Summer 7-31-2021

Integrated Machine Learning Approaches to Improve Classification performance and Feature Extraction Process for EEG Dataset

Mohammad Masum

Integrated Machine Learning Approaches to Improve Classification performance and Feature Extraction Process for EEG Dataset

by

Mohammad Masum

Thesis Advisors: Dr. Hossain Shahriar and Dr. Hisham Haddad
Committee Member: Dr. Jing (Selena) He

Presented to the Faculty of the College of Computing and Software Engineering of Kennesaw State University in Partial Fulfillment of the Requirements for the degree of

DOCTOR OF PHILOSOPHY

Kennesaw State University

July 2021

DocuSign Envelope ID: 3C53F7F0-B4E9-4F12-9BB3-109F94591512

KENNESAW STATE UNIVERSITY
COLLEGE OF GRADUATE AND PROFESSIONAL EDUCATION

Dissertation Defense Outcome

Name Mohammd Masum KSU ID 000823105

Email mmasum@students.kennesaw.edu Phone Number 470-900-8624

Program Analytics and Data Science, PhD

Title

Integrated Machine Learning Approaches to Improve Classification Performance and Feature Extraction Process for EEG Dataset

Dissertation Defense: Date 6/29/2021

☑ Passed ☐ Failed ☐ Passed With Revisions (attach revisions)

Signatures

		Date
Dissertation Chair	DocuSigned by: Hossain Shahriar; DocuSigned by: Hisham Haddad	July 1, 2021
Committee Member	DocuSigned by: [signature]	July 2, 2021
Committee Member		
Committee Member		
Committee Member		
Program Director	DocuSigned by: Jennifer Priestley	July 2, 2021
Department Chair	DocuSigned by: Sherrill Hayes	July 2, 2021
Graduate College Approval	DocuSigned by: Nikki Palamiotis	July 2, 2021

Updated 8/24/20

Dedicated to my dearest mother Tajkera Begum and father Hossain Ahmed.

They never stopped showering me with their perpetual love and affection.

My father passed away 22^{nd} February 2010. His memories inspired me to accomplish this lofty academic goal.

Abstract

Epileptic seizure or epilepsy is a chronic neurological disorder that occurs due to brain neurons' abnormal activities and has affected approximately 50 million people worldwide. Epilepsy can affect patients’ health and lead to life-threatening emergencies. Early detection of epilepsy is highly effective in avoiding seizures by intervening treatment. The electroencephalogram (EEG) signal, which contains valuable information of electrical activity in the brain, is a standard neuroimaging tool used by clinicians to monitor and diagnose epilepsy. Visually inspecting the EEG signal is an expensive, tedious, and error-prone practice. Moreover, the result varies with different neurophysiologists for an identical reading. Thus, automatically classifying epilepsy into different epileptic states with a high accuracy rate is an urgent requirement and has long been investigated. This PhD thesis contributes to the epileptic seizure detection problem using Machine Learning (ML) techniques.

Machine learning algorithms have been implemented to automatically classifying epilepsy from EEG data. Imbalance class distribution problems and effective feature extraction from the EEG signals are the two major concerns towards effectively and efficiently applying machine learning algorithms for epilepsy classification. The algorithms exhibit biased results towards the majority class when classes are imbalanced, while effective feature extraction can improve classification performance.

In this thesis, we presented three different novel frameworks to effectively classify epileptic states while addressing the above issues. Firstly, a deep neural network-based framework exploring different sampling techniques was proposed where both traditional and state-of-the-art sampling techniques were experimented with and evaluated for their capability of improving the

imbalance ratio and classification performance. Secondly, a novel integrated machine learning-based framework was proposed to effectively learn from EEG imbalanced data leveraging the Principal Component Analysis method to extract high- and low-variant principal components, which are empirically customized for the imbalanced data classification. This study showed that principal components associated with low variances can capture implicit patterns of the minority class of a dataset. Next, we proposed a novel framework to effectively classify epilepsy leveraging summary statistics analysis of window-based features of EEG signals. The framework first denoised the signals using power spectrum density analysis and replaced outliers with k-NN imputer. Next, window level features were extracted from statistical, temporal, and spectral domains. Basic summary statistics are then computed from the extracted features to feed into different machine learning classifiers. An optimal set of features are selected leveraging variance thresholding and dropping correlated features before feeding the features for classification.

Finally, we applied traditional machine learning classifiers such as Support Vector Machine, Decision Tree, Random Forest, and k-Nearest Neighbors along with Deep Neural Networks to classify epilepsy. We experimented the frameworks with a benchmark dataset through rigorous experimental settings and displayed the effectiveness of the proposed frameworks in terms of accuracy, precision, recall, and F-beta score.

Acknowledgements

First and foremost, I want to thank the Almighty Allah for giving me the strength to reach this far.

I want to express my heartfelt gratitude to my committee members. Firstly, I am very thankful to my committee chair, Dr. Hossain Shahriar, who has advised me over the last two years. His desire to continually watch out for my welfare, as well as his support in research, and job hunts, is highly appreciated. Next, I would like to thank my co-advisor, Dr. Hisham Haddad, for his advice, particularly for reviewing several manuscripts. I would also like to thank Dr. Selena He for her guidance in putting together final dissertation.

I want to convey my heartfelt thanks to Dr. Jennifer Priestley, who has been a tremendous supporter of mine over the last four years and has made this all possible. Her guidance and assistance in job searches are most appreciated. I want to also thank Dr. Mohammed Chowdhury for his supports and suggestions in job search.

Finally, I acknowledge and thank my family and friends for their endless love, support, and prayers. I would like to express my appreciation to my wife Faiza Aktar for her unwavering supports and encouragement. Without my family’s constant support, encouragement, and understanding it would not have been possible to achieve this academic goal.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

This chapters discusses an overview of this thesis. We briefly discuss the problem definition, objectives of the study, proposals to solutions and finally, the outline of the dissertation.

1.1 Motivation

Epilepsy or Epileptic Seizure is a common chronic neurological disorder affecting approximately 50 million people worldwide, with over 100 million patients experiencing a seizure at least once in their lifetime [1]. Experiencing more than one seizure is one of the primary symptoms of epilepsy, while the consequences vary based on the starting location of the seizure in the brain. Seizures can occur unexpectedly and can cause sudden breakdown affecting motor, sensory, and automatic functions of the body, leading to disturbing the patients' consciousness, cognition, and memory [2]. Accurately seizure detection enables medical professionals to monitor seizures and diagnose epilepsy, which is still a challenging task for researchers [3].

The Electroencephalogram (EEG) has long been used to investigate electrical activities of the brain and diagnose epilepsy due to its affordable cost and efficiency in temporal resolution of long-term monitoring [2]. EEG evaluates voltage variations across electrodes throughout subjects' scalp leveraging ionic currents flowing through brain neurons, providing temporal and geographical information regarding electrical activities in the brain [4, 5]. Thus, the underlying patterns contained in an EEG signal during seizure differ from the patterns contained in non-epileptic persons' EEG signal [4]. Consequently, analyzing and developing models based on the EEG data allows detect seizure and classify different epileptic states: normal, pre-ictal and inter-ictal stages. An EEG signal recorded from a healthy person is the normal phase, while an EEG recorded preceding a seizure and during a seizure refers to pre-ictal and inter-ictal stages of epilepsy, respectively. Distinguishing among different states of epilepsy using EEG data lead to predict the onset of seizures [6]. A visual scanning of EEG signal is one of the traditional practices by clinicians to classify different categories of epileptic states. However, visual scanning for long EEG readings is an expensive, time-consuming, error-prone exercise and is neurophysiologists dependent [2]. Therefore, developing an automatic and effective model for epilepsy detection using EEG signals is an urgent need.

Classification of epileptics' states is a well-known challenge for more than 30 years. The unpredictability of seizures hamper the management of chronic epilepsy. Recent effort emphasized on seizure detection using EEG signal obtained from real patient [4]. With the advancements of machine learning algorithms, many sophisticated and automatic systems have been implemented to improve the performance of EEG-based seizure detection. Classical ML approaches like Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), Naïve Bayes, and Decision Tree (DT) along with Deep Neural Network (DNN) have been applied to classify epilepsy [6, 7]. Despite the prevalence of ML techniques, machine learning algorithms experience biased results towards majority class and reduced performance when it comes to train imbalanced data, making the epilepsy detection a critical challenge.

A dataset is said to be imbalanced or skewed if there are relatively or significantly a smaller number of training instances in one class compared to the other class for a binary class classification problem. The class that contains more observations is called the majority class while the other class containing relatively or significantly less observations is called the minority class [8]. Many of the real-world datasets are imbalanced such as the epileptic seizure dataset that presented and experimented in this work. Class imbalance involves difficulties in learning since most of the ML classifiers are biased towards the majority class [8]. Thus, the correct prediction for minority class can be significantly dropped. Therefore, developing an automatic method addressing the inherent class imbalance problem towards epilepsy classification is necessary.

On the other hand, effective feature extraction from the EEG signal data plays an important role to improve the classification performance of ML classifiers since the classifiers, in general, are used the extracted features to train [9, 10]. The feature extraction process reduces dimensionality and complexity of the data, provides interpretability to the model by extracting meaningful features, and improves model performance towards epilepsy classification [11]. Feature extraction is the process of defining a feature vector from a regular vector (e.g., EEG signal or a segment of EEG signal) where the features are distinctive measurements or structural components of the regular vector [6]. Time-domain, frequency-domain, and time-frequency domain features are generally extracted from EEG signal for epilepsy classification. Finally, an effective feature extraction process facilitates model development, provides interpretability, and improve performance towards epilepsy classification.

1.2 Objectives

The primary objective of this thesis is to develop an automatic and effective machine learning-based novel framework for epilepsy classification addressing the issues mentioned earlier. Therefore, in the process, our goal is to–

1. Address the inherent imbalanced class problem in different epileptic states of the EEG data by exploring combinations of different sampling techniques and different ML classifiers.

2. Address the inherent imbalanced class problem by leveraging high- and low-variants principal components of the original EEG data and showing that principal components associated with low-variants can learn the underlying pattern of the minority class of the data.
3. Develop an effective feature extraction process to reduce the dimensionality and provide meaningful features to feed the ML classifiers, leading to overall performance improvement.

1.3 Contributions

In this thesis, we presented three different novel ML-based frameworks to achieve objectives 1, 2 and 3, respectively. The contributions are listed below in sequence of the objectives:

1. We presented an integrated machine learning approach for epilepsy detection to effectively learn from imbalanced data by experimenting several sampling techniques and evaluating their capability of improving the imbalance ratio. Different classical ML classifiers, along with a deep neural network-based framework, were applied with different class ratio, indicating performance improvement in classifying seizures [12].
2. We presented an integrated machine learning approach for epilepsy detection that can effectively learn from imbalanced data. The approach utilizes PCA at the first stage to extract both high- and low-variant principal components, which are empirically customized for imbalanced data classification. We hypothesized that principal components associated with low variances can capture the implicit pattern of minor class of a dataset and can contribute to improving the performance of a model [13].

3. We proposed an integrated ML-based epilepsy classification framework involving a novel feature extraction process. We pre-processed the signals by denoising and imputing outliers. We extracted summary statistics of window-based statistical, temporal, and spectral features. Feature selection criteria are also applied to select an optimal set of discriminative features. We showed the effectiveness of our proposed method by comparing the classification performance to other recent advanced studies [14].

Finally, this thesis can aid practitioners in adopting a low-cost model of classification with stable and high accuracy in the obtained results to apply in the clinical practice and research environment.

1.4 Dissertation Outline

This thesis is divided into six chapters. Chapter 1 introduces an overview of the subject, motivation, objectives, and proposals of the thesis. Chapter 2 discusses related works. Chapter 3 presents a work explaining objective 1, titled- Analysis of Sampling Techniques Towards Epileptic Seizure Detection from Imbalanced Dataset. Chapter 4 presents a work explaining objective 2, titled- Epileptic Seizure Detection from Imbalanced Dataset using an Integrated Machine Learning Approach. Chapter 5 presents a work explaining objective 2, titled- A Statistical Summary Analysis of Window-Based Extracted Features for EEG Signal Classification. Chapter 6 concludes by discussing the findings, limitations, and future directions.

