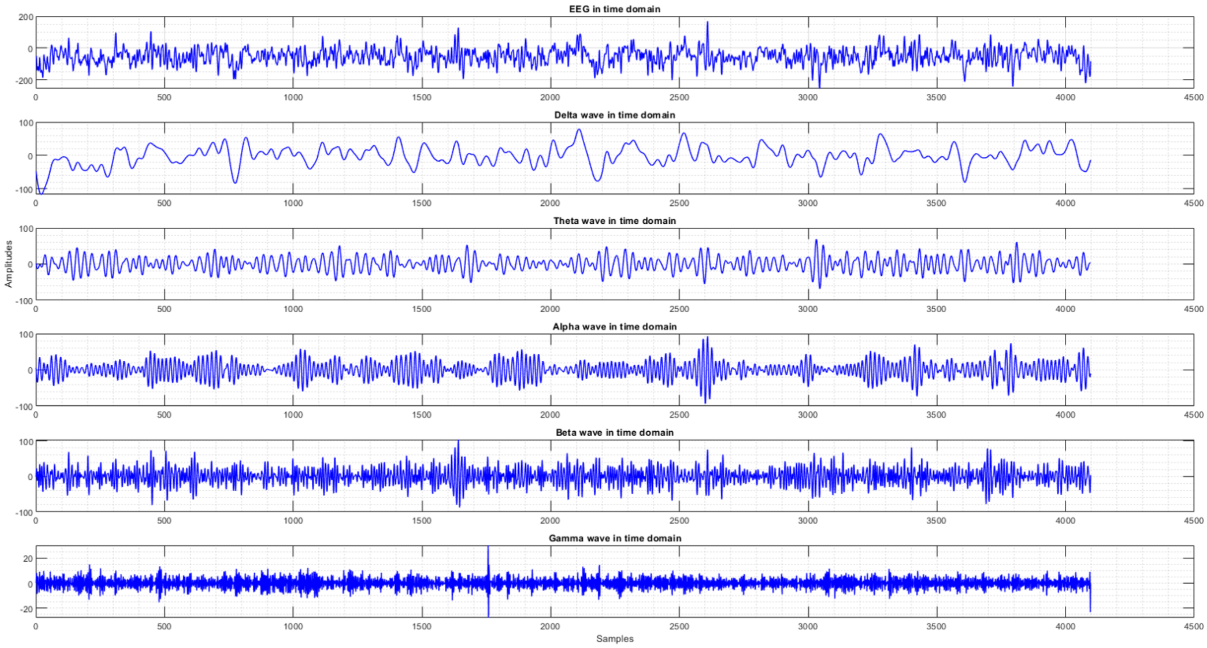


Figure 5. Preparation for Time Domain Feature Extraction from a Trial of a Set

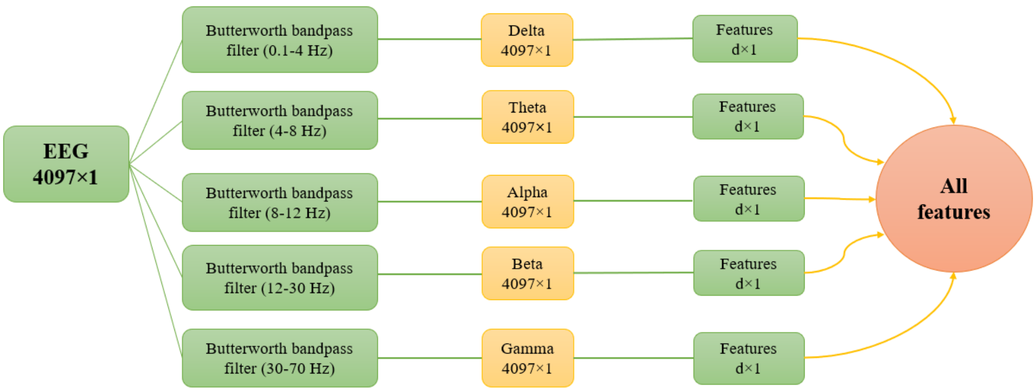


Figure 6 Time Domain Feature Extraction Steps

In calculating the statistical properties [40] of the signal in the time domain, the d value was determined as 6 by using mean, variance, skewness, kurtosis, entropy, and signal power. Mean contains information about the median of the fact data. It is represented by the parameter μ and shows the average of all signal samples. In the x signal with size N×1, the mean is calculated by Equation 1. The parameter N refers to the number of samples.

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1}$$

The variance shows the distribution of the fact data around the mean and is denoted by σ^2. Standard deviation (σ) is the square root of the variance. Since it contains the same information as variance, it can be used for analysis. In the x signal with size N×1, the variance is calculated by Equation 2.

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \tag{2}$$

Skewness [41] determines information about the asymmetry of the normal distribution. This criterion, which defines the degree of asymmetry, has made decision-making in research areas more successful and provided the opportunity for accurate modeling. This type of distribution is expressed by Equation 3.

$$Skewness = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^3}{(N-1)\sigma^3} \tag{3}$$

Kurtosis determines the peak degree of the normal distribution of the event. Thanks to this statistical criterion, the degree of tailing of the distribution is clearly shown compared to the normal distribution. This type of distribution is represented by Equation 4.

$$\boldsymbol{Kurtosis} = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^4}{(N-1)\sigma^4} \tag{4}$$

The calculation of entropy and signal power can be used for features. Shannon Entropy [42] determines the randomness of a phenomenon. Whether the behavior is random or factual in a case is determined by Shannon Entropy. The mathematical representation of these parameters is shown in Equations 5 and 6, respectively. By including these two features in the study, the total number of features was increased to 6 for each set.

$$\boldsymbol{H}(\boldsymbol{x}) = -\sum_{i=1}^{N} \boldsymbol{p}(\boldsymbol{x}_i)\,\boldsymbol{log}\,(\boldsymbol{p}(\boldsymbol{x}_i)) \tag{5}$$

$$\boldsymbol{P} = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x}_i)^2 \tag{6}$$

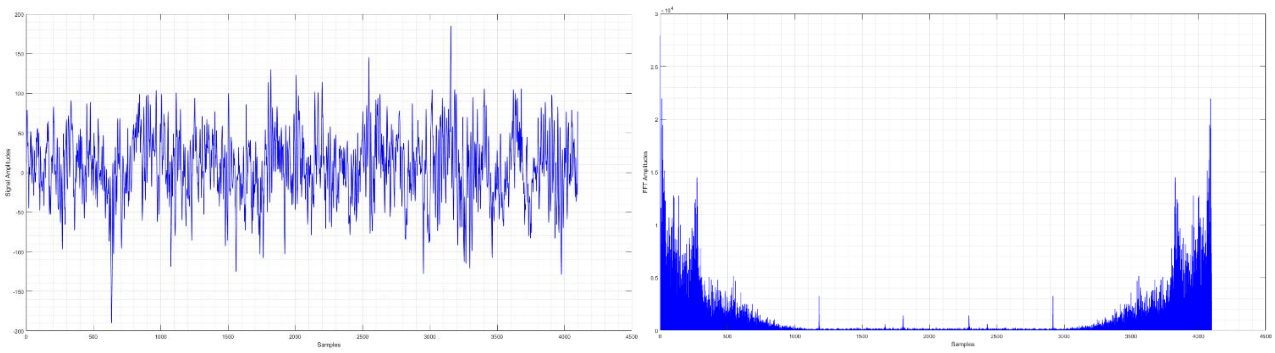
The information provided by the time domain indicates how the amplitude of the signal changes over time. In the rhythm separation process in the time domain, rhythms were obtained as a result of FFT calculation from the signal. Finally, by taking the inverse FFT, the signal in the time domain was the same size as the main signal but was directed to the feature step by obtaining a signal that would only include the coefficients of that rhythm.

In the present study, feature extraction in the frequency domain [43] plays a major role in diagnosing epilepsy. The information provided by the frequency domain shows in which frequency ranges the signal contains information. It shows the frequency band content of the EEG signal. As it is known, when FFT is calculated, the output is complex numbers. These complex numbers have an amplitude and a phase. Each signal is made of an infinite number of sinusoidal signals. Now, the frequency domain shows which sine signals the signal is made of and how this sine contributes to the formation of the main signal. Each sinusoidal wave has an amplitude and a phase.

