ALGORITHMS FOR EEG-BASED MONITORING OF EPILEPTIC SEIZURES

by

Javad Birjandtalab Golkhatmi

APPROVED BY SUPERVISORY COMMITTEE:

Dr. Mehrdad Nourani, Chair

Dr. Farokh B. Bastani

Dr. Lakshman S. Tamil

Dr. Won Namgoong

To my lovely wife Somayeh, and my cute son Liam

For your love and endless support. They have been by my side throughout this PhD, living every single minute of it. This journey would not have been possible if not for them, and I dedicate this milestone to them.

ALGORITHMS FOR EEG-BASED MONITORING OF EPILEPTIC SEIZURES

by

JAVAD BIRJANDTALAB GOLKHATMI, MS

DISSERTATION

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN

ELECTRICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT DALLAS

May 2020

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my supervisor, Professor Mehrdad Nourani, for his excellent guidance, caring, patience and for providing me an outstanding atmosphere throughout my research. I am very grateful for the opportunity to work in his research lab. I would like to thank my committee members, Dr. Lakshman Tamil, Dr. Farokh Bastani, and Dr. Won Namgoong, for serving on my PhD committee. I also want to thank you for letting my defense be an enjoyable moment, and for your brilliant comments and suggestions. I am thankful to my collaborators at UT Southwestern Medical Center, specially Dr. Jay Harvey for his helpful comments. I also thank to my peers at The University of Texas at Dallas for their encouragement and support.

I would also like to thank my parents for their love, kindness and support. They were always supporting me and encouraging me with their best wishes.

November 2019

ALGORITHMS FOR EEG-BASED MONITORING OF EPILEPTIC SEIZURES

Javad Birjandtalab Golkhatmi, PhD
The University of Texas at Dallas, 2020

Supervising Professor: Dr. Mehrdad Nourani, Chair

Millions of people around the world suffer from epilepsy. Approximately 0.1 percent of epileptic patients die from unexpected deaths. It is of a great value if technology can provide a method to efficiently monitor the seizures and alert the caregivers to help patients. An Electroencephalography (EEG) signal is able to discover any neuron's misfiring or excessive neural activity which can be a sign of a neurological disorder. It is proven that EEG signals are the best markers for detection and diagnosis of the epileptic seizures. Frequency domain features (like normalized in-band power spectral density) are known as most informative attributes to extract meaningful information from EEG signals.

In this work, we addressed three main challenges in the area of epileptic seizure monitoring. First, we proposed a channel selection method which selects the most informative EEG channels out of full EEG channel set. We embedded high dimensional spectral features into the low dimension space to improve the accuracy seizure detection. Second, we suggested two novel imbalance learning techniques to address the problem of class imbalance in the seizure dataset. Using this approach the classification models can better get trained and learn more from seizure samples. Third, we proposed a personalized seizure prediction methodology to extract footprint of seizure and identify pre-seizure attributes based on each patient's response that time. Using this approach, the accuracy of seizure prediction is improved since only the most informative portion of pre-seizure data is used for prediction.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Epileptic Seizure Monitoring

Epilepsy is a neurological disorder which can, if not controlled, potentially cause unexpected injuries and even death. It is extremely crucial to have accurate automatic pattern recognition and data mining techniques to detect the onset of seizures and inform care-givers to help the patients. EEG signals are the preferred biosignals for diagnosis of epileptic patients [13]. If possible, everyday monitoring and risk assessment of the approximately 1.3 million Americans who suffer from intractable epileptic seizures would significantly improve their quality of life. While Epilepsy Monitoring Units (EMUs) have been effective for pre-surgery evaluation, and for identifying origin of seizure and proper anticonvulsants, priceless patient-specific knowledge, buried within the large volume of EMU data, remain unused. This is attributed to the challenges associated with traditional methods of analysis. In spite of recent advances in noninvasive wearable devices, the critical barrier of developing reliable data analysis still limits their clinical viability for providing early seizure alerts. New breakthroughs in machine intelligence theory present an emerging opportunity to address this roadblock to apply advanced data analytic, which we believe will allow us to successfully monitor epileptic seizures.