Chapter 2: Related Work

This chapters discusses related works to this study. We proposed three different novel frameworks to achieve the objectives as mentioned earlier. The discussion of the related works is included successive order to the objectives and proposals:

2.1 Related works for Analysis of Sampling Techniques Towards Epileptic Seizure Detection from Imbalanced Dataset

The primary goal of seizure prediction is to identify a time when seizures are likely approaching and occurring. In general, the duration of non-seizure periods in an EEG recording is too long; on the contrary, the seizure signal lasts for a few seconds, resulting in the EEG data becomes imbalanced [65]. Therefore, the real-world epilepsy detection dataset suffers from a class imbalance problem causing less performance in prediction. Though ML algorithms have been efficiently used in the healthcare area, the algorithms have shown reduced performance when training with imbalanced data, making epilepsy detection a critical challenge [66]. Therefore, researchers proposed in the literature several methods for handling imbalanced datasets.

Many sampling techniques including undersampling, oversampling, and combined approaches have been applied to overcome class imbalance problem for improving classification performance. A multiple layer of intelligent signal classifier for brain EEG data was proposed where in the first layer an oversampling technique, SMOTE, was used to solve the class imbalance problem, and finally different machine learning classifiers was applied for epilepsy detection [35]. A simple technique was proposed where the imbalanced dataset at first is converted to a balanced dataset using under sampling, oversampling, and synthetic minority oversampling technique (SMOTE), and SVM was applied later to classify the class of the epilepsy [19].

A weighted Extreme Learning Machine (ELM) method was proposed for seizure detection with imbalanced EEG data distribution [20]. KMeans method combined with SVM was applied to breast cancer diagnosis and showed improved performance in terms of G-mean and accuracy metrics [32]. Simple oversampling was applied to each of the clusters of KMeans to balance the data, and later SVM was applied for classification. SMOTE was applied to oral cancer and erythemato-squamous diseases dataset to produce a balanced training data that lead to better accuracy score for the classification problem [33]. A combination of KMeans and Boosted C5.0 was proposed for prediction of imbalanced breast cancer data where KMeans clusters observations from both minority and majority classes and subsequently select similar number of samples from each of the clusters to deal with class imbalance of the data [34].

In this work, we experimented with different combinations of sampling techniques and ML methods to handle imbalanced data sets' problem and to achieve better classification. We utilized the real-world Epileptic Seizure Recognition dataset with these combinations. The experimental results show the effectiveness of using different sampling techniques.

2.2 Related works for Epileptic Seizure Detection from Imbalanced Dataset Using and Integrated Machine Learning Approach

The primary goal of seizure prediction is to identify a time when seizures are likely approaching and when they are occurring. Earlier works focused on Frequency-based methods, nonlinear dynamics (Chaos), and statistical analysis of EEG signals to predict seizures [10, 62]. These approaches rely on transforming input signals using mathematical transformations (e.g., Fourier transform). Binary programming and dynamic system approach to predict seizures are explored in [63, 64].

In recent years, machine learning algorithms have been applying in seizure classification using the EEG data and showing promising results. A combined approach was proposed to predict seizure status by extracting the features from the EEG signal using discrete wavelet transform method and later used the features as input to SVM classifier for classification of the signal [15]. A two-layer seizure detection classifier was proposed wherein the upper layer, a dimension reduction technique was used, and SVM was then applied to assign the class of epilepsy [16]. PCA, Independent Components Analysis (ICA), and Linear Discriminant Analysis (LDA) was applied to reduce the dimension of the data. An Improved Correlation-based Feature Selection method (ICFS) combined with RF classifier was proposed for detecting epilepsy status from EEG signals [17]. The ICFS method primarily applied to the EEG dataset to extract most important time domain-, frequency domain-, and entropy-based features that are consequently fed into RF classifier. Deep learning networks have been receiving increasing attention in epilepsy classification problems for improving model performance. A deep learning neural network was implemented on extracted frequency domain features from EEG signals [18].

Machine learning algorithms exhibit reduced performance when training imbalanced data, making epilepsy detection a critical challenge. Researchers proposed several methods for handling imbalanced datasets. A simple technique was proposed where the imbalanced dataset at first converted to a balanced dataset by using under sampling, oversampling, and synthetic minority oversampling technique, and then SVM was applied to classify the class of the epilepsy [19]. A weighted extreme learning machine (ELM) was proposed for seizure detection with imbalanced EEG data distribution [20]. Our proposed integrated ML method can handle imbalanced data sets' problem by utilizing high- and low-variant principal components in feature extraction process. We experimented our model by applying it to a real-world Epileptic Seizure Recognition dataset. The experimental results show the robustness and effectiveness of our model.

2.3 Related works for A Statistical Summary Analysis of Window-Based Extracted Features for EEG Signal Classification

Significant research has been accomplished to correctly classify EEG signals for epilepsy, where many combined approaches, including feature extraction methods and ML classifiers, are proposed. A correlation-based feature selection method (ICFS) combined with Random Forest (RF) classifier was proposed for detecting epileptic states, where time-, frequency- and entropy-based features were extracted [2]. A two-layer seizure detection classifier was proposed wherein the upper layer, a dimension reduction technique, including principal component analysis, independent component analysis, and linear discriminant analysis, were applied, and SVM was then applied to assign the class of epilepsy [46]. A random forest classifier with grid search hyper-parameter tuning was applied to extract features (e.g., mean, energy, and standard deviation) of the Bonn dataset for epilepsy detection [47]. Another recent study based on the Bonn dataset extracted altogether 15 statistical features from the EEG signal, followed by a correlation-based feature selection method for epilepsy classification [48]. The study applied five different classifiers (RF, Logistic Tree Model, k-NN, SVM, and NB), where the RF classifier provided the best accuracy. A multi-feature fusion approach was presented where an ensemble decision tree classifier was applied to a selected number of features for epilepsy classification using EEG signal data [49]. The fusion approach applied a Pearson correlation-based feature selection method on the extracted temporal (5-features), spectral (5-features), and temporal-spectral (6-features) features. A combined approach was developed for EEG classification where an SVM classifier was applied to temporal and spectral features that were extracted using empirical mode decomposition [50].

Many fusion approaches were proposed for EEG classification, where different ML classifiers were applied to extracted features from the original signal. This work presented a novel framework for EEG classification extracting window-based features and considered summary statistics of those features. In addition, we implemented a rigorous signal preprocessing step before the feature extraction and a feature selection process after the extraction. Finally, we applied different advanced ML classifiers to classify epileptic states to the Bonn dataset.

Chapter 3: Analysis of Sampling Techniques Towards Epileptic Seizure Detection from Imbalanced Dataset

This chapter discusses the impact of sampling techniques with varying class size ratios towards balancing the imbalanced dataset and overall epilepsy classification. This work was proposed to achieve the goal of the thesis associated with objective 1. It contains an introduction, methodology, experiments and results, and a conclusion section.

3.1 Introduction

This work investigates different sampling techniques along with ML classifiers to effectively balance and train the imbalanced data and improve performance for seizure detection. In the first stage, we applied different traditional and state-of-the-art sampling techniques, such as SMOTE, ADASYN, Random undersampling, Random Oversampling, SMOTEENN, SMOTETomek, Cluster centroids, and NearMiss to the original dataset and generated a number of new datasets. Random oversampling, and Random undersampling are conventional, easy to implement techniques. However, these techniques suffer from overfitting, and loss of valuable information problems, respectively. Therefore, some state-of-the-art sampling techniques were examined in this work.

Different combinations of sampling methods were applied to the original dataset to generate new datasets, which are then fed into several ML classifiers to measure the performance of the prediction. The comparisons among different combinations of datasets and ML methods are presented to show the effectiveness of applying sampling techniques for prediction performance. Our results indicate that it is possible to predict more accurately seizure from EEG data with an oversampled or undersampled dataset, instead of using the original data. Furthermore, this work can aid practitioners to adopt a more accurate model of classification with stable and high accuracy in the obtained results to apply in the clinical practice and research environment.

3.2 Methodology

3.2.1 Undersampling techniques

Undersampling techniques balance the class distribution for a classification problem dataset by eliminating observations from the majority class of the training dataset. Many undersampling methods have been implemented to balance the class ratio of the data. Random undersampling, NearMiss, Edited nearest neighbors, and cluster centroids are the undersampling techniques that are explored in this work.

Random undersampling

It is the simplest method of balancing the imbalanced data by randomly removing observations from the majority class. The method may lead to the loss of valuable information about the data since it randomly eliminates data from the majority class [8]. If observations of majority class are close to each other, then the method might produce a good performance for classification problems [27].

NearMiss

NearMiss sampling technique performs undersampling on the majority class of the data by considering the distance of a data point in majority class to the data points in the minority class. Three different versions of this method are applied to balance class ratio: 1. NearMiss-1: Data points in majority class are eliminated that have a minimum average distance to *k* number of data points in minority class, where *k* is a hyperparameter. 2. NearMiss-2: Contrary to the NearMiss-1, NearMiss-2 drops the majority class data points that have the maximum average distance to k number of data points in the minority class where k is a hyperparameter. 3. NearMiss-3: It only remains the data points in the majority class that are on the decision boundary, *i.e.*, the data points with the lowest distance to each of the data points of the minority class are only kept. Hence, the size of the majority class directly controlled by the number of the minority class.

Cluster centroid

It is another undersampling technique that forms *n* number of clusters of data points in the majority class at first where n is a tunable parameter and then replaces the data points of a cluster by the cluster centroid. The method leverages the KMeans clustering technique in the process of clustering.

Edited Nearest Neighbor (ENN)

It is another similarity based undersampling technique that removes observation from the majority class if class of the observation differs from one of its nearest neighbors.

3.2.2 Oversampling techniques

Oversampling is a non-heuristic technique to balance the class distribution of training data. Random oversampling, Synthetic Minority Oversampling, and Adaptive synthetic sampling techniques are explored and implemented in this work.

Random oversampling

It is a simple oversampling technique that randomly replicates observations of minority class repetitively until the desired class ratio is achieved. It is the most used sampling technique among practitioners due to the simplicity and ease of application [28]. However, the major drawback is that it suffers from overfitting because of the random duplication of data [29].

Synthetic Minority Oversampling Technique (SMOTE)

SMOTE was proposed based on the similarities between data points of minority class to overcome the limitation of the random oversampling. Instead of random duplication, the method generates new synthetic observations by linearly interpolating between a randomly selected observation from minority class and its closest k observations in the minority class where k is a hyperparameter [28]. Assume, x is the randomly selected data from the minority class and y is one of the k-nearest neighbors of x. Then, a synthetic data z is generated by interpolating x and y:

$$z = x + w(y - x)$$

where x and y are vectors and w is a random weight in [0,1]. However, the method suffers from some downsides as well though it effectively overcomes the limitation of random oversampling. The method may introduce noise and within-class class imbalance if the randomly selected data from the minority class located among the majority class observations [30].

Adaptive synthetic (ADASYN) sampling

ADASYN utilizes a similar idea of SMOTE, and additionally distributes weights to the minority class samples according to their complexity level in the training process. The samples that are difficult to learn are given more weights and more synthetic data are generated from those samples as well. Hence, applying ADASYN lower the bias introduced by the imbalance class and shifts the classification decision boundary towards the harder samples.

3.2.3 Mixed sampling techniques

Combination of undersampling and oversampling techniques are applied to handle imbalance class ratio problem. The combined approaches can overcome some of the limitations of the separate sampling approaches. Two of the combined approaches SMOTEENN, and SMOTETomek are discussed.

SMOTEENN

SMOTEENN is a combined approach to balance the class ratio of an imbalanced dataset. It utilizes SOMTE to over samples from the minority class and simultaneously, implements ENN to eliminates samples from both classes [31]. Hence, it provides in depth data cleaning.

SMOTETomek

It is another combining of oversampling and undersampling methods. SMOTE is applied to minority class for over sampling, but it may introduce noise and class imbalance in the dataset as discussed earlier. Hence, to overcome the challenges of SMOTE, a data cleaning method Tomek links is applied to the over-sampled synthetic samples. Tomek links is one of the neighbor-based undersampling techniques where Tomek links are formed by pairs of opposite class observations who are their own nearest neighbors. In the process of SMOTETomek, it removes Tomek links containing observations from both classes [31]. Subsequently, the method produces a balance data with well-defined class clusters.