After FFT, half of the FFT coefficients are considered for feature extraction. The first frequency is 0 Hz, and the frequency of the last coefficient is $F_{sampling}/2$ Hz. Frequency resolution (FR) was taken into account to calculate other frequencies. The time domain feature extraction steps are presented in detail in Figure 7.

Following the collaboration between medicine and engineering, the analysis of biosignals has gained momentum for early disease diagnosis [44]. EEG classification is the process of identifying a specific medical condition or classifying a specific activity by taking EEG signals. These signals are used to measure brain activity and diagnose conditions such as epilepsy, sleep disorders, ADHD [45], or in brain-computer interface (BCI) systems [46]. Classification algorithms predict whether features belong to a particular class. Common classification algorithms include support vector machines (SVM) [47], artificial neural networks (ANN) [40], decision trees [48], and k nearest neighbors (k-NN).

In the proposed EEG-based machine learning model design, the decision phase was carried out after pre-processing and time/frequency domain feature extraction in diagnosing epilepsy disease and healthy subjects. Examining how time and frequency domain features affect the classification step and which one is superior in terms of successful classification results constitute the basic framework of the study. The classification steps followed in the article are illustrated in the flowchart in Figure 8.



(a) (b)

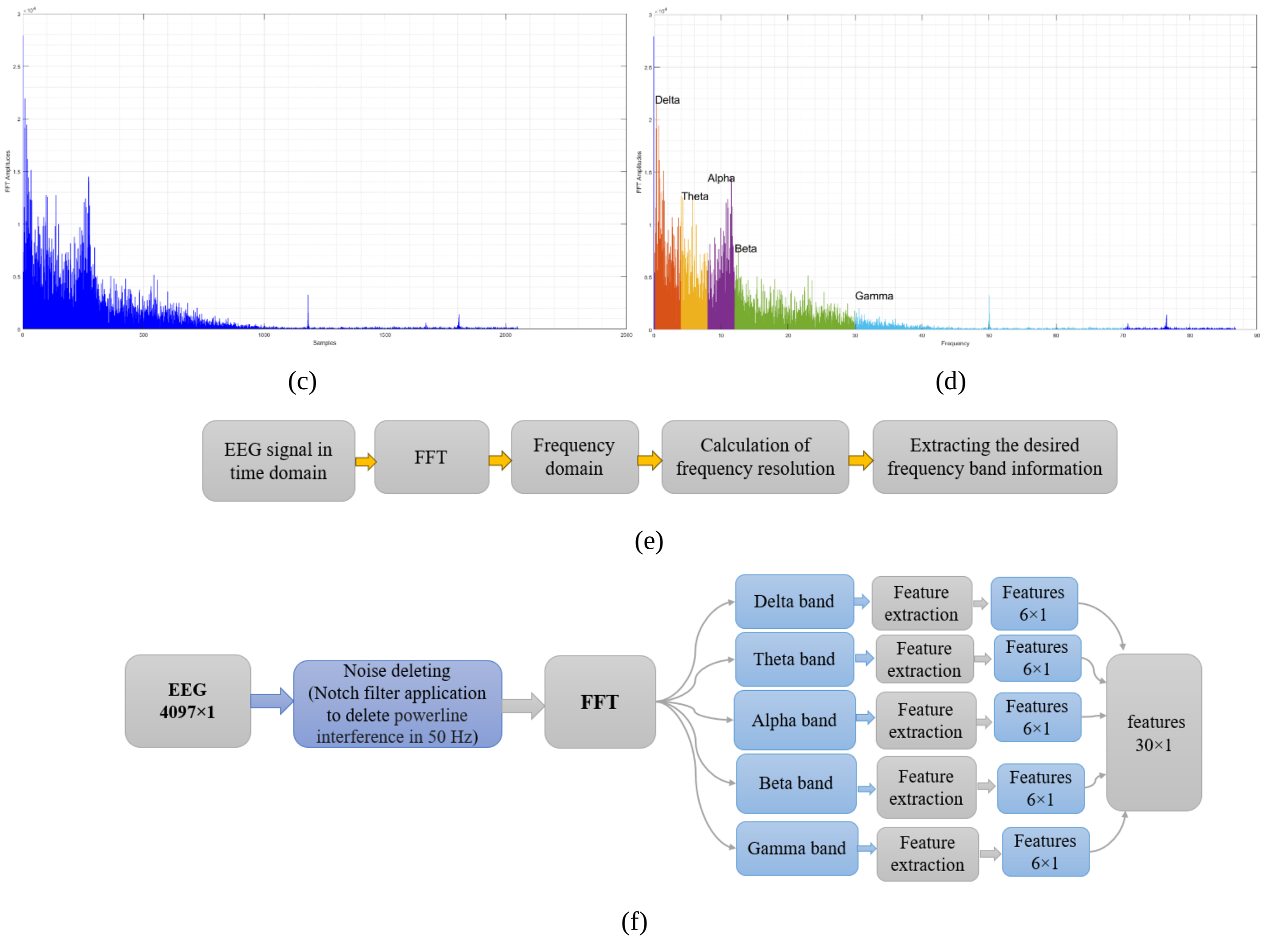


Figure 7. (a) EEG signal in the time domain, (b) Entire spectrum, (c) Half of the spectrum (select half of the coefficients), (d) Determination of rhythm coefficients, (e) and (f) General and detailed representation of feature extraction steps in the frequency domain

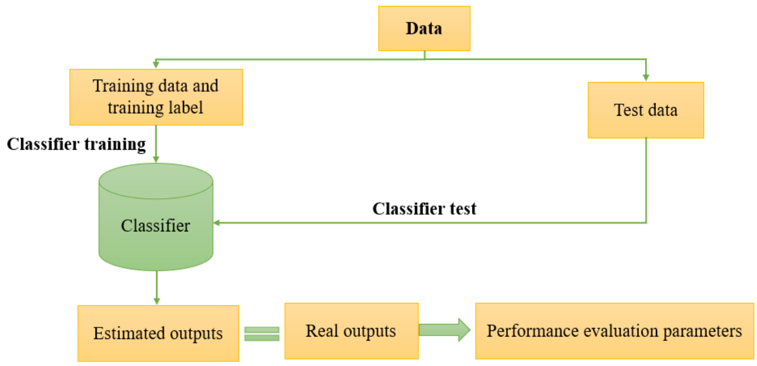


Figure 8. Classification Steps

After the necessary preprocessing on the EEG data, the classification step is started by extracting separate statistical features in the time and frequency domain. As a classification, the KNN algorithm [49] has become the chosen algorithm in many machine learning studies due to its easy use and successful performance [50]. Despite its simplicity, it is very effective and accomplished for determining an observation label. The KNN algorithm, which has the advantages of simplicity, easy understandability, and applicability, can be computationally costly and prone to overfitting when working on large datasets. Therefore, when using KNN, it is important to set the K parameter well and consider the characteristics of the dataset [51].

Although the non-parametric KNN classifier has an easy working principle, it provides a very successful and strong performance in noisy data. The KNN algorithm, which is also the focus of attention in the scientific world, is developing with the emergence of new methods. The KNN algorithm, which is frequently used in both regression and classification problems, is distance-oriented and works based on the assumption that the samples in the dataset are close to each other. In this context, the Euclidean distance definition, widely used in the proposed model, was chosen for the algorithm [52]. If the

Euclidean distance between two points with (x_1, y_1) and (x_2, y_2) coordinates is denoted by D, this distance is presented in Equation 7.

$$\boldsymbol{D} = \sqrt{(\boldsymbol{x_2} - \boldsymbol{x_1})^2 + (\boldsymbol{y_2} - \boldsymbol{y_1})^2} \qquad (7)$$

Some of the solvable challenges of KNN include dealing with excessive examples, unbalanced classes, overlapping class regions, noise from irrelevant features, and outliers. Prototype development, correct adjustment of the K parameter, creation of artificial data, filter application, and outlier analysis are among the solutions that mitigate KNN difficulties.