1.2 State of knowledge

Figure 1.1 shows different epileptic seizure monitoring approaches. They can be grouped into invasive and non-invasive categories. The invasive category is grouped into depth EEG and subdural EEG. The depth EEG is implemented inside the brain while the subdural EEG sits over the surface of the brain. The non-invasive category is grouped into Neuro-imaging, Multi-channel EEG monitoring and wearable monitoring. Multi-channel EEG monitoring is

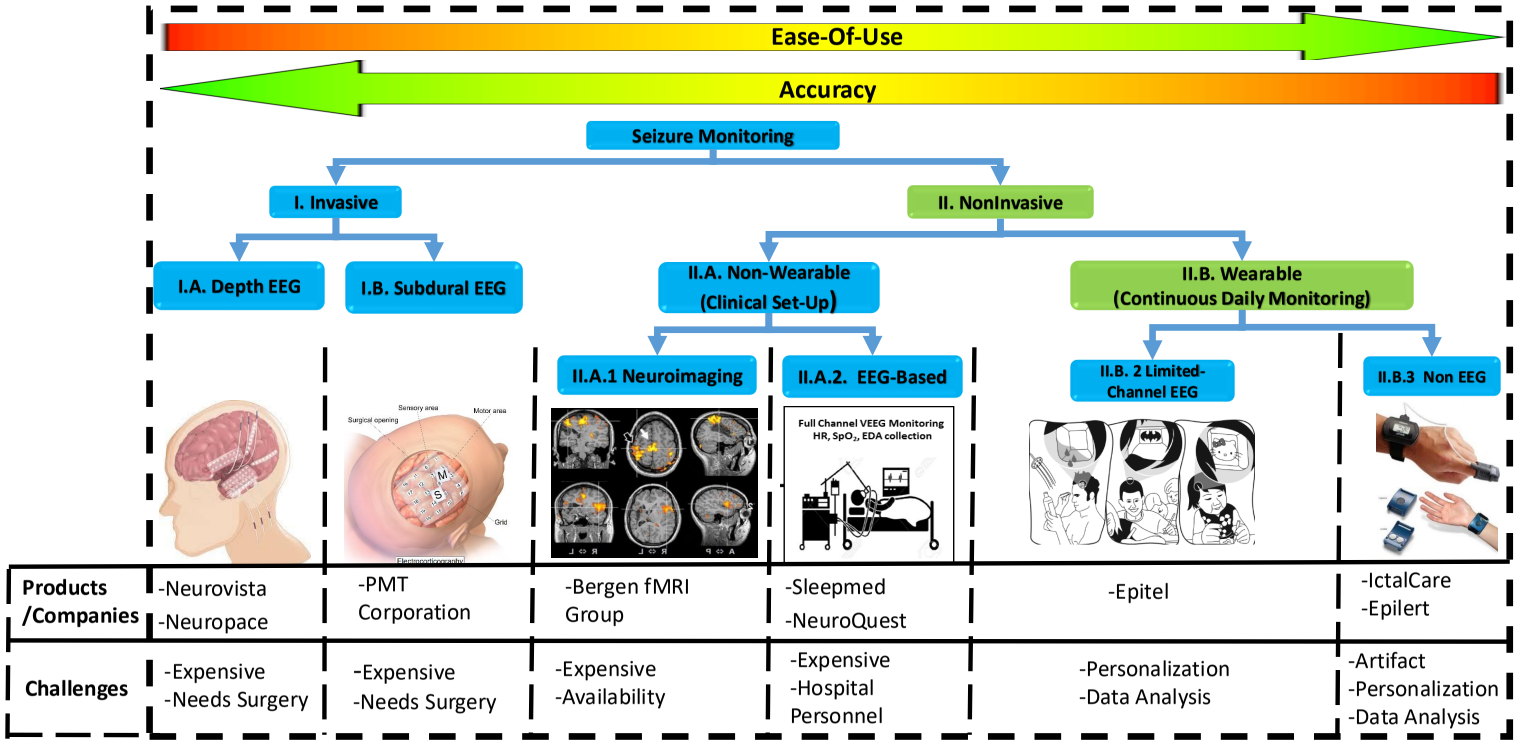


Figure 1.1. Epileptic seizure monitoring methods.

the gold standard for seizure characterization and localizing changes within the brain which is indicative of seizure focus. Such monitoring is, however, impractical for use in daily life because it requires the use of either a special hat/headset worn in just the right way or sensors implanted in the brain. The wearable monitoring itself is divided into non-EEG analysis and limited channel analysis. Use of non-EEG signals to analyze seizures presents its own set of challenges. The impact of seizures on biosignals differs among patients and seizure types; consequently, no single modal system will be helpful for every patient. As shown in Figure 1.1, we advocate an EEG based analysis approach that utilizes the limited number of EEG channels.