3.3 Machine learning classifiers

Once the class imbalance problem is solved, machine learning classifiers are applied to the balanced dataset for classification. Here, we investigated Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, and Deep neural network classifiers for epilepsy detection.

3.3.1 Support Vector Machine

SVM is a well-known supervised learning technique to analyze high-dimensional data. SVM searches for an optimal hyperplane in the input space that categorizes two classes given training data. Therefore, the hyperplane is used to classify new data [39].

3.3.2 Decision Tree

Decision Tree is a well-known supervised machine learning technique for classification. It builds a classification model in the shape of a tree structure through a process known as binary recursive partitioning [7]. It iteratively splits the data into smaller and smaller subsets (branches) until each of the branches achieves homogeneous partitions. Therefore, it finally creates a tree with decision nodes and leaf nodes where the decision nodes contain two or more branches and leaf node assigns a class or decision.

3.3.3 Random Forest

Random forest is an ensemble of multiple individual decision trees. In the training period, it produces a class prediction for each of the decision trees and the class with the majority votes becomes the methods' prediction class [25].

3.3.4 Neural Network

At present, neural networks are widely used for many applications due to the capability of highly non-linear systems and flexibility in architecture design. The neural network's basic architecture contains input layers, one or more hidden layers, and output layers where each of the layers includes a certain number of neurons. Weighted linear combination of neurons of a layer is computed and then used as input to another neuron in the succeeding layer. To capture the nonlinearity of the data, a non-linear function, called activation function, can be applied to the weighted sums of neurons. All the weights of a neural network are set to random values at the initial stage of training. Data is fed into the input layer of the network, then it travels through the hidden layers, and finally output is produced in the output layer. The network continually updates the weights applying backpropagation based on the output and desired target of the neural network. The network consequently reduces the error between the output and target in each iteration [17]. In the process, a loss function is used to calculate the error of the network and the error is minimized by applying optimization function during backpropagation.

3.4 Experiments and Results

3.4.1 Dataset specification

We performed all experiments on Epileptic Seizure Recognition dataset to evaluate model performance for using different combination of sampling techniques and machine learning algorithms. The dataset is publicly available on UCI's machine learning repository [21]. The dataset represents a recording of brain activity which includes 4097 EEG readings over 23.6 seconds for a single subject/patient, with 25 patients overall. Each patients' 4097 readings were then divided and shuffled into 23 chunks where each chunk contains 178 readings for 1 second. Each of the 23 chunks of a single patient were then translated into one row of the dataset where each row contains 178 columns (readings). Collectively, there are $23 \times 500 = 11{,}500$ rows, and 178 columns in the dataset. The response variable contains five different categories: 1. Healthy and Eyes Open, 2. Healthy and Eyes Closed, 3. Epileptic, Inter-ictal, 4. Epileptic, Inter-ictal, and 5. Epileptic, Ictal. The patients in category 5 (Epileptic, Ictal) have epileptic seizures, and patients falling in the rest of the classes did not have an epileptic seizure with distinctive characteristics.

Each of the classes contains 20% data of the total dataset. We transformed classes 2,3,4, and 5 (no having seizure) into a single class to prepare the dataset for binary classification. Hence, the dataset became imbalanced and consists of two classes: class 1 (Epileptic seizure), and class 0 (no seizure) where class 1 contains 20% of data and rest 80% of the data are in class 0.

3.4.2 Model evaluation metrics

The dataset is become imbalanced after the transformation. Therefore, we ought not to consider the "accuracy" metric to assess the performance of the models. Thus, the following performance measurements are considered in the assessment of the models [10].

1. Recall: Recall is the the quantity of correct positive predictions among all the positive samples. Mathematically:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Where, TP is True Positive (quantity of correct positive predictions) and FN is False Negative (quantity of misclassified positive predictions)

2. Precision: Precision is the proportion of the correctly identified positives to all the predicted positives. Mathematically:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

3. F_1 score: F_1 score is the harmonic mean of Precision and Recall. F_1 score is a better performance metric than the accuracy metric for imbalanced data [10].

$$\text{F}_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

The F-beta score is the weighted harmonic mean of precision of recall where F-beta value at 1 means perfect score (perfect precision and recall) and 0 is worst.

$$F_{\beta} = (1 + \beta^2) \frac{Precision \ \times Recall}{(\beta^2 \times \ Precision) + Recall}$$

When $\beta = 1$, F-beta is F_1 $score$. The β parameter determines the weight of precision and recall. $\beta < 1$ can be picked, if we want to give more weight to precision, while $\beta > 1$ values give more weight to recall. Since we want to identify maximum number of seizure cases, we give more weights to recall and utilize $\beta > 1$ values. Hence, the F-beta score was considered the principal performance metric to evaluate models in our experiments.

3.4.3 Experimental Design

Sampling techniques were used to generate new datasets. Eight different sampling techniques in total were implemented with attaining three different class ratios. The sampling techniques include Random undersampling, NearMiss, and cluster centroids as undersampling techniques. Random oversampling, SMOTE, and ADASYN as oversampling techniques and SMOTEENN, and SMOTETomek as combined approach were implemented on the epilepsy dataset for balancing class ratio. 0.5, 0.75, and 1 class ratio between majority and minority class was attained for all eight different techniques. Altogether $3 \times 8 = 24$ datasets were used. Machine learning algorithms: Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression (LR), and Deep Neural Network (DNN) based framework were applied to the 24 variants of datasets for classifying epileptic seizure. In total, $5 \times 24 = 110$ were experimented in this work. All the experiments were performed with 10-fold cross validation.

Table 1:Distribution of samples in majority and minority classes

Ratio	# of sample in No seizure	# of samples in Seizure
0.5	9200	4600
0.75	9200	6900
1.00	9200	9200

Table 2: Performance evaluation of original dataset

Classifiers	F-beta	Precision	Recall
RF	0.966	0.942	0.888
DT	0.939	0.862	0.823
SVM	0.972	0.956	0.904
LR	0.817	0.966	0.890
DNN	0.970	0.955	0.894

The original data consists of 9,200 of no seizure and remaining 2,300 seizure samples. In the process of oversampling from minority class and undersampling from majority class, we opted to attain 0.5, 0.75, and 1.00 ratios between minority and majority class. We wanted to minimum number of synthetic or duplicate data since both kinds of techniques have some limitations. Hence, we select the final model that achieve maximum performance by including minimum number of synthetic or duplicate data and maximum. Table 1 shows the number of samples in minority and majority classes when different ratios are picked.

All the datasets were randomly split into training and test data while maintaining the class ratio between seizure and no seizure samples. The training data was used to train each of the models we experimented with while the test data was used for evaluating the performance of the models. To verify the consistency of the model, we experimented with each of the models with 10-fold cross-validation. The SVM, DT, LR, and RF were implemented using Python scikit-learn library with default hyperparameter options.

The DNN consists of four layers: one input layer, two hidden layers, and one output layer. We used 'ReLu' activation function in the hidden layer and 'sigmoid' function in the output layer to train the DNN. 'Adam' and 'binary cross-entropy' were used for optimizer and loss function respectively. We implemented an early stopping method to stop training once the model performance stops improving on the test data. The initial learning rate was set to 0.001 with a decay of $1e - 5$ in every epoch. All the parameters and hyperparameters used in the model were optimized by grid search. The 'beta' parameter in calculating the F-beta score was set to 50 to give more weight to recall so that the maximum number of seizures can be identified.

The experiments are carried out on a Windows 10 Intel(R) Core (TM) i7-8565U CPU 1.80 GHz with 16.0 GB RAM and NVIDIA GeForce MX250 2GB GDDR5. We implemented our experiment on Keras framework in Python 3.7 version [22].

3.4.4 Experimental Results

F-beta score, precision, and recall were used to evaluate the models' performance. We applied five different classifiers (SVM, RF, DT, LR, and DNN) on 24 different datasets to detect seizures. 10-fold cross-validation was performed for each of the experiments. The same configuration was applied to each experiment for maintaining consistency.

All the classifiers were trained on 90% of data and tested on the remaining 10% of the data. Table 2 illustrates the experimental results of using the original dataset without any sampling techniques implementation. Table 2 shows the results for using original dataset without any sampling techniques. SVM achieves the maxum F-beta score of 0.972 while our proposed DNN based framework achieved the second maximum F-beta score of 0.970. Though LR shows highest precisoin, but it fails to attain a good recall score, i.e., it poorly predicts on the seizure cases.

Tables 3, 4, and 5 illustrate the experimental results of using different combinations of sampling techniques and ML classifiers when class ratio between minority and majority classes is 0.5, 0.75, and 1.00, respectively. From Table 3: RF, SVM, and our presented DNN based framework demonstrate similar kind of performance with insignificant margin. The NearMiss undersampling technique with SVM achieves the highest F-beta (0.996), precision (0.999), and recall (0.989) indicating that more seizures can be correctly detected by using sampling techniques instead of original dataset. On the other hand, LR fails to provide good performance for with any combination with different ML classifiers. Random oversampling with DNN based framework shows second best F-beta score of 0.990. From Table 4, NearMiss with RF classifier outperforms other combinations in terms of F-beta and precision score while highest recall score is achieved by Random oversampling + RF. Our presented DNN-based framework achieves rewarding performance as well. All the experiments in Table 5 were done with equal number of observations in each class. Once again, NearMiss undersampling method outclasses other sampling techniques. NearMiss combines with RF achieves the highest F-beta and recall score. Overall, sampling the original dataset for balancing the class ratio helps ML classifiers to better learning from the minority class and result in better performance in epileptic seizure detection.

Table 3: Experimental results of using different combination of sampling techniques and ML classifiers when class ratio is 0.5.

Class Ratio: 0.5									
Classifiers	Random Oversampling			SMOTE			ADASYN		
	F-beta	Precision	Recall	F-beta	Precision	Recall	F-beta	Precision	Recall
RF	0.983	0.972	0.978	0.970	0.956	0.953	0.962	0.944	0.943
DT	0.956	0.917	0.955	0.929	0.894	0.892	0.916	0.871	0.886
SVM	0.971	0.969	0.945	0.967	0.966	0.936	0.963	0.955	0.937
LR	0.725	0.958	0.183	0.719	0.946	0.168	0.700	0.931	0.130
DNN	0.979	0.972	0.967	0.970	0.967	0.944	0.970	0.956	0.957
Classifiers	Random Undersampling			Cluster Centroids			NearMiss		
	F-beta	Precision	Recall	F-beta	Precision	Recall	F-beta	Precision	Recall
RF	0.957	0.946	0.924	0.945	0.935	0.899	0.992	0.995	0.983
DT	0.917	0.885	0.864	0.896	0.859	0.826	0.966	0.958	0.942
SVM	0.962	0.965	0.921	0.946	0.954	0.882	**0.996**	**0.999**	**0.989**
LR	0.711	0.964	0.141	0.708	0.940	0.133	0.724	0.997	0.173
DNN	0.961	0.958	0.925	0.952	0.953	0.900	0.990	0.997	0.973
Classifiers	SMOTEENN			SMOTETomek					
	F-beta	Precision	Recall	F-beta	Precision	Recall			
RF	0.976	0.967	0.954	0.966	0.954	0.943			
DT	0.941	0.899	0.910	0.928	0.894	0.890			
SVM	0.977	0.974	0.950	0.967	0.966	0.935			
LR	0.749	0.960	0.180	0.718	0.952	0.162			
DNN	0.985	0.978	0.973	0.973	0.965	0.952			

Table 4: Experimental results of using different combination of sampling techniques and ML classifiers when class ratio is 0.75.