The dataset must be divided into two sets, test and training, to carry out a classification project (as presented in Figure 8). Then, the classifier training will begin. The model will be trained with the training data and training label at this stage. Next, the trained classifier is tested with the test dataset. The most important step is to evaluate the model using performance parameters. Classifier accuracy, sensitivity, specificity, and confusion matrix were calculated as performance parameters. A detailed mathematical description of accuracy, sensitivity, and specificity is shown in Equations 8, 9, and 10, respectively. Additionally, Table 2 represents the binary problem confusion matrix detail. In the proposed epilepsy machine learning binary classification problem, *TP* belongs to the first class and represents the correctly classifying examples. *FP* shows examples that belong to the first class but are misclassified. *TN* belongs to the second class and shows correctly classifying examples, and finally *FN* belongs to the second class and shows incorrectly classifying examples.

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{TP+TN}}{\boldsymbol{TP+TN+FP+FN}} \qquad (8)$$

$$\boldsymbol{Sensitivity} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}} \qquad (9)$$

$$\boldsymbol{Specificity} = \frac{\boldsymbol{TN}}{\boldsymbol{TN+FP}} \qquad (10)$$

Table 2. Binary Problem Confusion Matrix [53]

		Predicted Class	
		Positive	**Negative**
Actual Class	**Positive**	TP	FN
	Negative	FP	TN

In our study, four validation methods were employed for model evaluation, namely, the holdout method, k-fold cross-validation method, Leave one out validation method, and Random subsampling, respectively.

The hold out method [53] is commonly used, where data are allocated only once for training and testing. The model is trained with the training data and tested with the test data; then, the performance is evaluated. Generally, 70% or 80% of all data is allocated to training. This method is suitable for large and crowded datasets. However, in small datasets, due to the limited size of training and test data, the model may not be adequately trained, and statistical problems may arise in the test data, leading to unreliable performance parameters. In the hold out method, training and test data are separated only once. When dealing with small datasets, model training may not be complete due to limited data availability. Performance evaluation aims to utilize the maximum amount of data for the training process, as the more training data available, the better the model learns. Conversely, it is more useful to use the maximum amount of data for testing to ensure statistically reliable results. However, the maximum data capacity is limited to the entire dataset. Thus, it is not possible to use the maximum data level for both testing and training since the same data cannot be utilized for both purposes. In the hold out method of the current study, training and testing data were separated using a 70% to 30% ratio, respectively.

In the leave-one-out validation method [54], which carries the same k-fold logic, there is 1 sample in each fold instead of several samples. The advantage of this method is that it allows the dataset to be used in the most beneficial way in small datasets. Calculating the performance for each stage does not make sense because there is only 1 sample in the test. Thanks to the test labels, after the prediction process is completed, the real labels will be compared with the prediction labels, and the performance will be calculated.

In the random subsampling model, the train and test process will repeat k times. Randomly, some are allocated as testing, and some as training, and the performance of each stage is calculated separately. This method has many repetitions. k is generally chosen as 100 or 200 [55]. Since the EEG data division process is random in this model, it will obtain even more reliable predictions. The use of a single training-testing in validation methods can sometimes be undesirable. Single train-test applications can diminish classifier performance by increasing the likelihood of bias and variability. The random subsampling verification method has acceptable performance in eliminating possible situations. Training and testing data were separated by 70% and 30% ratios, respectively.

Extensive use of the dataset is possible by using some techniques. k fold validation [56] is a special and common method that uses the maximum level of the data to test. Approximately 90% of the data is reserved for training. It allocates a small portion of the data for testing, but in fact, it uses the maximum level of the data for testing. It divides the data into k equal parts. k has a gradual repetition process. In article research, k is generally chosen as 5, 10, 15, 20 [57]. The data is divided into k parts: 1 is used for testing, and (k-1) is used for training. The method will continue until all folds are used as tests once. After taking the performance of each step separately, the average performance of all steps is calculated. For example, if we have 100 data, only 30 are used for testing in the hold out method. Nevertheless, k-fold will pass all 100 tests step by step. In the k-fold verification method, k = 5 was chosen for all stages.

4. Results and Discussion

With time and frequency domain feature extraction methods, KNN classification result percentages were calculated for two values of K, representing the number of neighbors.

In the tables,

- A/C and A/D represent healthy subjects (recording with eyes open) vs. subjects with epilepsy without seizures.
- A/E shows healthy subjects (recording with eyes open) vs. subjects with epilepsy with seizures.
- B/C and B/D show healthy subjects (recording with eyes closed) vs. subjects with epilepsy without seizures.
- B/E represents healthy subjects (recording with eyes closed) vs. subjects with epilepsy and seizures.

Considering six binary classes for classification, the accuracy, sensitivity, and specificity for the four evaluation models are presented in Tables 3, 4, 5, and 6 for the time and frequency features for the number of neighbors K = 3 and K = 5, respectively. To avoid the crowding of figure presentations, the confusion matrix for a single strategy method in both time and frequency domain feature extraction has been added for six binary classes (leave one out for time domain, and the hold out for frequency domain). The confusion matrix for the six binary classes under the leave one out and the hold out strategies for the KNN algorithm (K=5) based on the features extracted in the time and frequency domain is given in Figures 9 and 10 respectively.

Table 3. KNN Classification Results (K=3) for Time Domain Feature Extraction Methods

Model evaluation	Performance parameters	A/C	A/D	A/E	B/C	B/D	B/E
The holdout	Accuracy	98.34	95	98.34	100	98.34	98.34
	Sensitivity	100	100	100	100	100	100
	Specificity	96.67	90	96.67	100	96.67	96.67
Leave one out	Accuracy	71.5	62.5	100	57	51	96.5
	Sensitivity	77	70	100	57	51	97
	Specificity	66	55	100	57	51	96
Random subsampling	Accuracy	66.31	60.23	100	58.01	54.45	97.13
	Sensitivity	70.93	66.95	100	56.84	55.35	96.97
	Specificity	62.21	54.38	100	59.82	54.18	97.42
k-fold	Accuracy	92.5	89.5	99.5	97	96	99.5
	Sensitivity	88	87	100	95	93	100
	Specificity	97	92	99	99	99	99

Table 4 KNN Classification Results (K=5) for Time Domain Feature Extraction Methods

Model evaluation	Performance parameters	A/C	A/D	A/E	B/C	B/D	B/E
The holdout	Accuracy	96.67	93.34	98.34	100	98.34	98.34
	Sensitivity	100	100	100	100	100	100
	Specificity	93.34	86.67	96.67	100	96.67	96.67
Leave one out	Accuracy	68	58.5	100	59.5	56	96.5
	Sensitivity	76	68	100	59	58	97
	Specificity	60	49	100	60	54	96
Random subsampling	Accuracy	62.65	57.3	99.98	58.65	55.73	97.05
	Sensitivity	69.06	63.1	100	59.85	63.35	97.37
	Specificity	56.86	51.98	99.97	58.34	48.77	96.77
k-fold	Accuracy	92.5	88	99.5	97	96	99.5
	Sensitivity	89	87	100	95	93	100
	Specificity	96	89	99	99	99	99

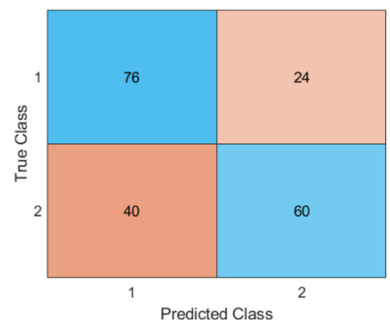
Table 5. KNN Classification Results (K=3) for Frequency Domain Feature Extraction Methods

Model evaluation	Performance parameters	A/C	A/D	A/E	B/C	B/D	B/E
The holdout	Accuracy	98.34	91.67	98.34	100	98.34	98.34
	Sensitivity	100	100	100	100	100	100
	Specificity	96.67	83.34	96.67	100	96.67	96.67
Leave one out	Accuracy	95.5	94	99.5	98.5	98	99.5
	Sensitivity	94	94	100	98	97	100
	Specificity	97	94	99	99	99	99
Random subsampling	Accuracy	94.46	92.88	99.5	98.53	97.83	99.35
	Sensitivity	92.78	93.34	100	98.07	96.2	100
	Specificity	96.16	92.68	99	99	99.5	98.68
k-fold	Accuracy	94.5	93.5	99.5	97.5	96.5	99.5
	Sensitivity	92	93	100	96	94	100
	Specificity	97	94	99	99	99	99