1.3 Motivation

Approximately 1 percent of world population suffer from the epilepsy. These neurological disorders may cause permanent damages if they are not managed properly. It is crucial to have techniques to accurately and automatically detect epileptic seizures. Many patients gain control of their seizures with anti-convulsant or anti-epileptic drugs (AEDs); However, approximately one-third of the patients prove intractable with medications [18]. Serious physical injury and death (known as Sudden Unexpected Death of Epilepsy or SUDEP) occur in patients with uncontrolled epilepsy. Risk of serious physical injury or even death is high for such epileptic patients. There is a need, therefore, for accurate, automated detection of seizures to alert patients/caregivers and to assist epileptologist provide effective therapeutic solutions and to improve the quality of life of epileptic patients [46][13].

Electroencephalography (EEG) signals carry informative features which are able to explain most of normal and abnormal brain activities, particularly epileptic seizures. Physicians use EEG signals as the ground truth for epileptic seizure detection for years in order to find the focal point of seizure and treat the injured brain tissues by medication and/or surgery. Electroencephalography (EEG) signal analysis, as a diagnostic test, has had an enormous

impact on epileptic seizure detection [52]. EEG signal analysis can improve our understanding of abnormal brain activities like neurons misfiring or excessive neural activity. However, EEG-based seizure detection is relatively complex since it involves voluminous amounts of recorded EEG samples coming from multiple scalp mounted electrodes [2].

Class imbalance problem is a well-known machine learning problem where the total number of data points in one class is far less than the total number of points in another class. In health-care, data is often predominately composed of normal samples with only a small percentage of abnormal ones, leading to the so-called class imbalance problems. In class imbalance problems, inputting all the data into the classifier to build up the learning model will usually lead to a learning bias toward the majority class and thus low accuracy.

The seizure dataset is often heavily under-represented and extremely unbalanced because such patients experience seizures only for a very short time. The overwhelming majority of the instances belong to the non-seizure class. Actual data collected in Epileptic Monitoring Unit (EMU) indicate more than 99.9% non-seizure and only 0.1% seizure. When the number of instances in one class far exceeds the other (e.g. $\times 10$ or more), many computational problems can arise. So, we apply methods to balance the imbalanced dataset. Data balancing prevent classification model to be overwhelmed by the majority (large) class and ignore minority (small) class and consequently improve classifier robustness [13].

Andres M. Kanner in his books talks about patients with epilepsy experiencing psychiatric and cognitive symptoms preceding and following the seizure, apart from the seizure event itself. The period before the seizure event is called the *preictal* (pre-seizure) period and after it is called the *postictal* (post-seizure) period. He observes that among all psychiatric phenomena, preictal, ictal and postictal psychiatric symptoms are the least investigated in our systematic research studies [26][33]. What we attempt to achieve in this work is to recognize patterns in brain activity of each patient personalized to him or her. Personalization for recognizing each patient's preictal patterns is made possible by the adoption of neural

networks after having trained on the data available for each of 23 patients in MIT dataset. Another reason for adoption of neural networks is ability to utilize more training data as collected. Ultimately, there is a need for wearable devices capable of detecting both convulsive and nonconvulsive seizures in everyday life. A wearable device, worn by a patient apart from sending out alarm signals to a caregiver, could also collect data which can be used to train the neural network and improve its performance over time. In addition, long term use of wearable devices by large numbers of epileptic patients would provide biosignal data for researchers seeking to better understand long term effects of seizures, and in particular the causes and risk factors of SUDEP [20].

1.4 Contribution

In this work, three techniques are proposed to be used in a platform for EEG-based epileptic seizure monitoring:

I. The first contribution is a new method to identify a limited number of EEG channels from full set of EEG channels. This method enables us to identify EEG channels with most informative profiles while avoiding those channels that may unnecessarily inject more noise and ambiguities in our analysis. We propose to train a random forest model using the spectral domain features and count the number of times that each channel is repeated in the forest. We also employed a dimension reduction approach to visualize the EEG data in low dimension. In comparison with others, the proposed method is able to provide higher detection accuracy.

II. Most of epileptic seizure data sets suffer from insufficient number of seizure events. In order to address this challenge, we advocate two techniques to improve the accuracy of the seizure detection. First, we take into account the non-seizure class in the first imbalance learning approach. This class is the majority class and practically includes

several normal activities itself. So, we proposed clustering the majority class and consider the minority class separately with each majority class cluster. This approach can decrease the imbalance ratio and ultimately improves the accuracy of seizure detection. Second, we used a customized neural network to detect epileptic seizures. We used the concept of weighted cost function taking into account the imbalance ratio of each EEG dataset. So, the importance of correctly classifying seizure events is higher that correctly classifying non-seizure events from neural network perspective. Using this approach the configuration of neural network classifier is different from patient to patient.