Class Ratio: 0.75									
Classifiers	Random Oversampling			SMOTE			ADASYN		
	F-beta	Precision	Recall	F-beta	Precision	Recall	F-beta	Precision	Recall
RF	0.987	0.977	**0.993**	0.970	0.962	0.968	0.962	0.946	0.966
DT	0.970	0.947	0.986	0.929	0.916	0.918	0.910	0.891	0.904
SVM	0.971	0.974	0.958	0.969	0.972	0.955	0.965	0.957	0.963
LR	0.691	0.906	0.313	0.686	0.875	0.322	0.662	0.799	0.289
DNN	0.983	0.978	0.983	0.974	0.970	0.970	0.971	0.961	0.973
Classifiers	Random Undersampling			Cluster Centroids			NearMiss		
	F-beta	Precision	Recall	F-beta	Precision	Recall	F-beta	Precision	Recall
RF	0.957	0.047	0.951	0.928	0.933	0.897	**0.992**	0.993	0.988
DT	0.911	0.905	0.886	0.869	0.857	0.834	0.963	0.966	0.947
SVM	0.959	0.970	0.935	0.927	0.955	0.870	0.996	0.999	0.989
LR	0.670	0.909	0.225	0.656	0.833	0.246	0.696	0.984	0.297
DNN	0.960	0.960	0.946	0.938	0.951	0.903	0.987	**0.999**	0.972
Classifiers	SMOTEENN			SMOTETomek					
	F-beta	Precision	Recall	F-beta	Precision	Recall			
RF	0.974	0.969	0.969	0.970	0.962	0.969			
DT	0.936	0.924	0.926	0.927	0.914	0.916			
SVM	0.972	0.978	0.959	0.967	0.971	0.953			
LR	0.690	0.878	0.305	0.682	0.872	0.304			
DNN	0.982	0.979	0.978	0.977	0.972	0.975			

Table 5: Experimental results of using different combination of sampling techniques and ML classifiers when class ratio is 1.

Class Ratio: 1.00									
Classifiers	Random Oversampling			SMOTE			ADASYN		
	F-beta	Precision	Recall	F-beta	Precision	Recall	F-beta	Precision	Recall
RF	0.990	0.982	0.997	0.970	0.962	0.968	0.965	0.953	0.977
DT	0.976	0.958	0.995	0.927	0.911	0.921	0.911	0.906	0.917
SVM	0.973	0.978	0.968	0.968	0.971	0.955	0.972	0.962	0.982
LR	0.646	0.698	0.516	0.638	0.679	0.522	0.617	0.637	0.534
DNN	0.985	0.980	0.989	0.974	0.973	0.976	0.975	0.967	0.0983
Classifiers	Random Undersampling			Cluster Centroids			NearMiss		
	F-beta	Precision	Recall	F-beta	Precision	Recall	F-beta	Precision	Recall
RF	0.952	0.949	0.954	0.923	0.937	0.910	**0.991**	0.993	**0.990**
DT	0.897	0.908	0.887	0.854	0.871	0.831	0.963	0.970	0.955
SVM	0.955	0.971	0.938	0.908	0.953	0.859	0.995	0.999	0.990
LR	0.590	0.629	0.443	0.571	0.593	0.457	0.658	0.746	0.480
DNN	0.960	0.969	0.95	0.926	0.944	0.907	0.987	**0.998**	0.976
Classifiers	SMOTEENN			SMOTETomek					
	F-beta	Precision	Recall	F-beta	Precision	Recall			
RF	0.975	0.972	0.978	0.972	0.967	0.978			
DT	0.933	0.928	0.938	0.928	0.925	0.932			
SVM	0.972	0.979	0.966	0.966	0.974	0.958			
LR	0.640	0.683	0.521	0.630	0.669	0.516			
DNN	0.981	0.980	0.983	0.979	0.975	0.983			

Fig. 1, 2, & 3 display the comparative result (Recall) analysis using boxplots with varying balancing ratios. We focus on comparing the recall score since minimizing the false negative relatively significant for epilepsy classification. Fig. 1 shows that the NearMiss sampling technique with RF classifier achieved the highest recall score for balancing ratio 0.50, while random oversampling with the RF classifier outperformed other combinations of sampling and classifiers for ratio 0.75 (Fig. 2). For the equal sample ratio between majority and minority class, the combination of random oversampling and random forest classifier achieved the highest recall score (Fig. 3). Fig. 4 exhibits the changes in performance (recall) while changing the balancing ratios. The performance of the random forest classifier increases while the ratios between majority and minority classes increase. The RF classifier achieved a recall score of 88.8 with the original dataset where the balancing ratio is 0.20, while the classifier achieved the highest recall score of 99.7 when the size of the minority class data is equal to the majority class size (ratio: 1.00).

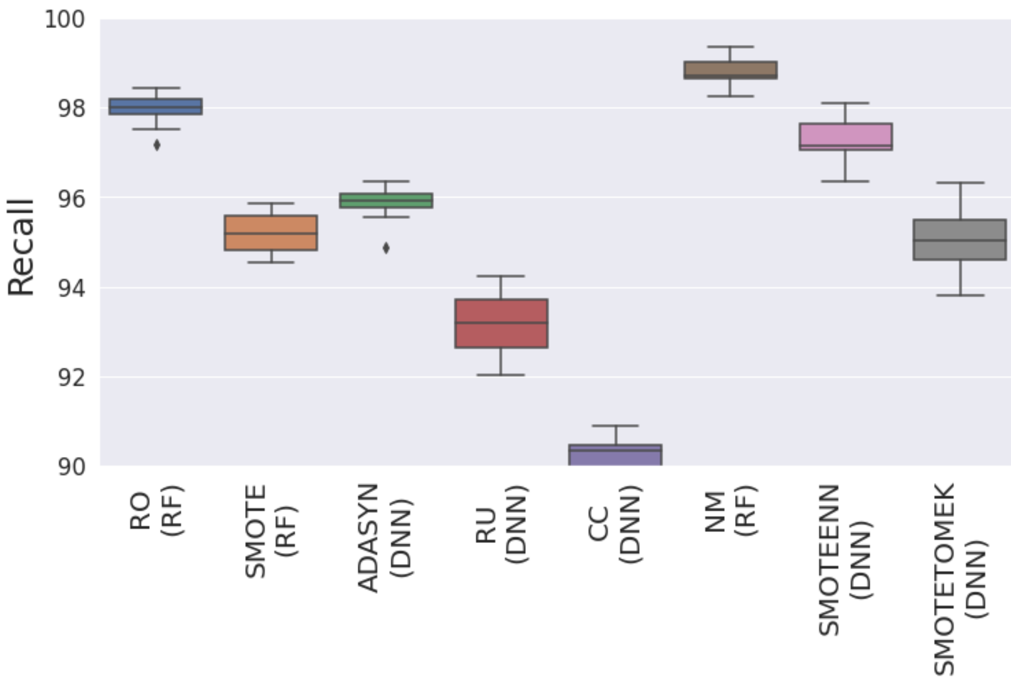

Figure 1: Comparative result analysis using boxplots for balancing ratio 0.50.

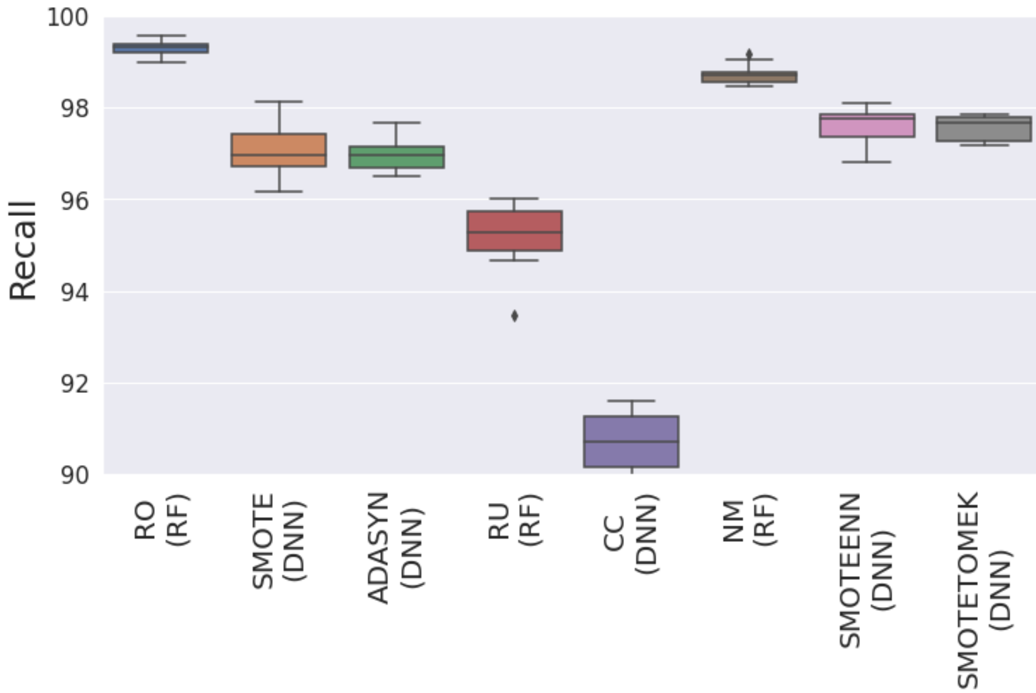

Figure 2: Comparative result analysis using boxplots for balancing ratio 0.75.

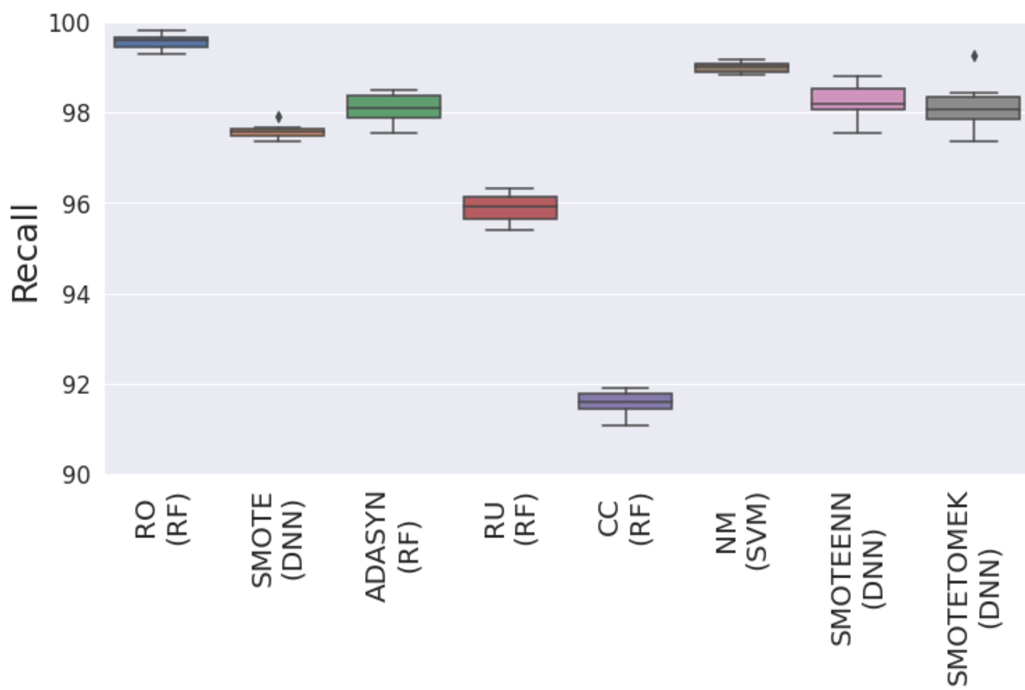

Figure 3: Comparative result analysis using boxplots for balancing ratio 1.00.

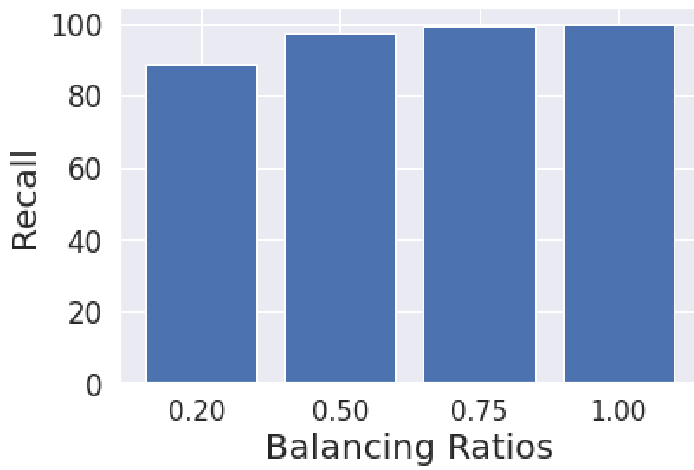

Figure 4: Changes of performance (Recall) for different balancing ratios

3.5 Conclusion

In this study, we applied an integrated machine learning approach for epilepsy detection that can effectively learn from imbalanced data. In this work, several sampling techniques have been experimented and evaluated for their capability of improving the imbalance ratio. Different classical machine learning classifiers along with a deep neural network-based framework are applied to all the new datasets that indicate performance improvement in classifying seizures. The NearMiss undersampling technique outperforms other sampling techniques while RF, SVM, and DNN demonstrate similar results. Finally, sampling techniques can be applied to imbalance dataset for balancing class ratio to improve the classification performance.