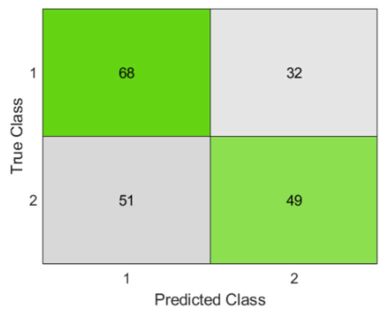
Table 6. KNN Classification Results (K=5) for Frequency Domain Feature Extraction Methods

Model evaluation	Performance parameters	A/C	A/D	A/E	B/C	B/D	B/E
The holdout	Accuracy	98.34	90	98.34	100	98.34	98.34
	Sensitivity	100	100	100	100	100	100
	Specificity	96.67	80	96.67	100	96.67	96.67
Leave one out	Accuracy	94.5	93	99.5	98.5	98	99.5
	Sensitivity	93	94	100	98	96	100
	Specificity	96	92	99	99	100	99
Random subsampling	Accuracy	93.76	91.13	99.43	98.51	97.61	99.51
	Sensitivity	92.96	94.44	100	97.82	95.86	99.97
	Specificity	94.79	88.11	98.87	99.17	99.4	99.02
k-fold	Accuracy	93	91	99.5	98	96	99.5
	Sensitivity	91	93	100	97	94	100
	Specificity	95	89	99	99	98	99

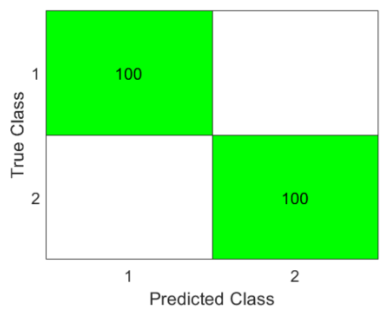
One of the biggest problems with machine learning model design is the issue of overfitting and underfitting. When training the model, the problem of overfitting occurs when maximum fit is achieved on the training data but generalization is not made to newly seen data. This problem was minimized by choosing the correct parameters in our proposed model. Thus, when we look at the results presented, except for the leave-one-out and random subsampling strategies in the time domain, the fit shown on the training data in other analyses shows the same consistent performance on the test data. Underfitting includes low accuracy on both training and validation/testing data. In our proposed model, it is aimed to address the issues of overfitting and underfitting by providing an adequate balance between information generalization and model complexity.



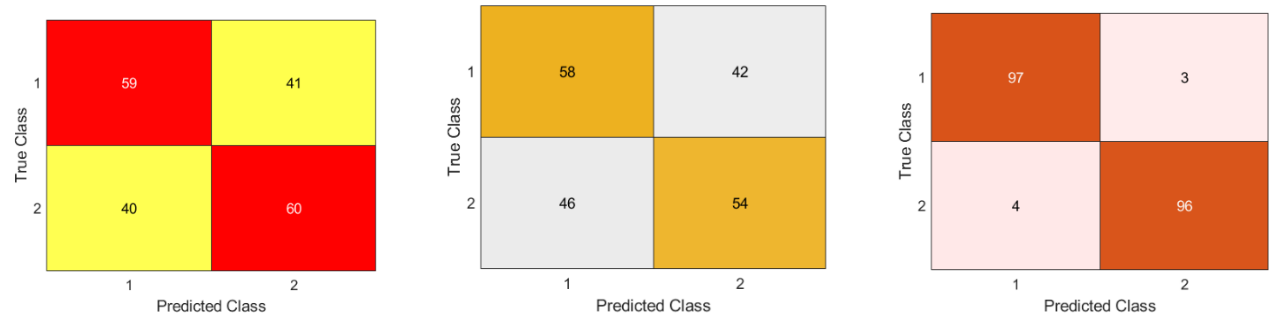
(A/C)



(A/D)

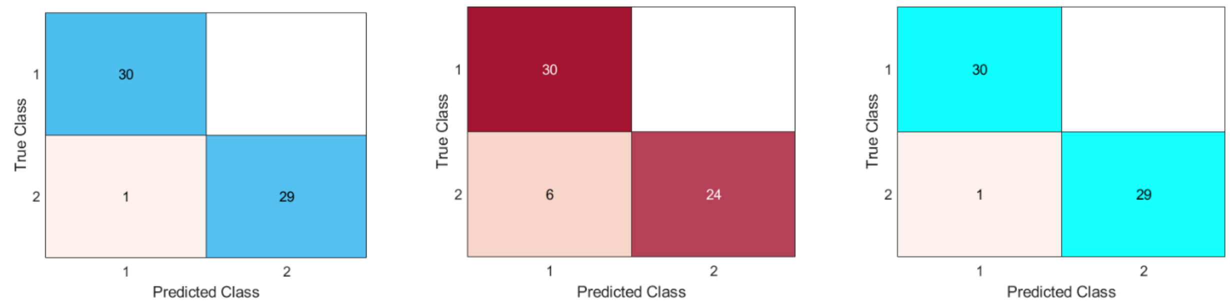


(A/E)

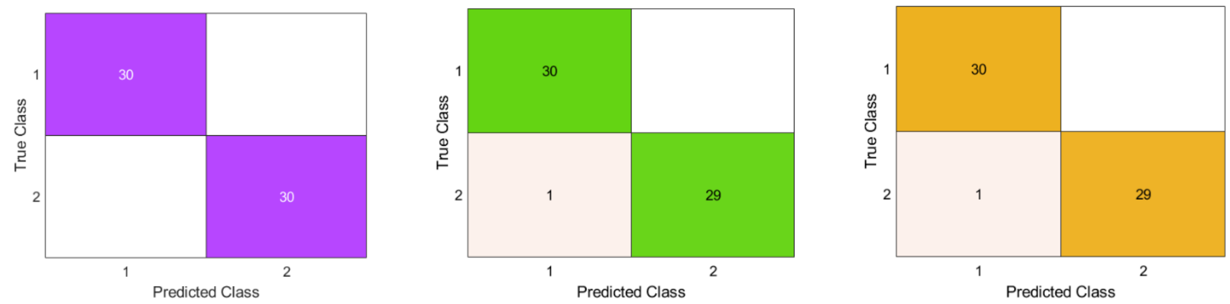


(B/C) (B/D) (B/E)

Figure 9. The Confusion Matrix for the Six Binary Classes Under the Leave One Out Strategy for the KNN Algorithm (K=5) (Time Domain Features)



(A/C) (A/D) (A/E)



(B/C) (B/D) (B/E)

Figure 10. The Confusion Matrix for the Six Binary Classes Under the Holdout Strategy for the KNN Algorithm (K=5) (Frequency Domain Features)

In general, there was a high degree of agreement between accuracy, sensitivity, and specificity in all analyses. Tables can be evaluated from two perspectives; the first is the high accuracy evaluation between classes, and the second is the successful model evaluation of each class in model performance methods.

Much detailed research has been conducted on the dataset we are examining. We aim to simplify the research process by guiding graduate students who are new to the field of machine learning and pattern recognition. In this study, we evaluated machine learning models using time and frequency domain features, employing four performance evaluation strategies during the classification phase to distinguish between epilepsy and healthy subjects. Based on the findings from our literature review, our classification model closely aligns with previously reported results. We demonstrate the success of our proposed model through metrics such as classification accuracy, sensitivity, specificity, and the confusion matrix.

From the first perspective, the highest achievement for each accuracy row is underlined in the tables. In this evaluation, the percentage of the most successful class in 6 binary class problems shows a very good agreement with time and frequency domain feature extraction and different K neighbor numbers. From a general perspective, leave one out and random subsampling, which had poor performance in extracting time domain features, experienced a significant increase in success with frequency domain features. As expected, successful classification was achieved in all model performances due to classification with seizure epilepsy, i.e., E, in healthy EEG sets A and B due to the pattern clearly displayed in the waveform.