III. To address the issue of personalized seizure prediction, in the third part, we developed a systematic technique to customize the pre-seizure data selection. Starting with a fixed amount of pre-seizure data, we ended up with a specific amount of pre-seizure data for each patient which are different from each other in terms of time-to-seizure and window size. We will show how personalizing the pre-seizure data factors can increase the quality of seizure prediction. Our proposed technique involves the concept of sliding windows of different window sizes to squeeze in the shortest monitoring time to accurately call an impending seizure. The proposed personalized method improves the accuracy of seizure prediction.

1.5 Related Works

There are two different approaches to handling the problem of high-dimensionality in EEG signals:

1) **Channel Selection**: These techniques select an effective subset of the original channels. A filter-based channel selection technique based on maximum variance criteria has been investigated in [48]. Authors used the difference in variance between seizure and non-seizure

events for selecting the best number of channels in [25]. The information gain of EEG channels is used in [20] to find the most informative channels for seizure detection. Authors calculated the accuracy of all individual EEG channels and selected the channels with the highest accuracy for limited channel seizure detection in [61]. However, these techniques may not always be efficient, particularly when there are a huge number of features extracted from a noisy environment like scalp EEG.

2) **Dimension Reduction**: Researchers have investigated how to map an original feature space into a smaller feature space representation in order to reduce the overwhelming number of features. Authors presented a semi-automated patient-dependent unsupervised technique using all EEG channels and Principal Component Analysis (PCA) as their dimension reduction technique. Authors in [34] applied autocorrelation to extract features and used Common Spatial Patterns (CSP) to decrease the feature space dimensionality. Authors in [61] extracted 26 features per channel and reduced the number to 8 using PCA. They reported the accuracy of EEG channels individually using an LDA classifier. Although PCA is able to reduce the number of features, it cannot properly preserve the distance relationships between data points in low dimensions. The number of patients investigated in [61] is too small to be conclusive (only six epileptic patients out of 23 available patients in the MIT database [30]). Moreover, the number of non-seizure events selected is too small for adequate training.

Most popular epileptic seizure detection systems using Machine Learning have been based on features extracted in time-frequency domain and classical classifiers like Support Vector Machines and Artificial Neural Networks (ANN). Authors in [55] have used SVM to build classifiers for seizure detection. Another work proposed using feature extraction on sample entropy combined with Extreme Learning Machine (ELM) where high accuracy along with fast learning speed are reported [56].

Coming to ANNs, researchers proposed a method for automatic seizure detection from EEG signals using a multistage nonlinear pre-processing filter in combination with a diagnostic

Large Memory Storage and Retrieval (LAMSTAR) Neural Network [44]. Authors in [29] used cosine Radial Basis Function Neural Network (RBFNN) for seizure detection. They have performed time-frequency analysis on selected segments of EEG signals to extract features and use them as input to ANN and achieved excellent results [59]. In all real patient's data, one of the problems is the high imbalance ratio between the majority and minority classes. There are several ways proposed to address the class imbalance problem. Researchers empirically employed oversampling, under-sampling, threshold-moving, hard-ensembling and soft-ensemble techniques in training cost-sensitive neural networks [63].

Several works have demonstrated different techniques for mitigating the class imbalance problem. The authors in [28] have provided an empirical comparison of a wide variety of machine learning algorithms for solving class imbalance problem. Authors in [62] present a new EEG classification approach based on the extreme learning machine (ELM) and nonlinear dynamical features and demonstrate its superior performance. In another work, authors investigated how different imbalance ratios affect clustering imbalanced data. Algorithm represented in [39], under-sample class imbalance data used clustering technique. Authors proposed an extension of Synthetic Minority Over-sampling Technique (SMOTE) for epilepsy monitoring [24].