Chapter 4: Epileptic Seizure Detection for Imbalanced Datasets using an Integrated Machine Learning Approach

This chapter explores high-and low-variant principal components to cope with inherent imbalanced class distribution problem, leading to an effective epilepsy classification model. This work was proposed to achieve the goal of the thesis associated with objective 2. It contains an introduction, methodology, experiments and results, and a conclusion section.

4.1 Introduction

We developed a new integrated analysis technique of PCA and ML classifiers to effectively train imbalanced data and improve performance for seizure detection. In the first stage, PCA was applied to the original dataset and extract both the high- and low-variant attributes or components. Conventionally, PCA is used for dimension reduction of a dataset leveraging principal components (PCs) with high variances. In this work, we show that PCs associated with low variances can capture the implicit pattern of minor class of a dataset. Our assumption is that the high variant PCs may effectively learn the underlying structure of the majority class of a dataset, but they may not be enough to represent the implicit pattern of minor class of the dataset. Based on the hypothesis, our proposed method selects both high- and low-variant PCs and combine them subsequently.

Different Combinations of high- and low-variant PCs are then fed into several ML classifiers and measure the performance of the prediction. The comparisons between selective components and all attributes in the original dataset show a wide difference in terms of performance in prediction. Our results indicate that it is possible to predict more accurately seizure from EEG data with a limited and selective number of attributes/components, instead of all attributes in the original data. Our contributions in this work include: (1) a novel integrated ML approach that can handle imbalanced data, and (2) a comprehensive assessment with rigorous experimental setting to assess our proposed models' performance with a publicly available real-world epileptic seizure detection dataset. Further, the experimental results show that the statistical significance of our proposed model. Finally, our work can aid practitioners to adopt a fast and low-cost model of classification with stable and high accuracy in the obtained results to apply in the clinical practice and research environment.

4.2 Methodology

Fig. 5 demonstrates the architecture of our proposed method. The method consists of two stages: in the first stage high- and low-variant features are extracted by applying PCA on the original EEG data. The extracted features associated with high variance are then combine with different chunks of low-variant components. The construction of chunks is described in Section IV(C). In the second step, the combination of high-low variant features is fed into different machine learning classifiers that classify the label of the dataset.

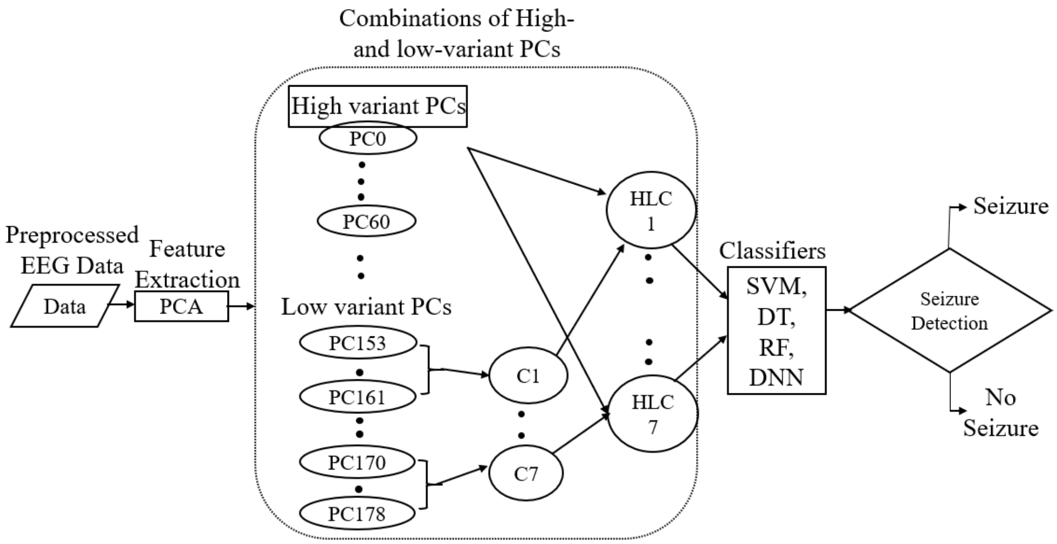

Figure 5: Architecture of our proposed Framework

4.3 Experiments and Results

4.3.1 Dataset Specification

We performed all experiments on Epileptic Seizure Recognition dataset to evaluate our proposed integrated approach. The description of the dataset is already previously discussed in Chapter 2 dataset specification section.

4.3.2 Model Evaluation Metrics

The dataset is become imbalanced after the transformation. Therefore, we ought not to consider the "accuracy" metric to assess the performance of the models. Thus, we utilize weighted F-beta score to measure the model performance and more weights were given to recall ($\beta > 1$, β is the weight parameter in F-beta score) to identify maximum number of seizure cases, Wilcoxon Rank Sum test is applied to evaluate statistical significance of the model performance.

4.3.3 Experimental Design

We extracted high- and low- variant principal components by applying PCA on the original dataset. The data was normalized with mean 0 before applying the PCA. The number of principal components is 178.

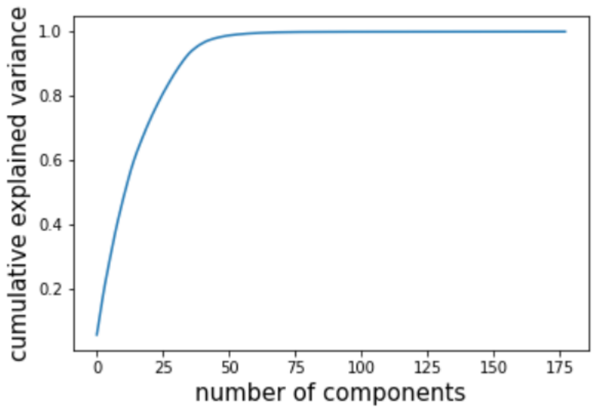

Figure 6: Cumulative variation explanation of original data

Table 6: Process of constructing combinations of high- and low-variant principal components.

Chunks	*PCs*	*Combinations*	*High- and Low-variant PCs*
C1	170-178	HLC1	1st 60 PCs + C1
C2	161-169	HLC2	1st 60 PCs + C2
C3	152-160	HLC3	1st 60 PCs + C3
C4	143-151	HLC4	1st 60 PCs + C4
C5	134-142	HLC5	1st 60 PCs + C5
C6	161-178	HLC6	1st 60 PCs + C6
C7	143-160	HLC7	1st 60 PCs + C7

as the number of features is 178 in the original dataset. Fig. 6 shows the cumulative variation explanation of original data by principal components. Approximately 99% of the variation of the original dataset is explained by the first 60 PCs. We considered rest of the PCs are associated with low variances. To experiment empirically, we took last 25% of the PCs and divided into 5 chunks where each chunk contains 5% of PCs. For instance, the 170-178 PCs last are the last 5% PCs which is a single chunk. We also added two more chunks where each fold contains 10% of PCs of last 20% PCs. In total, 7 folds of low variance PCs were then combined with the high-variant first 60 PCs that result in 7 different datasets. Table 6 shows the chunks and different combinations of datasets. For example, the first combination HLC1 is made up of the first 60 PCs and 170-178 PCs.

The 7 different datasets of PCs consequently fed into different ML classifiers like SVM, RF, DT, and DNN. Finally, the different combined datasets performance was evaluated by comparing with original dataset's performance.

All the datasets were randomly split into training and test data while maintaining the class ratio between seizure and no seizure samples. The training data was used to train each of the models we experimented with while the test data was used for evaluating the performance of the models. To verify the consistency of the model, we experimented with each of the models with 10-fold cross-validation. The SVM, DT, and RF were implemented using Python scikit-learn library with default hyperparameter options. The DNN consists of four layers: one input layer, two hidden layers, and one output layer. We used 'ReLu' activation function in the hidden layer and 'sigmoid' function in the output layer to train the DNN. 'Adam' and 'binary cross-entropy' were used for optimizer and loss function, respectively. We implemented an early stopping method to stop training once the model performance stops improving on the test data. The initial learning rate was set to 0.001 with a decay of $1e-5$ in every epoch. All the parameters and hyperparameters used in the model were optimized by grid search.

The 'beta' parameter in calculating the F-beta score was set to 50 to give more weight to recall so that the maximum number of seizures can be identified. We implemented our experiment on Keras framework in Python 3.7 version [22].

4.3.4 Experimental Results

F-beta score, FPR, and TPR are used to evaluate the models' performance. We applied four different classifiers (SVM, RF, DT, and DNN) on 8 different datasets including the original dataset to detect seizures. 10-fold cross-validation was performed for each of the experiments. The same configuration was applied to each experiment for maintaining consistency.

All the classifiers were trained on 90% of data and tested on the remaining 10% of the data. Table 7 illustrates the experimental results of using different datasets. The four classifiers achieved maximum F-beta score by using combination of high- and low-variant dataset. SVM achived highest F-beta score of 0.9786 by using HLC3 (1st 60 high-variant PCs + low-variant PCs of 152-160) combination. RF shows a substantial imrpovement of F-beta score for using high- and low-variant PCs. RF achieved 97.31% F-beta score using HLC4 combination while 92.84% F-beta score was achieved by using the original dataset. DT and DNN achieved maximum F-beta score of 95.06% and 97.34% by using HLC5 and HLC7 dataset, respectively. RF, and DT classifiers show considerable improvement of performance for using combination high- and low-variant PCs compare to other two classifiers.

Tables 8 and 9 illustrate the results of TPR and FPR, respectively. From Table 8, maximum TPR was achieved by using high- and low-variant combinations for all the four classifier which shows the effectiveness of our proposed model. The highest TPR of 95.08% was achieved by applying SVM on HLC5 dataset. From Table 9, the lowest FPR of 0.91% was achieved by DNN for HLC6 combination. Though SVM shows lowest FPR for original dataset, other two classifiers RF and DT present better FPR for high- and low-variant combination of PCs. Table 10 presents the statistical significance of the performance of the models. SVM, RF, and DT demonstrate a statistically better F-beta score for using different combinations than using original dataset considering 0.05 significance level. DNN does not show statistically significant imrpovement in terms of F-beta score.

Table 7: Experimental results of different classifiers on different datasets using F-beta score as performance matric.

Datasets	***SVM***	***RF***	***DT***	***DNN***
Original	0.9747	0.9284	0.9387	0.9709
HLC1	0.9784	0.9705	0.9484	0.9727
HLC2	0.9784	0.9711	0.9483	0.9713
HLC3	**0.9786**	0.9719	0.9472	0.9721
HLC4	0.9784	**0.9731**	0.9486	0.9716
HLC5	0.9786	0.9702	**0.9506**	0.9726
HLC6	0.9783	0.9691	0.9477	0.9718
HLC7	0.9781	0.9699	0.9498	**0.9734**

Table 8: Experimental results of different classifiers on different datasets using TPR as performance matric.

Datasets	***SVM***	***RF***	***DT***	***DNN***
Original	0.9182	0.9008	0.8330	0.8965
HLC1	0.9504	0.9139	0.8443	0.9013
HLC2	0.9500	0.9178	0.8508	0.8960
HLC3	0.9491	0.9117	0.8556	0.9004
HLC4	0.9495	0.9143	0.8552	0.9004
HLC5	**0.9508**	0.9108	**0.8586**	0.9073
HLC6	0.9469	0.9121	0.8578	**0.9082**
HLC7	0.9465	**0.9221**	0.8500	**0.9082**

Table 9: Experimental results of different classifiers on different datasets using FPR as performance matric.