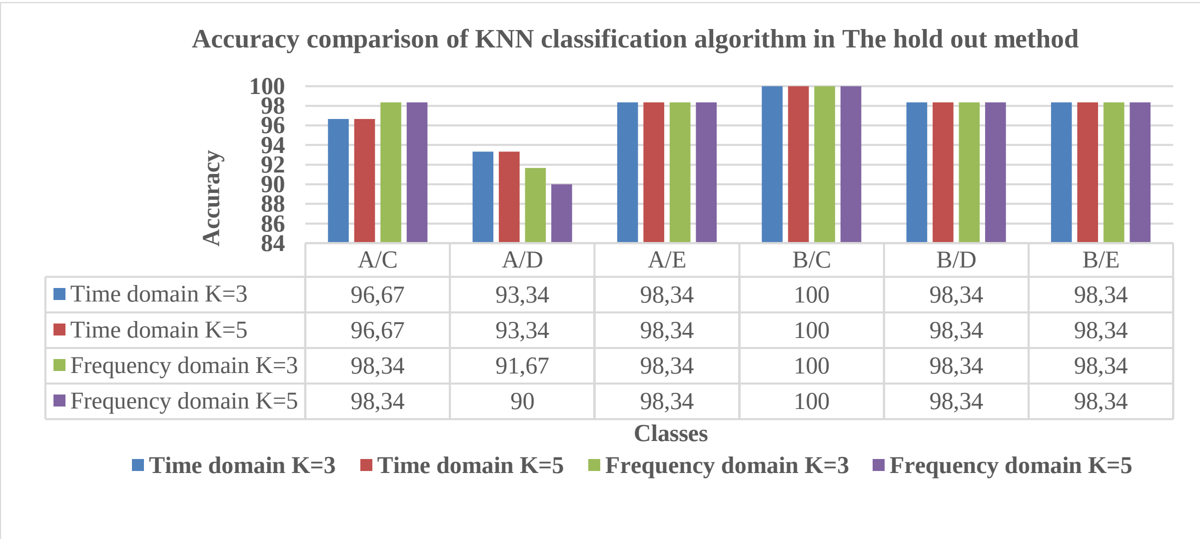
To compare the classification result in a proper and understandable manner, the accuracy success percentages for each evaluation model are shown in detail in Figures 11, 12, 13, and 14 for the hold out, the leave one out, random subsampling, and k-fold methods, respectively.

In Figure 11, the B/C healthy and seizure-free epilepsy classification has achieved 100% success in the hold out validation model in the KNN classification made with both time and frequency domain feature methods. On the other hand, in the healthy and seizure epilepsy classification analysis (A/E, B/E), the proposed feature extraction techniques achieved classification with a rate of 98.34%. The class that had the lowest success in the binary class groups was obtained as 90% in the classification of healthy (recording with eyes open) and non-seizure epilepsy, with the frequency domain feature extraction technique in KNN with the number of neighbors K = 5. In this way, frequency domain analysis provides a more successful outcome than time analysis in A/C classification. However, the time domain yields a more successful result in A/D classification.

In Figures 12 and 13, the highest success belongs to the healthy and seizure epilepsy classification (A/E, B/E). For this high success, time and frequency domain feature extraction techniques demonstrated approximately similar performance. Here, it cannot be overlooked that the number of K neighbors for KNN sometimes directly influences the classification outcome. The K parameter can be selected optimally using a few trial-and-error methods. In both time and frequency domains, the number of neighbors of K = 5 provides a better accuracy rate in many classes than K = 3. For the same number of K neighbors (K=5), the feature extraction method in two domains yields similar results in the leave-one-out model evaluation. In general terms, the lowest success is also evident in the healthy and seizure-free epilepsy classification (A/D, B/D).

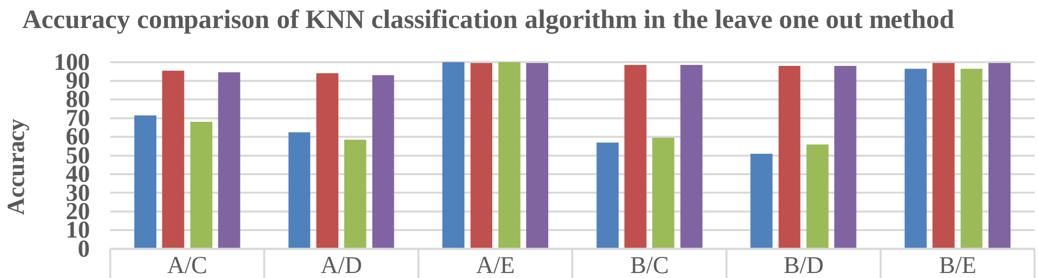
In the k-fold model evaluation in Figure 14, the time and frequency domain classification results offer similar results in all binary classes, such as the retention model. Healthy and seizure-free epilepsy classification (A/D) has a lower success rate compared to other classes. The highest success belongs to the healthy and seizure epilepsy classification (A/E, B/E), with a success rate of 99.5%. The effect of changing the K parameter in the KNN algorithm manifests itself in a small amount in this model for the frequency domain and K = 5. If a general result is summarized, the hold out and k-fold cross-validation models distinguish healthy and epileptic conditions with a higher success rate than the other two models. The success of k-fold has already been proven in many research articles [58]. In the hold out model evaluation strategy, KNN demonstrates 100% success in both the time and frequency domains in the B/C subgroup classification across all classification tables. As known, in this strategy, the dataset is divided into training and testing sets. The model is trained on the training set and evaluated on the test set. In this method, whose performance varies depending on how the data is divided [59], the division ratio that yields the best results is selected using a trial-and-error method. This selection results in the highest success rates in both the time and frequency domains.

This educational and small-scale EEG study revealed that, based on the figures, choosing the right cross-validation method depends on many factors, such as the calculation steps, the nature and size of the dataset, and the ideal level of sensitivity in the designed model.



	A/C	A/D	A/E	B/C	B/D	B/E
Time domain K=3	96,67	93,34	98,34	100	98,34	98,34
Time domain K=5	96,67	93,34	98,34	100	98,34	98,34
Frequency domain K=3	98,34	91,67	98,34	100	98,34	98,34
Frequency domain K=5	98,34	90	98,34	100	98,34	98,34

Figure 11. Accuracy Comparison of the KNN Classification Algorithm in the Hold Out Method

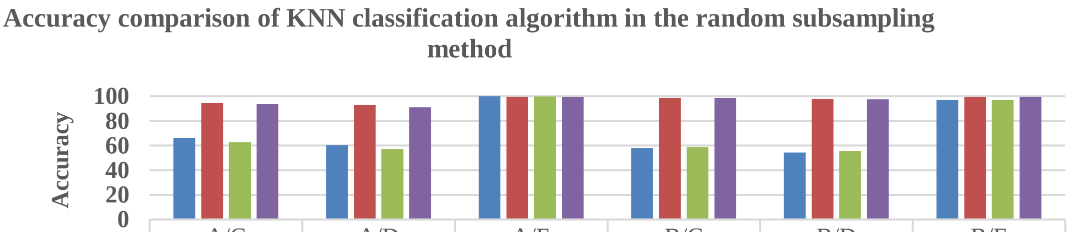


	A/C	A/D	A/E	B/C	B/D	B/E
Time domain K=3	71,5	62,5	100	57	51	96,5
Time domain K=5	95,5	94	99,5	98,5	98	99,5
Frequency domain K=3	68	58,5	100	59,5	56	96,5
Frequency domain K=5	94,5	93	99,5	98,5	98	99,5

Classes

Time domain K=3 **Time domain K=5** **Frequency domain K=3** **Frequency domain K=5**

Figure 12. Accuracy Comparison of the KNN Classification Algorithm in the Leave One Out Method

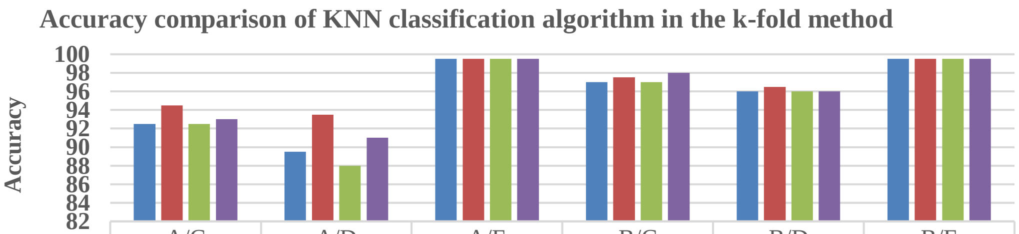


	A/C	A/D	A/E	B/C	B/D	B/E
Time domain K=3	66,31	60,23	100	58,01	54,45	97,13
Time domain K=5	94,46	92,88	99,5	98,53	97,83	99,35
Frequency domain K=3	62,65	57,3	99,98	58,65	55,73	97,05
Frequency domain K=5	93,76	91,13	99,43	98,51	97,61	99,51