Datasets	***SVM***	***RF***	***DT***	***DNN***
Original	**0.0111**	0.0848	0.0346	0.0110
HLC1	0.0142	0.0166	0.0277	0.0109
HLC2	0.0143	0.0154	0.0277	0.0104
HLC3	0.0143	0.0157	0.0266	0.0104
HLC4	0.0143	0.0150	0.0272	0.0107
HLC5	0.0143	0.0158	0.0283	0.0110
HLC6	0.0136	0.0150	0.0276	**0.0091**
HLC7	0.0139	**0.0147**	**0.02771**	0.0117

Table 10: Statistical significance of classifiers using different datasets.

Datasets	***SVM***	***RF***	***DT***	***DNN***
Original vs. HLC1	0.041*	0.0001*	0.001*	0.650
Original vs. HLC2	0.041*	0.0001*	0.002*	0.545
Original vs. HLC3	0.034*	0.0001*	0.003*	0.405
Original vs. HLC4	0.041*	0.0001*	0.008*	0.570
Original vs. HLC5	0.034*	0.0001*	0.000*	0.198
Original vs. HLC6	0.041*	0.0001*	0.000*	0.705
Original vs. HLC7	0.058	0.0001*	0.000*	0.212

**Statistical significance considering 0.05 significance level.*

4.4 Conclusion

In this chapter, we present an integrated machine learning approach for epilepsy detection that can effectively learn from imbalanced data. The approach utilizes PCA at the first stage to extract both high- and low-variant principal components (PCs) which are empirically customized for imbalanced data classification. We hypothesized that PCs associated with low variances can capture the implicit pattern of minor class of a dataset and can contribute to improving the performance of a model. We experimented with different combinations of high- and low-variant components on the Epileptic Seizure Recognition dataset to evaluate our proposed model. The experimental results show statistically significant performance improvement that strongly support our hypothesis.

Chapter 5: A Statistical Summary Analysis of Window-Based Extracted Features for EEG Signal Classification

This chapter presents an effective feature extraction process from EEG signal dataset for epilepsy classification. This work was proposed to achieve the goal of the thesis associated with objective 3. It contains an introduction, methodology, experiments and results, and a conclusion section.

5.1 Introduction

We propose an automatic and effective framework for EEG signal classification towards epilepsy, where we mainly leverage summary statistics analysis of window-based statistical, temporal, and spectral features. A recent experimental study showed that window-based feature extraction outperformed traditional feature extraction from original signal [45]. Our contributions to the proposed frameworks are:

1. Window-based features extraction from statistical, temporal, and spectral domains
2. Applying a robust signal pre-processing step including denoising signals with power spectrum density analysis, identifying outliers with the z-score method, and replacing outliers with k-NN imputer
3. Summary statistics analysis of window level features
4. Developing ML classifiers with a significantly smaller number of meaningful features compare to original signals
5. A rigorous experimental setting to assess the performance of the proposed framework with a benchmark epileptic seizure dataset (University of Bonn)

Finally, this work can aid practitioners in adopting a low-cost model of classification with stable and high accuracy in the obtained results to apply in the clinical practice and research environment.

5.2 Methodology

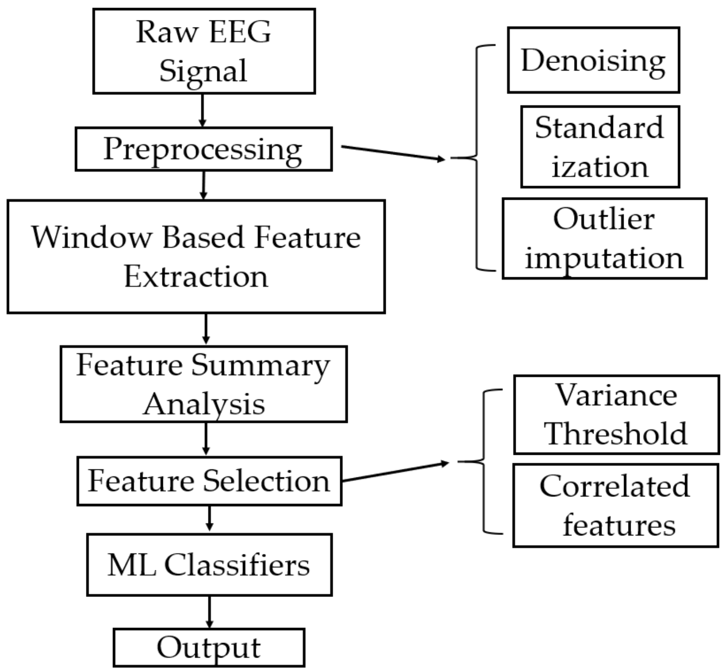

Figure 7: Flowchart of the proposed framework for EEG classification

Fig. 7 shows the flowchart of the proposed framework. The raw EEG signals are pre-processed using three different processes: denoising, standardization, and outlier imputation. Since the original EEG signals are recorded on human scalps using sensors, they are prone to noise (e.g., EEG artifacts) and may have a low signal-to-noise ratio [43]. Thus, denoising is a necessary step to be taken before the signals are analyzed to reveal the characteristics of EEG signals. Though both Wavelets and Fourier transformation have been using for transforming signals into power spectrum, an experimental study showed the superiority of Fourier transformation in noise analysis

[53]. Thus, we applied a power spectrum threshold denoising method using Fast Fourier Transformation (FFT). Notably, power per frequency is calculated in the power spectrum by applying FFT to the raw EEG signal. A threshold is used in the power spectrum to keep all the frequencies with large power (spectra) and zero out all other frequencies related to low power. Finally, the inverse FFT is applied to achieve a cleaned and filtered signal. Fig. 8 demonstrates the process of the denoising process using the power spectrum threshold (green line) method.

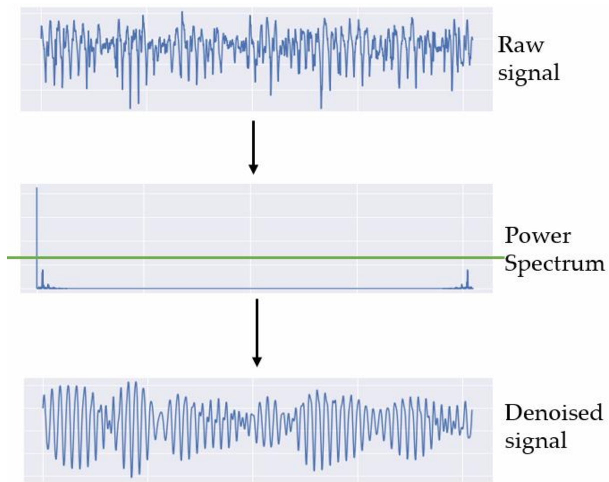

Figure 8: Process of the denoising process using the power spectrum threshold method.

In the second step of pre-processing, the signal is standardized using the z-score standardization method. Outliers are identified using the z-score and then replaced using the k-Nearest Neighbors (KNN) imputation technique.

A non-overlapping sliding window is used to segment the filtered EEG signal, and then different features in statistical, spectral, and temporal domains are extracted from each of the segments. Basic summary statistics (e.g., mean, mode, median, minimum, maximum, and standard deviation) are calculated from features of all segments of a single signal. The summary statistics of the window-based features can capture more implicit and consistent patterns of the signals. Fig. 9 shows the process of collecting summary statistics (e.g., mean, median, minimum, maximum, and standard deviation) as features from each of the signals.

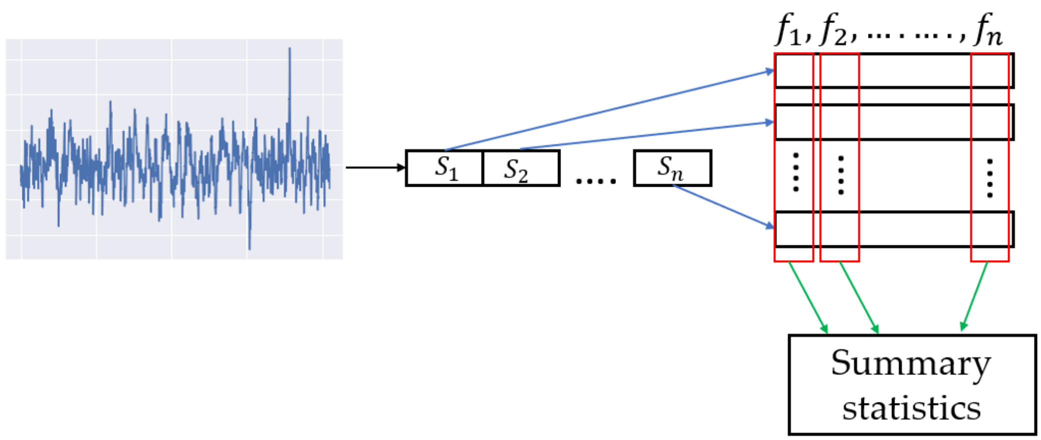

Figure 9: Process of generating summary statistics from window-based features of EEG signal.

The new set of features are then passed through a feature selection process where variance thresholding and correlated feature methods are applied. Feature's– variance lower than a threshold and higher than a correlation coefficient is removed. Finally, ML classifiers are applied to the selected features for epilepsy classification.

5.3 Experiments and Results

5.3.1 Dataset Specification

The EEG database used in this analysis consists of five EEG datasets (Set A-E) and developed by the Department of Epileptology, University of Bonn [51]. Each dataset contains 100 single channels to represent recordings of brain activities, where each channel includes 4097 EEG readings over 23.6 seconds. Fig. 10 shows five random samples from each of the datasets.

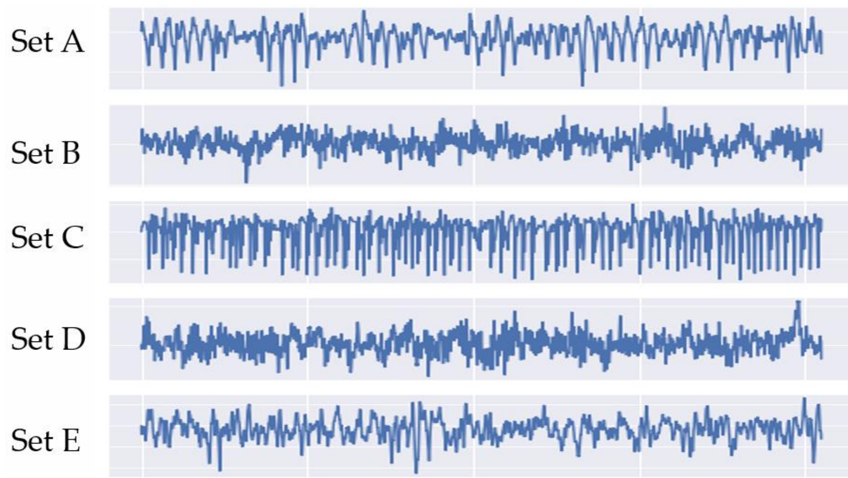

Figure 10: Visualization of EEG signals from each of the datasets. Sets (A-D) display samples of no seizure EEG signal, while Set E displays a seizure EEG signal.

While selecting the signals, different artifacts such as muscle activities and eye movements were considered for quality check. Table 11 displays summary of the five EEG data with patients' state, electrode type and placements, and the number of channels [2]. Set A and Set B contain surface EEG readings of five healthy awake volunteers with eyes open and closed, respectively. On the other hand, Sets C, D, and E contain EEG readings of five epileptic patients with state seizure-free (inter-ictal) for sets C and D and seizure activity (ictal) for set E. Sets A and B recording are captured by the international 10-20 electrode placement scheme. Set C readings are captured by placing electrode opposite to epileptogenic zone, while the recordings of Sets D and E are captured by placing electrodes within epileptogenic zone.