Classes

Time domain K=3 **Time domain K=5** **Frequency domain K=3** **Frequency domain K=5**

Figure 13. Accuracy Comparison of the KNN Classification Algorithm in the Random Subsampling Method



	A/C	A/D	A/E	B/C	B/D	B/E
Time domain K=3	92,5	89,5	99,5	97	96	99,5
Time domain K=5	94,5	93,5	99,5	97,5	96,5	99,5
Frequency domain K=3	92,5	88	99,5	97	96	99,5
Frequency domain K=5	93	91	99,5	98	96	99,5

Classes

Time domain K=3 **Time domain K=5** **Frequency domain K=3** **Frequency domain K=5**

Figure 14. Accuracy Comparison of the KNN Classification Algorithm in the K-Fold Method

The dataset of the University of Bonn, which has balanced classes [60], continues to establish strong ties between the medical world and machine learning through its application for diagnosing many EEG-based epileptic diseases. Based on a literature review, the final results of this educational study are largely compatible with the findings obtained so far. In the machine learning-based epilepsy study conducted in 2021, a high success rate was achieved with the KNN algorithm using the PCA feature reduction technique [61]. After the dominant features are obtained, the KNN algorithm, which is useful in detecting abnormalities in epileptic seizures in EEG signals, achieved 97.5% success thanks to the features obtained by the wavelet transform feature extraction technique [62]. Considering the statistical feature extraction method, which has been used in many studies and has provided successful results, in the proposed high-performance machine learning model, energy, Shannon entropy, and variance significantly contribute to highlighting the dominant feature in seizure capture [63]. In a comprehensive 2021 study, various linear and non-linear features were extracted using the Bonn and Freiburg [64] datasets, and acceptable results were obtained with 10-fold cross-validation thanks to the deep learning approach. The proposed model demonstrated 99.71% and 99.13% accuracy for the Bonn and Freiburg datasets [65]. In a study conducted in 2023 with the Bonn dataset [66], an automatic epilepsy seizure detection model was designed. In this model, classification was carried out by applying a discrete wavelet transform, extracting four mixed features, and using a convolutional neural network algorithm. The classification success of this designed model was reported as 100%. Introducing a new feature extraction method [67], complex network logic was used for automatic epilepsy seizure detection. In this study, Henon and Logistic maps in the Bonn dataset were considered to demonstrate the validity of the methodology. In the detailed study, epilepsy seizure diagnosis was presented with 100% accuracy. In this epilepsy seizure diagnosis model, which is useful to expert neurologists, successful performance was demonstrated using different classifiers such as support vector machine and linear discriminant analysis. In a multiple classification study conducted in 2023 using deep learning techniques, epileptic seizures were diagnosed in the Bonn dataset [68]. Seizure diagnosis was achieved with a 99.5% success rate using scalogram and spectrogram images and multiple classification techniques.

References

[1] N. Manshouri, M. Maleki, and T. Kayikcioglu, "An EEG-based stereoscopic research of the PSD differences in pre and post 2D&3D movies watching," *Biomed Signal Process Control*, vol. 55, Jan. 2020, doi: 10.1016/j.bspc.2019.101642.

[2] M. Soufineyestani, D. Dowling, and A. Khan, "Electroencephalography (EEG) Technology Applications and Available Devices," *Applied Sciences 2020, Vol. 10, Page 7453*, vol. 10, no. 21, p. 7453, Oct. 2020, doi: 10.3390/APP10217453.

[3] M. Melek, N. Manshouri, and T. Kayikcioglu, "Low-Cost Brain-Computer Interface Using the Emotiv Epoc Headset Based on Rotating Vanes," *Traitement du Signal*, vol. 37, no. 5, pp. 831–837, Nov. 2020, doi: 10.18280/ts.370516.

[4] E. Maiorana, "Deep learning for EEG-based biometric recognition," *Neurocomputing*, vol. 410, pp. 374–386, Oct. 2020, doi: 10.1016/J.NEUCOM.2020.06.009.

[5] P. Arnau-Gonzalez, S. Katsigiannis, M. Arevalillo-Herraez, and N. Ramzan, "BED: A New Data Set for EEG-Based Biometrics," *IEEE Internet Things J*, vol. 8, no. 15, pp. 12219–12230, Aug. 2021, doi: 10.1109/JIOT.2021.3061727.

[6] M. Maleki and T. Kayikçioglu, "A new brain-computer interface system using the gaze on rotating vane.," *Biomedical Research-tokyo*, 2016.

[7] B. T. Klassen *et al.*, "Quantitative EEG as a predictive biomarker for Parkinson disease dementia," *Neurology*, vol. 77, no. 2, pp. 118–124, Jul. 2011, doi: 10.1212/WNL.0B013E318224AF8D/SUPPL_FILE/KLASSEN.PDF.

[8] C. Melissant, A. Ypma, E. E. E. Frietman, and C. J. Stam, "A method for detection of Alzheimer's disease using ICA-enhanced EEG measurements," *Artif Intell Med*, vol. 33, no. 3, pp. 209–222, Mar. 2005, doi: 10.1016/J.ARTMED.2004.07.003.

[9] E. Cainelli, L. Vedovelli, B. Carretti, and P. Bisiacchi, "EEG correlates of developmental dyslexia: a systematic review," *Annals of Dyslexia 2022 73:2*, vol. 73, no. 2, pp. 184–213, Nov. 2022, doi: 10.1007/S11881-022-00273-1.

[10] A. R. Clarke, R. J. Barry, S. J. Johnstone, R. McCarthy, and M. Selikowitz, "EEG development in Attention Deficit Hyperactivity Disorder: From child to adult," *Clinical Neurophysiology*, vol. 130, no. 8, pp. 1256–1262, Aug. 2019, doi: 10.1016/J.CLINPH.2019.05.001.

[11] R. J. Barry, A. R. Clarke, S. J. Johnstone, R. McCarthy, and M. Selikowitz, "Electroencephalogram theta/beta ratio and arousal in attention-deficit/hyperactivity disorder: evidence of independent processes," *Biol Psychiatry*, vol. 66, no. 4, pp. 398–401, Aug. 2009, doi: 10.1016/J.BIOPSYCH.2009.04.027.

[12] C. Baumgartner and J. P. Koren, "Seizure detection using scalp-EEG," *Epilepsia*, vol. 59, pp. 14–22, Jun. 2018, doi: 10.1111/EPI.14052.

[13] S. Nasehi and H. Pourghassem, "Seizure detection algorithms based on analysis of EEG and ECG signals: A survey," *Neurophysiology*, vol. 44, no. 2, pp. 174–186, Jun. 2012, doi: 10.1007/S11062-012-9285-X/METRICS.

[14] T. Rowberry *et al.*, "Implementation and Early Evaluation of a Quantitative Electroencephalography Program for Seizure Detection in the PICU*," *Pediatric Critical Care Medicine*, vol. 21, no. 6, pp. 543–549, Jun. 2020, doi: 10.1097/PCC.0000000000002278.

[15] E. Milne, R. Gomez, A. Giannadou, and M. Jones, "Atypical EEG in autism spectrum disorder: Comparing a dimensional and a categorical approach," *J Abnorm Psychol*, vol. 128, no. 5, pp. 442–452, Jul. 2019, doi:

10.1037/ABN0000436.
[16] M. E. Santarone *et al.*, “EEG Features in Autism Spectrum Disorder: A Retrospective Analysis in a Cohort of Preschool Children,” *Brain Sciences 2023, Vol. 13, Page 345*, vol. 13, no. 2, p. 345, Feb. 2023, doi: 10.3390/BRAINSCI13020345.
[17] M. M. Siddiqui, G. Srivastava, and S. H. Saeed, “Diagnosis of insomnia sleep disorder using short time frequency analysis of PSD approach applied on EEG signal using channel ROC-LOC,” *Sleep Science*, vol. 9, no. 3, pp. 186–191, Jul. 2016, doi: 10.1016/J.SLSCI.2016.07.002.
[18] C. Spironelli, M. Manfredi, and A. Angrilli, “Beta EEG band: A measure of functional brain damage and language reorganization in aphasic patients after recovery,” *Cortex*, vol. 49, no. 10, pp. 2650–2660, Nov. 2013, doi: 10.1016/J.CORTEX.2013.05.003.
[19] S. Ballanti *et al.*, “EEG-based methods for recovery prognosis of patients with disorders of consciousness: A systematic review,” *Clinical Neurophysiology*, vol. 144, pp. 98–114, Dec. 2022, doi: 10.1016/J.CLINPH.2022.09.017.
[20] J. A. Micoulaud-Franchi, C. Jeunet, A. Pelissolo, and T. Ros, “EEG Neurofeedback for Anxiety Disorders and Post-Traumatic Stress Disorders: A Blueprint for a Promising Brain-Based Therapy,” *Curr Psychiatry Rep*, vol. 23, no. 12, pp. 1–14, Dec. 2021, doi: 10.1007/S11920-021-01299-9/FIGURES/5.
[21] D. A. Moscovitch, D. L. Santesso, V. Miskovic, R. E. McCabe, M. M. Antony, and L. A. Schmidt, “Frontal EEG asymmetry and symptom response to cognitive behavioral therapy in patients with social anxiety disorder,” *Biol Psychol*, vol. 87, no. 3, pp. 379–385, Jul. 2011, doi: 10.1016/J.BIOPSYCHO.2011.04.009.
[22] E. Netzer, A. Frid, and D. Feldman, “Real-time EEG classification via coresets for BCI applications,” *Eng Appl Artif Intell*, vol. 89, p. 103455, Mar. 2020, doi: 10.1016/J.ENGAPPAI.2019.103455.
[23] M. Shen, P. Wen, B. Song, and Y. Li, “Real-time epilepsy seizure detection based on EEG using tunable-Q wavelet transform and convolutional neural network,” *Biomed Signal Process Control*, vol. 82, p. 104566, Apr. 2023, doi: 10.1016/J.BSPC.2022.104566.
[24] T. A. Milligan, “Epilepsy: A Clinical Overview,” *Am J Med*, vol. 134, no. 7, pp. 840–847, Jul. 2021, doi: 10.1016/J.AMJMED.2021.01.038.
[25] T. Adewumi, E. Oladipo, and A. O. Adewuya, “Public perception and attitude towards people living with epilepsy in Nigeria,” *Epilepsy & Behavior*, vol. 106, p. 107033, May 2020, doi: 10.1016/J.YEBEH.2020.107033.
[26] S. Ibrahim, R. Djemal, and A. Alsuwailem, “Electroencephalography (EEG) signal processing for epilepsy and autism spectrum disorder diagnosis,” *Biocybern Biomed Eng*, vol. 38, no. 1, pp. 16–26, Jan. 2018, doi: 10.1016/J.BBE.2017.08.006.
[27] W. O. Tatum *et al.*, “Clinical utility of EEG in diagnosing and monitoring epilepsy in adults,” *Clinical Neurophysiology*, vol. 129, no. 5, pp. 1056–1082, May 2018, doi: 10.1016/J.CLINPH.2018.01.019.
[28] F. Rosenow, K. M. Klein, and H. M. Hamer, “Non-invasive EEG evaluation in epilepsy diagnosis,” *Expert Rev Neurother*, vol. 15, no. 4, pp. 425–444, Apr. 2015, doi: 10.1586/14737175.2015.1025382.
[29] C. A. M. Guerreiro, “Epilepsy: Is there hope?,” *Indian J Med Res*, vol. 144, no. 5, p. 657, Nov. 2016, doi: 10.4103/IJMR.IJMR_1051_16.
[30] M. Diykh, Y. Li, and P. Wen, “Classify epileptic EEG signals using weighted complex networks based community structure detection,” *Expert Syst Appl*, vol. 90, pp. 87–100, Dec. 2017, doi: 10.1016/J.ESWA.2017.08.012.
[31] A. K. Jaiswal and H. Banka, “Epileptic seizure detection in EEG signal using machine learning techniques,” *Australas Phys Eng Sci Med*, vol. 41, no. 1, pp. 81–94, Mar. 2018, doi: 10.1007/S13246-017-0610-Y/TABLES/12.
[32] H. U. Amin, M. Z. Yusoff, and R. F. Ahmad, “A novel approach based on wavelet analysis and arithmetic coding for automated detection and diagnosis of epileptic seizure in EEG signals using machine learning techniques,” *Biomed Signal Process Control*, vol. 56, p. 101707, Feb. 2020, doi: 10.1016/J.BSPC.2019.101707.
[33] M. S. Farooq, A. Zulfiqar, and S. Riaz, “Epileptic Seizure Detection Using Machine Learning: Taxonomy, Opportunities, and Challenges,” *Diagnostics 2023, Vol. 13, Page 1058*, vol. 13, no. 6, p. 1058, Mar. 2023, doi: 10.3390/DIAGNOSTICS13061058.
[34] M. T. Akhtar, W. Mitsuhashi, and C. J. James, “Employing spatially constrained ICA and wavelet denoising, for automatic removal of artifacts from multichannel EEG data,” *Signal Processing*, vol. 92, no. 2, pp. 401–416, Feb. 2012, doi: 10.1016/J.SIGPRO.2011.08.005.
[35] V. Harpale and V. Bairagi, “An adaptive method for feature selection and extraction for classification of epileptic EEG signal in significant states,” *Journal of King Saud University - Computer and Information Sciences*, vol. 33, no. 6, pp. 668–676, Jul. 2021, doi: 10.1016/J.JKSUCI.2018.04.014.
[36] T. Zhang, W. Chen, and M. Li, “Generalized Stockwell transform and SVD-based epileptic seizure detection in EEG using random forest,” *Biocybern Biomed Eng*, vol. 38, no. 3, pp. 519–534, Jan. 2018, doi: 10.1016/J.BBE.2018.03.007.
[37] 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. 061907, Nov. 2001, doi: 10.1103/PhysRevE.64.061907.
[38] A. K. Singh and S. Krishnan, “Trends in EEG signal feature extraction applications,” *Front Artif Intell*, vol. 5, p. 1072801, Jan. 2023, doi: 10.3389/FRAI.2022.1072801/BIBTEX.
[39] I. Stancin, M. Cifrek, and A. Jovic, “A Review of EEG Signal Features and Their Application in Driver Drowsiness

Detection Systems," *Sensors (Basel)*, vol. 21, no. 11, Jun. 2021, doi: 10.3390/S21113786.

[40] H. Choubey and A. Pandey, "A combination of statistical parameters for the detection of epilepsy and EEG classification using ANN and KNN classifier," *Signal Image Video Process*, vol. 15, no. 3, pp. 475–483, Apr. 2021, doi: 10.1007/S11760-020-01767-4/TABLES/7.

[41] J. Xiang *et al.*, "Kurtosis and skewness of high-frequency brain signals are altered in paediatric epilepsy," *Brain Commun*, vol. 2, no. 1, Jan. 2020, doi: 10.1093/BRAINCOMMS/FCAA036.

[42] W. Garcia-González, W. Flores-Fuentes, O. Sergiyenko, J. C. Rodríguez-Quiñonez, J. E. Miranda-Vega, and D. Hernández-Balbuena, "Shannon Entropy Used for Feature Extractions of Optical Patterns in the Context of Structural Health Monitoring," *Entropy 2023, Vol. 25, Page 1207*, vol. 25, no. 8, p. 1207, Aug. 2023, doi: 10.3390/E25081207.

[43] A. Mert and A. Akan, "Seizure onset detection based on frequency domain metric of empirical mode decomposition," *Signal Image Video Process*, vol. 12, no. 8, pp. 1489–1496, Nov. 2018, doi: 10.1007/S11760-018-1304-Y/TABLES/2.