Table 11: Summary of the Bonn Dataset

Set	Patient State	Electrode Type	Electrode Placement	Channels
A	Healthy, Awake, and Eyes Open	Surface	International 10-20 system	100
B	Healthy, Awake, and Eyes Closed	Surface	International 10-20 system	100
C	Epileptic, Inter-ictal	Intracranial	Opposite to epileptogenic zone	100
D	Epileptic, Inter-ictal	Intracranial	Within epileptogenic zone	100
E	Epileptic, Ictal	Intracranial	Within epileptogenic zone	100

In this work, we considered three classification cases for epilepsy classification– Case 1: Healthy (AB) vs. Seizure (E); Case 2: inter-ictal (CD) vs. ictal (E); and Case 3: non-seizure (ABCD) vs. seizure (E)

5.3.2 Model Evaluation Metrics

Different performance evaluation metrics– accuracy, precision, recall, and F-beta score, are used to assess the models' performance which are mainly used in biomedical research. Recall accounts for the proportion of correctly classified ictal out of total ictal samples, while precision is the proportion of correctly classified non-ictal out of the total number of non-ictal samples. F-beta score is the weighted harmonic mean of sensitivity and specificity, and accuracy is the ratio of correctly classified EEG signals vs. the total number of EEG signals.

5.3.3 Experimental Design

We evaluated the proposed framework for EEG signal classification using one of the benchmark datasets: The University of Bonn. The raw EEG signals of the dataset were denoised using the power spectral density analysis. A threshold value of 10 was selected to filter the signals. The z-score standardization was applied to each signal for standardizing the values and identify outliers of the signal. A z-score value outside of 3-standard deviation was considered as an outlier and replaced with neighbor values using a k-NN imputation technique, where the value of $k = 3$ was chosen for imputation. As the EEG signal contains 4097 readings over 23.6 seconds, where each second consists of 178 data points, we segmented the signal where each segment is a 1 seconds of EEG readings. Conventional ML classifiers such as Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and k-Nearest Neighbors (k-NN) cannot be directly applied to the original signal as these methods do not consider temporal dependencies while training the model. Hence, features can be extracted to feed into the classifiers. We extracted statistical, temporal, and spectral features from each segment of a signal. TSFEL, a python package, was used for non-overlapping window-level feature extraction from the signals [52]. Table 12 shows the list of statistical, spectral, and temporal features that are extracted for analysis.

Each feature contains 23 different values generated by 23 segments of a signal. Summary statistics− mean, median, minimum, maximum, and standard deviation, are calculated for each of the features.

Table 12: List of extracted statistical, spectral, and temporal features

Domain	Features
Statistical	FFT mean coefficient, Wavelet absolute mean, Wavelet standard deviation, Wavelet variance, Spectral distance, Fundamental frequency, Maximum frequency, Median frequency, Spectral maximum peaks, Maximum Power Spectrum, Spectral Centroid, Decrease, Kurtosis, Skewness, Spread, Slope, Variation, Spectral Roll-off, Roll-on, Human Range Energy, MFCC, LPCC, Power Bandwidth, Spectral Entropy, Wavelet Entropy, and Wavelet Energy
Temporal	Autocorrelation, Centroid, Mean absolute differences, Mean differences, Median absolute differences, Median differences, Distance, Sum of absolute differences, Total energy, Entropy, Peak to peak distance, Area under the curve, Absolute energy, Maximum peaks, Minimum peaks, Slope, Zero crossing rate,
Spectral	Histogram, Interquartile range, mean absolute deviation, Median absolute deviation, Root mean square, Standard deviation, Variance, ECDF percentile count, ECDF slope, Kurtosis, Skewness, Maximum, Minimum, Mean, Median, ECDF, ECDF, and Percentile

We applied feature selection methods on the overall features as some of the features are unnecessary due to highly correlated to each other and low variances. Pairwise correlation of features is computed using the Pearson method, and then, highly correlated features are removed applying a threshold. Variance threshold method is applied to the features to drop all the low variant features considering a threshold value. Optimal thresholds 0.98 and 0.80 were selected using grid-search hyperparameter tuning for correlation coefficients and variance, respectively. Using the features selection process, the number of features significantly reduced to 151 from 4097 features of the original signal. Moreover, the selected features are meaningful to further analyze the EEG signal. Finally, the optimal subset of features was fed into ML classifiers for EEG classification.

Ten-fold cross validation was performed to check performance consistency, where each fold contains 90% of the signals as training data and remaining 10% signals as test data. Class ratio was preserved in the process of splitting the data into training and testing in all three cases that we experimented. Moreover, training parameters were used to standardize both training and test data to avoid data leakage. The training data was used to train each of the models, while the test data was used for evaluating the performance of the models.

5.3.4 Experimental Results

Accuracy, F-beta score, Precision, and Recall are used to evaluate the models' performance. We applied four different ML classifiers (SVM, RF, DT, and k-NN) on the University of Bonn dataset for EEG classification. 10-fold cross-validation was performed for each of the experiments. The same configuration was applied to each experiment to maintain consistency. The hyperparameters for the ML classifiers were selected using the grid search technique. Our focus is on recall rates as our goal is to minimize the number of false negatives. Table 13 displays the experimental results of applying different ML classifiers on the Bonn Dataset. RF, SVM, and k-NN classifiers produced identical results (accuracy:99.7%; recall:99.9% & F-beta: 99.9%), while the DT classifier performed slightly less. For Case 1: the RF classifier outperformed the other three classifiers in terms of accuracy and recall by producing an accuracy and recall score of 98.4% and 98.0%, respectively. k-NN achieved the highest precision score of 98.1% for Case 2. RF classifier achieved the maximum accuracy and recall score (98.8% and 97.0%, respectively) for Case 3, while the nearest accuracy 98.4% and recall 96.0% achieved by the k-NN method.

We evaluated effectiveness of the proposed framework by comparing it to some other advanced methods where the Bonn datasets were used for experiments. Table 14 illustrates comparative results of our proposed framework and other existing methods. The experimental results demonstrated that the proposed framework achieved higher accuracies than most of the listed methods. For instance, the proposed framework achieved second highest accuracy for Case 3 (ABCD vs. E) among the listed 9 recent studies. For Case 1 (AB vs. E): the proposed framework jointly achieved the second highest accuracy (99.7%) with [20], while our framework ranked fourth considering accuracy for Case 2 (CD vs. E).

Table 13: Performance comparison of different classifiers for EEG classification

	Case 1				Case 2				Case 3			
	Acc.	Recall	Precision	F-beta	Acc.	Recall	Precision	F-beta	Acc.	Recall	Precision	F-beta
RF	99.7 ± 0.01	99.0 ± 0.02	1.0 ± 0.0	99.99 ± 0.00	98.4 ± 0.01	98.0 ± 0.02	97.4 ± 0.05	97.4 ± 0.05	98.8 ± 0.01	97.0 ± 0.04	96.99 ± 0.02	96.99 ± 0.02
DT	99.0 ± 0.01	98.0 ± 0.02	99.05 ± 0.02	99.05 ± 0.02	97.3 ± 0.02	96.0 ± 0.05	96.2 ± 0.03	96.2 ± 0.03	98.2 ± 0.02	95.0 ± 0.08	96.05 ± 0.04	95.0 ± 0.08
SVM	99.7 ± 0.01	99.0 ± 0.02	1.0 ± 0.0	99.99 ± 0.00	98.0 ± 0.01	97.0 ± 0.02	97.04 ± 0.02	97.04 ± 0.02	98.2 ± 0.01	93.4 ± 0.07	97.04 ± 0.02	97.04 ± 0.02
k-NN	99.7 ± 0.01	99.0 ± 0.02	1.0 ± 0.0	99.99 ± 0.00	97.6 ± 0.02	95.0 ± 0.03	98.1 ± 0.03	98.1 ± 0.03	98.4 ± 0.01	96.0 ± 0.04	96.04 ± 0.02	96.04 ± 0.02

Table 14: Comparison between proposed and other methods

Studies	Classifiers	Cases	Accuracy
Swami et al. [15] (2016)	DNN	Case 1 Case 3	99.2 95.2
Zhang et al. [16] (2016)	SVM	Case 3	98.9
Wang et al. [26] (2018)	RF	Case 1 Case 2 Case 3	100 98.2 98.5
Singh et al. [18] (2018)	DNN	Case 1 Case 2 Case 3	89.0 99.3 95.6
Mursalin et al. [9] (2019)	RF	Case 1 Case 2 Case 3	98.6 96.2 96.9
Raghu et al. [17] (2019)	DNN	Case 1 Case 2 Case 3	97.1 96.8 97.2
Gupta et al. [19] (2019)	LS-SVM	Case 2 Case 3	99.0 98.6
Mamli et al. [20] (2019)	SVM	Case 1 Case 2 Case 3	99.7 99.6 97.4
This Study	RF	Case 1 Case 2 Case 3	99.7 98.4 98.8

5.4 Conclusion

This chapter presented a novel classification framework involving feature extraction, feature selection, and employing ML classifiers for automatically and effectively classifying EEG signals. We pre-processed the signals by denoising using power spectrum density analysis and imputing outliers with k-NN methods. Moreover, we extracted summary statistics of window-based statistical, temporal, and spectral features. Feature selection criteria: variance thresholding and correlated features removal are also applied to select an optimal set of discriminative features. We showed the effectiveness of our proposed method by comparing the classification performance to other recent advanced studies. Finally, our work can aid practitioners in adopting a fast and low-cost classification model with stable and high accuracy in the results of the clinical practice and research environment.

Chapter 6: Conclusion

Epileptic seizure or epilepsy is a chronic neurological disorder that occurs due to brain neurons' abnormal activities and has affected approximately 50 million people worldwide. Automatically classify epilepsy into different epileptic states with a high accuracy rate is an urgent requirement. Though, machine learning algorithms has long been studied, imbalance class distribution problems and effective feature extraction from the EEG signals are still major challenges toward developing effective epilepsy classification methods. In this thesis, we presented three different novel frameworks to effectively classify epileptic states addressing the challenges by exploring state-of-the-art sampling techniques, empirically customizing the high- and low-variant principal components of PCA and leveraging summary statistics analysis of window-based features of EEG signals, respectively. The first two analysis contribute to effective epilepsy classification explaining imbalanced class distribution problem, while the third study exhibited the impact of effective feature extraction process. The experimental results demonstrated the effectiveness of our proposed novel, integrated machine learning-based, frameworks using a benchmark- University of Bonn EEG dataset.

In this thesis, our experimentation was limited to University of Bonn EEG dataset containing a single channel. An extended research with other EEG datasets containing multiple channels can be performed to support our findings. Automatic channel selection is an active research field, which was not included in this study as the dataset we used contained single channel. In future, we can extend our research using multiple channel EEG datasets: Upenn and Mayo Clinics' seizure detection dataset, and CHB-MIT Scalp EEG dataset [60, 61].

In the first study, we were limited to experiment different ratio size empirically, for oversampling the minority class or undersampling the majority class. Similarly in the second study, we empirically selected the cut for high-and low-variant principal components. These limitations can be avoided in future studies by developing an automatic selection criterion. We plan to automate the selection criteria using the Bayesian Optimization technique.

Interpretability of machine learning is crucial for predictive analytics as it explains why the model operates. Though machine learning methods have been successfully applied, providing interpretability towards epilepsy classification yet to be explored. We plan to develop a window-based interpretable machine learning framework that can provide valuable information towards epilepsy classification.

Reference

[1] Siddiqui, M. K., Morales-Menendez, R., Huang, X., & Hussain, N. (2020). A review of epileptic seizure detection using machine learning classifiers. Brain informatics, 7, 1-18.

[2] Mursalin, M., Zhang, Y., Chen, Y., & Chawla, N. V. (2017). Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier. Neurocomputing, 241, 204-214.

[3] Faust, O., Acharya, U. R., Adeli, H., & Adeli, A. (2015). Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis. Seizure, 26, 56-64.

[4] Misulis, K. E. (2013). Atlas of EEG, seizure semiology, and management. Oxford University Press.

[5] Pachori, R. B., & Patidar, S. (2014). Epileptic seizure classification in EEG signals using second-order difference plot of intrinsic mode functions. Computer methods and programs in biomedicine, 113(2), 494-502.

[6] Al-Fahoum, A. S., & Al-Fraihat, A. A. (2014). Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains. International Scholarly Research Notices, 2014.

[7] Q. Yuan, W. Zhou, S. Li, and D. Cai, “Epileptic EEG classification based on extreme learning machine and nonlinear features,” Epilepsy Res., vol. 96, nos. 1–2, pp. 29–38, Sep. 2011.

[8] Longadge, R., & Dongre, S. (2013). Class imbalance problem in data mining review. arXiv preprint arXiv:1305.1707.

[9] S. Nasehi, H. Pourghassem, and A. Etesami, “Online epilepsy diagnosis based on analysis of EEG signals by hybrid adaptive filtering and higherorder crossings,” in Proc. Int. Conf. Intell. Comput. Bio-Med. Instrum., Dec. 2011, pp. 192–195.

[10] De Clercq W, Lemmerling P, Van Huffel S, Van Paesschen W. Anticipation of epileptic seizures from standard EEG recordings. Lancet. 2003 Mar 15;361(9361):971.

[11] Kang Lövgren, S., & Rosquist, C. (2019). Machine Learning Methods for EEG-based Epileptic Seizure Detection.

[12] Masum, M., Shahriar, H., & Haddad, H. (2020, July). Analysis of Sampling Techniques Towards Epileptic Seizure Detection from Imbalanced Dataset. In 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC) (pp. 684-692). IEEE.

[13] Masum, M., Shahriar, H., & Haddad, H. M. (2020, July). Epileptic Seizure Detection for Imbalanced Datasets Using an Integrated Machine Learning Approach. In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 5416-5419). IEEE.

[14] Masum, M., Shahriar, H., & Haddad, H. M. (2021, September). A Statistical Summary Analysis of Window-Based Extracted Features for EEG Signal Classificaiton. In 2021 International Conference on Digital Health (ICDH). IEEE. (Under review).

[15] Kumar, A., & Kolekar, M. H. (2014, November). Machine learning approach for epileptic seizure detection using wavelet analysis of EEG signals. In 2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom) (pp. 412-416). IEEE.

[16] Subasi, A., & Gursoy, M. I. (2010). EEG signal classification using PCA, ICA, LDA and support vector machines. Expert systems with applications, 37(12), 8659-8666.

[17] Mursalin, M., Zhang, Y., Chen, Y., & Chawla, N. V. (2017). Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier. Neurocomputing, 241, 204-214.

[18] Birjandtalab, J., Heydarzadeh, M., & Nourani, M. (2017, August). Automated EEG-based epileptic seizure detection using deep neural networks. In 2017 IEEE International Conference on Healthcare Informatics (ICHI) (pp. 552-555). IEEE.

[19] Coşgun, E., Çelebi, A., & Güllü, M. K. (2019, October). Epileptic Seizure Prediction for Imbalanced Datasets. In 2019 Medical Technologies Congress (TIPTEKNO) (pp. 1-4). IEEE.

[20] Yuan, Q., Zhou, W., Zhang, L., Zhang, F., Xu, F., Leng, Y., ... & Chen, M. (2017). Epileptic seizure detection based on imbalanced classification and wavelet packet transform. Seizure, 50, 99-108.

[21] https://archive.ics.uci.edu/ml/index.php

[22] https://keras.io/

[23] Jolliffe I.T. Principal Component Analysis, Series: Springer Series in Statistics, 2nd ed., Springer, NY, 2002, XXIX, 487 p. 28 illus. ISBN 978-0-387-95442-4

[24] Afrin, R., Haddad, H., & Shahriar, H. (2019, July). Supervised and Unsupervised-Based Analytics of Intensive Care Unit Data. In 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC) (Vol. 2, pp. 417-422). IEEE

[25] Ho, T. K. (1995, August). Random decision forests. In Proceedings of 3rd international conference on document analysis and recognition (Vol. 1, pp. 278-282). IEEE.

[26] J. R. Villar, M. Menéndez, E. de La Cal, J. Sedano, and V. M. González, “Identification of abnormal movements with 3D accelerometer sensors for seizure recognition,” J. Appl. Logic, vol. 24, pp. 54–61, Nov. 2017.

[27] More, A. (2016). Survey of resampling techniques for improving classification performance in unbalanced datasets. arXiv preprint arXiv:1608.06048.

[28] Last, F., Douzas, G., & Bacao, F. (2017). Oversampling for imbalanced learning based on k-means and smote. arXiv preprint arXiv:1711.00837.

[29] Chawla, N. V., Japkowicz, N., and Drive, P. (2004). Editorial: Special issue on learning from imbalanced data sets. ACM SIGKDD Explorations Newsletter, 6(1):1–6.

[30] Bunkhumpornpat, C., Sinapiromsaran, K., and Lursinsap, C. (2009). Safe-level-smote: Safe-levelsynthetic minority over-sampling technique for handling the class imbalanced problem. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), volume 5476 LNAI, pages 475–482.

[31] Batista, G. E., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD explorations newsletter, 6(1), 20-29.

[32] Zhang, J., & Chen, L. (2019). Clustering-based undersampling with random over sampling examples and support vector machine for imbalanced classification of breast cancer diagnosis. Computer Assisted Surgery, 24(sup2), 62-72.

[33] Mohd, F., Jalil, M. A., Noora, N. M. M., Ismail, S., Yahya, W. F. F., & Mohamad, M. (2019, December). Improving Accuracy of Imbalanced Clinical Data Classification Using Synthetic Minority Over-Sampling Technique. In International Conference on Computing (pp. 99-110). Springer, Cham.

[34] Zhang, J., Chen, L., & Abid, F. (2019). Prediction of Breast Cancer from Imbalance Respect Using Cluster-Based Undersampling Method. Journal of healthcare engineering, 2019.

[35] Wu, J. M. T., Tsai, M. H., Hsu, C. T., Huang, H. C., & Chen, H. C. (2019, March). Intelligent Signal Classifier for Brain Epileptic EEG Based on Decision Tree, Multilayer Perceptron and Over-Sampling Approach. In Future of Information and Communication Conference (pp. 11-24). Springer, Cham.

[36] Levin Kuhlmann, Klaus Lehnertz, Mark P. Richardson, Björn Schelter & Hitten P. Zaveri, Seizure prediction — ready for a new era, Nature Reviews Neurology volume 14, pages618–630(2018).

[37] Y. Zhang, Y. Zhang, J. Wang, and X. Zheng, “Comparison of classification methods on EEG signals based on wavelet packet decomposition,” Neural Comput. Appl., vol. 26, no. 5, pp. 1217–1225, Jul. 2015.

[38] Duckrow RB, Spencer SS. Regional coherence and the transfer of ictal activity during seizure onset in the medial temporal lobe. Electroencephalogr Clin Neurophysiol. 1992 Jun;82(6):415–22.

[39] K. Fu, J. Qu, Y. Chai, and Y. Dong, "Classification of seizure based on the time-frequency image of EEG signals using HHT and SVM," Biomed. Signal Process. Control, vol. 13, pp. 15–22, Sep. 2014.

[40] Hassan, K. M., Islam, M. R., Tanaka, T., & Molla, M. K. I. (2019, October). Epileptic Seizure Detection from EEG Signals Using Multiband Features with Feedforward Neural Network. In 2019 International Conference on Cyberworlds (CW) (pp. 231-238). IEEE.

[41] Vaitheeshwari, R., & SathieshKumar, V. (2019, February). Performance Analysis of Epileptic Seizure Detection System Using Neural Network Approach. In 2019 International Conference on Computational Intelligence in Data Science (ICCIDS) (pp. 1-5). IEEE

[42] Faust, O., Acharya, U. R., Adeli, H., & Adeli, A. (2015). Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis. Seizure, 26, 56-64.

[43] Gao, Y., Gao, B., Chen, Q., Liu, J., & Zhang, Y. (2020). Deep convolutional neural network-based epileptic electroencephalogram (EEG) signal classification. Frontiers in neurology, 11.

[44] Shoeb, A., Pang, T., Guttag, J., & Schachter, S. (2009). Non-invasive computerized system for automatically initiating vagus nerve stimulation following patient-specific detection of seizures or epileptiform discharges. International Journal of Neural Systems, 19(03), 157-172.

[45] Banerjee, S., & Chatterjee, R. (2021). Temporal Window based Feature Extraction Technique for Motor-Imagery EEG Signal Classification. bioRxiv.

[46] Subasi, A., & Gursoy, M. I. (2010). EEG signal classification using PCA, ICA, LDA and support vector machines. Expert systems with applications, 37(12), 8659-8666.

[47] Wang, X., Gong, G., Li, N., & Qiu, S. (2019). Detection analysis of epileptic EEG using a novel random forest model combined with grid search optimization. Frontiers in human neuroscience, 13, 52.

[48] Mursalin, M., Islam, S. S., Noman, M. K., & Al-Jumaily, A. A. (2019). Epileptic seizure classification using statistical sampling and a novel feature selection algorithm. arXiv preprint arXiv:1902.09962.

[49] Radman, M., Moradi, M., Chaibakhsh, A., Kordestani, M., & Saif, M. (2020). Multi-Feature Fusion Approach for Epileptic Seizure Detection from EEG Signals. IEEE Sensors Journal.

[50] Riaz, F., Hassan, A., Rehman, S., Niazi, I. K., & Dremstrup, K. (2015). EMD-based temporal and spectral features for the classification of EEG signals using supervised learning. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 24(1), 28-35.

[51] Andrzejak, R. G., Lehnertz, K., Mormann, F., Rieke, C., 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(6), 061907.

[52] Barandas, M., Folgado, D., Fernandes, L., Santos, S., Abreu, M., Bota, P., ... & Gamboa, H. (2020). Tsfel: Time series feature extraction library. SoftwareX, 11, 100456.

[53] Fan, N., Balan, R. V., & Rosca, J. (2004, November). Comparison of wavelet-and FFT-based single-channel speech signal noise reduction techniques. In Wavelet Applications in Industrial Processing II (Vol. 5607, pp. 127-138). International Society for Optics and Photonics.

[54] Swami, P., Gandhi, T. K., Panigrahi, B. K., Tripathi, M., & Anand, S. (2016). A novel robust diagnostic model to detect seizures in electroencephalography. Expert Systems with Applications, 56, 116-130.

[55] Zhang, T., & Chen, W. (2016). LMD based features for the automatic seizure detection of EEG signals using SVM. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 25(8), 1100-1108.

[56] Raghu, S.; Sriraam, N.; Hegde, A.S.; Kubben, P.L. A novel approach for classification of epileptic seizuresusing matrix determinant.Expert Syst. Appl.2019,127, 323–341

[57] Singh, N.; Dehuri, S. Usage of Deep Learning in Epileptic Seizure Detection Through EEG Signal. InProceedings of the 3rd Nanoelectronics, Circuits and Communication Systems (NCCS), Ranchi, India,11–12 November 2017; pp. 219–228

[58] Gupta, V.; Pachori, R.B. Epileptic seizure identification using entropy of FBSE based EEG rhythms.Biomed.Signal Process. Control2019,53, 10156

[59] Mamli, S.; Kalbkhani, H. Gray-level co-occurrence matrix of Fourier synchro-squeezed transform for epilepticseizure detection.Biocybern. Biomed. Eng.2019,39, 87–99

[60] Usman, S. M., Khalid, S., & Aslam, M. H. (2020). Epileptic seizures prediction using deep learning techniques. Ieee Access, 8, 39998-40007.

[61] Zhang, Z., & Parhi, K. K. (2015, August). Seizure detection using regression tree based feature selection and polynomial SVM classification. In 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 6578-6581). IEEE.

[62] Fisher NK, Talathi SS, Cadotte A, Carney PR. Epilepsy detecting and monitoring. In: Tong S, Thakor NV, editors. Quantitative EEG analysis methods and clinical applications. Artech House; 2009. pp. 141–65.

[63] Iasemidis LD, Pardalos P, Sackellares JC, Shiau DS. Quadratic binary programming and dynamical system approach to determine the predictability of epileptic seizures. J Comb Optim. 2001;5:9–26.

[64] Iasemidis LD, Shiau DS, Chaivalitwongse W, Sackellares JC, Principe JC, Carney PR, Prasad A, Veeramani B, Tsakalis K. Adaptive epileptic seizure prediction system. IEEE Trans Biomed Eng. 2003;50:616–27.

[65] Alotaiby, T. N., Alshebeili, S. A., Alshawi, T., Ahmad, I., & Abd El-Samie, F. E. (2014). EEG seizure detection and prediction algorithms: a survey. *EURASIP Journal on Advances in Signal Processing*, *2014*(1), 1-21.

[66] Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., & Herrera, F. (2011). A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42(4), 463-484.