[44] M. L. Vicchietti, F. M. Ramos, L. E. Betting, and A. S. L. O. Campanharo, "Computational methods of EEG signals analysis for Alzheimer's disease classification," *Scientific Reports 2023 13:1*, vol. 13, no. 1, pp. 1–14, May 2023, doi: 10.1038/s41598-023-32664-8.

[45] M. M. Lansbergen, M. Van Dongen-Boomsma, J. K. Buitelaar, and D. Slaats-Willemse, "ADHD and EEG-neurofeedback: A double-blind randomized placebo-controlled feasibility study," *J Neural Transm*, vol. 118, no. 2, pp. 275–284, Feb. 2011, doi: 10.1007/S00702-010-0524-2/FIGURES/1.

[46] S. Abenna, M. Nahid, H. Bouyghf, and B. Ouacha, "EEG-based BCI: A novel improvement for EEG signals classification based on real-time preprocessing," *Comput Biol Med*, vol. 148, Sep. 2022, doi: 10.1016/J.COMPBIOMED.2022.105931.

[47] M. Melek, "Diagnosis of COVID-19 and non-COVID-19 patients by classifying only a single cough sound," *Neural Computing and Applications 2021*, pp. 1–12, Jul. 2021, doi: 10.1007/S00521-021-06346-3.

[48] M. Melek *et al.*, "An automatic EEG-based sleep staging system with introducing NAoSP and NAoGP as new metrics for sleep staging systems," *Cogn Neurodyn*, vol. 15, no. 3, pp. 405–423, Jun. 2021, doi: 10.1007/S11571-020-09641-2.

[49] T. Kayikcioglu, M. Maleki, and K. Eroglu, "Fast and accurate PLS-based classification of EEG sleep using single channel data," *Expert Syst Appl*, vol. 42, no. 21, pp. 7825–7830, Jun. 2015, doi: 10.1016/j.eswa.2015.06.010.

[50] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction," *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/S41598-022-10358-X.

[51] S. Zhang, D. Cheng, Z. Deng, M. Zong, and X. Deng, "A novel kNN algorithm with data-driven k parameter computation," *Pattern Recognit Lett*, vol. 109, pp. 44–54, Jul. 2018, doi: 10.1016/J.PATREC.2017.09.036.

[52] P. K. Syriopoulos, N. G. Kalampalikis, S. B. Kotsiantis, and M. N. Vrahatis, "kNN Classification: a review," *Ann Math Artif Intell*, pp. 1–33, Sep. 2023, doi: 10.1007/S10472-023-09882-X/METRICS.

[53] G. Mezzadri, T. Laloë, F. Mathy, and P. Reynaud-Bouret, "Hold-out strategy for selecting learning models: Application to categorization subjected to presentation orders," *J Math Psychol*, vol. 109, p. 102691, Aug. 2022, doi: 10.1016/J.JMP.2022.102691.

[54] H. Cheng, D. J. Garrick, and R. L. Fernando, "Efficient strategies for leave-one-out cross validation for genomic best linear unbiased prediction," *J Anim Sci Biotechnol*, vol. 8, no. 1, pp. 1–5, May 2017, doi: 10.1186/S40104-017-0164-6/TABLES/5.

[55] J. P. Cruz-Tirado, J. M. Amigo, D. F. Barbin, and S. Kucheryavskiy, "Data reduction by randomization subsampling for the study of large hyperspectral datasets," *Anal Chim Acta*, vol. 1209, p. 339793, May 2022, doi: 10.1016/J.ACA.2022.339793.

[56] H. Ling, C. Qian, W. Kang, C. Liang, and H. Chen, "Combination of Support Vector Machine and K-Fold cross validation to predict compressive strength of concrete in marine environment," *Constr Build Mater*, vol. 206, pp. 355–363, May 2019, doi: 10.1016/j.conbuildmat.2019.02.071.

[57] N. Manshouri, M. Melek, and T. Kayıkcıoglu, "Detection of 2D and 3D Video Transitions Based on EEG Power," *Comput J*, Sep. 2020, doi: 10.1093/COMJNL/BXAA116.

[58] D. François, F. Rossi, V. Wertz, and M. Verleysen, "Resampling methods for parameter-free and robust feature selection with mutual information," *Neurocomputing*, vol. 70, no. 7–9, pp. 1276–1288, Mar. 2007, doi: 10.1016/J.NEUCOM.2006.11.019.

[59] C. Beleites *et al.*, "Variance reduction in estimating classification error using sparse datasets," *Chemometrics and Intelligent Laboratory Systems*, vol. 79, no. 1–2, pp. 91–100, Oct. 2005, doi: 10.1016/J.CHEMOLAB.2005.04.008.

[60] S. Wong *et al.*, "EEG datasets for seizure detection and prediction— A review," *Epilepsia Open*, vol. 8, no. 2, pp. 252–267, Jun. 2023, doi: 10.1002/EPI4.12704.

[61] S. NAHZAT and M. YAĞANOĞLU, "Classification of Epileptic Seizure Dataset Using Different Machine Learning Algorithms and PCA Feature Reduction Technique," *Journal of Investigations on Engineering and Technology*, vol. 4, no. 2, pp. 47–60, Dec. 2021, Accessed: Feb. 14, 2024. [Online]. Available: https://dergipark.org.tr/en/pub/jiet/issue/67435/1002958

[62] W. Mardini, M. M. Bani Yassein, R. Al-Rawashdeh, S. Aljawarneh, Y. Khamayseh, and O. Meqdadi, "Enhanced detection of epileptic seizure using EEG signals in combination with machine learning classifiers," *IEEE Access*, vol.

8, pp. 24046–24055, 2020, doi: 10.1109/ACCESS.2020.2970012.
[63] P. Boonyakitanont, A. Lek-uthai, K. Chomtho, and J. Songsiri, "A review of feature extraction and performance evaluation in epileptic seizure detection using EEG," *Biomed Signal Process Control*, vol. 57, p. 101702, Mar. 2020, doi: 10.1016/J.BSPC.2019.101702.
[64] "EEG Database — Seizure Prediction Project Freiburg." Accessed: Feb. 14, 2024. [Online]. Available: https://epilepsy.uni-freiburg.de/freiburg-seizure-prediction-project/eeg-database/
[65] A. Malekzadeh, A. Zare, M. Yaghoobi, H. R. Kobravi, and R. Alizadehsani, "Epileptic Seizures Detection in EEG Signals Using Fusion Handcrafted and Deep Learning Features," *Sensors 2021, Vol. 21, Page 7710*, vol. 21, no. 22, p. 7710, Nov. 2021, doi: 10.3390/S21227710.
[66] W. Chen *et al.*, "An automated detection of epileptic seizures EEG using CNN classifier based on feature fusion with high accuracy," *BMC Med Inform Decis Mak*, vol. 23, no. 1, pp. 1–17, Dec. 2023, doi: 10.1186/S12911-023-02180-W/TABLES/5.
[67] S. Supriya, S. Siuly, H. Wang, and Y. Zhang, "New feature extraction for automated detection of epileptic seizure using complex network framework," *Applied Acoustics*, vol. 180, p. 108098, Sep. 2021, doi: 10.1016/J.APACOUST.2021.108098.
[68] M. Varlı and H. Yılmaz, "Multiple classification of EEG signals and epileptic seizure diagnosis with combined deep learning," *J Comput Sci*, vol. 67, p. 101943, Mar. 2023, doi: 10.1016/J.JOCS.2023.101943.

Author(s) Contributions

The article is written by a single author. I hereby declare that I have prepared the article alone for the authorship declaration.

Conflict of Interest Notice

The author declares that there are no potential conflicts of interest.

Support/Supporting Organizations

This research did not receive any specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethical Approval and Informed Consent

It is declared that during the preparation process of this study, scientific and ethical principles were followed, and all the references are provided in the bibliography. This study used the EEG dataset purchased by Andrzejak et al. [37]. and belonging to the University of Bonn. In this case, the author of the cited study did not apply for ethical approval.

Availability of data and material

The datasets analyzed during the current study are available from https://www.ukbonn.de/epileptologie/arbeitsgruppen/ag-lehnertz-neurophysik/downloads/.

Plagiarism Statement

This article has been scanned by iThenticate ™.