In order to predict the epileptic seizures, Acharya, et. al, employ a 13 layer Deep Convolutional Neural Networks (CNN) with a 10-fold cross validation strategy all of which requires significant resources to train and run. The authors claim that their research is the first implementation of CNN for EEG signal processing in general and seizure detection in particular. This work involves detection of normal, preictal and seizure EEG signals without performing feature extraction and selection [1]. Another seizure prediction based machine learning publication proposed the use of Cost-sensitive Support Vector Machines (CSVM) which treats signals from preictal more significantly than those from *interictal* (the period between seizures). Unlike the CNN proposal, this involves prepossessing the patient data,

extracting features before training the classifier. It uses the k-of-n analysis among consecutive windows as part of post-processing in the 5 minute horizon to classify if the signals are preictal or interictal [42]. This work also advocates loading an externally trained model on the wearable or implantable device. However, it does not consider personalization for seizure prediction parameters.

1.6 Dissertation Organization

In Chapter 2, the proposed method for limited channel epileptic seizures detection using EEG signals is presented. This chapter describes the features extracted from EEG signals and the proposed dimension reduction technique and classification method. Chapter 3 presents two imbalance learning techniques based on clustering approach and customized neural network for detecting epileptic seizures. These proposed techniques improve the class imbalance problem in seizure data. A personalized technique is developed and explained in Chapter 4 to predict epileptic seizures. The proposed method customize the pre-seziure data for each patient. Finally, Chapter 5 summarizes the dissertation and the key points and possible future works. This chapter explains the limitations of proposed methods.

CHAPTER 2

AUTOMATED SEIZURE DETECTION USING LIMITED-CHANNEL EEG AND NON-LINEAR DIMENSION REDUCTION[1]

2.1 Overview

Electroencephalography (EEG) signals are the best indicators of epileptic seizures. Epileptiform EEG patterns such as spikes and sharp waves can assist in the diagnosis and in classifying seizures [27]. However, full-channel EEG signals recorded from 18-23 electrodes on the scalp is neither wearable nor computationally effective. This work presents advantages of both channel selection and nonlinear dimension reduction for accurate automatic seizure detection. We first extract the frequency domain features from the full-channel EEG signals. Then, we use a random forest algorithm to determine which channels contribute the most in discriminating seizure from non-seizure events. Next, we apply a non-linear dimension reduction technique to capture the relationship among data elements and map them in low dimension. Finally, we apply a KNN classifier technique to discriminate between seizure and non-seizure events. The experimental results for 23 patients show that our proposed approach outperforms other techniques in terms of accuracy. It also visualizes long-term data in 2D to enhance physician cognition of occurrence and disease progression.

2.1.1 Motivation

In order to tackle the problems related to the complexity of EEG-based seizure detection, two techniques can be employed: First, select a limiting number of EEG channels instead of using all EEG (often 18-23) electrodes. Full-channel EEG monitoring is only practical in clinics since it is expensive, time consuming, uncomfortable and stigmatizing. Moreover,

[1]Birjandtalab, J., Pouyan, M. B., Cogan, D., Nourani, M., Harvey, J. (2017). Automated seizure detection using limited-channel EEG and non-linear dimension reduction. Computers in biology and medicine, 82, 49-58.

irrelevant channels add noise to the feature space and decrease the seizure detection accuracy. Limited-channel configurations can be implemented with newly developed discrete wearable patches (e.g., Epitel nodes [31]) which are able to record seizure activity in daily life using individual nodes. There are recently developed techniques which make these wearable devices more energy-efficient and practical for daily life applications [53]. Second, map the high dimensional dataset into a lower dimension by preserving the neighborhood information of data points. Since the feature space of EEG signals has a non-linear structure, non-linear data embedding (that preserves distance information among neighboring data points) techniques are the best option for reducing the number of feature space dimensions. In this work, we combine both channel selection and dimension reduction to provide a low dimensional feature space for epileptic seizure detection.

2.1.2 Main Contribution

To the best of our knowledge, our proposed method is the first work which combines channel-selection and dimension reduction for EEG-based epileptic seizure detection. The general view of our proposed epileptic seizure detection model is shown in Figure 2.1. First, we use power spectral analysis to extract features per channel per subject for each time window. Then, we use a random decision forest for selecting a limited number of channels. To do this, we first generate a large number of random decision trees using features from all the EEG channels. Next, we investigate the number of times each channel appears in the forest. The channels with the highest contribution to the entire forest are chosen for limited channel seizure detection. The best channels are selected by voting among all the decision trees in the forest, thereby making our method robust against noisy channels - the most important advantage of this technique. Next, we use t-distributed stochastic neighbor embedding (t-SNE) to embed and represent data in a lower dimension by preserving the relationship of data points in high dimensional feature space. Finally, we use a KNN classifier model to differentiate between seizure and non-seizure events in a patient-specific manner.

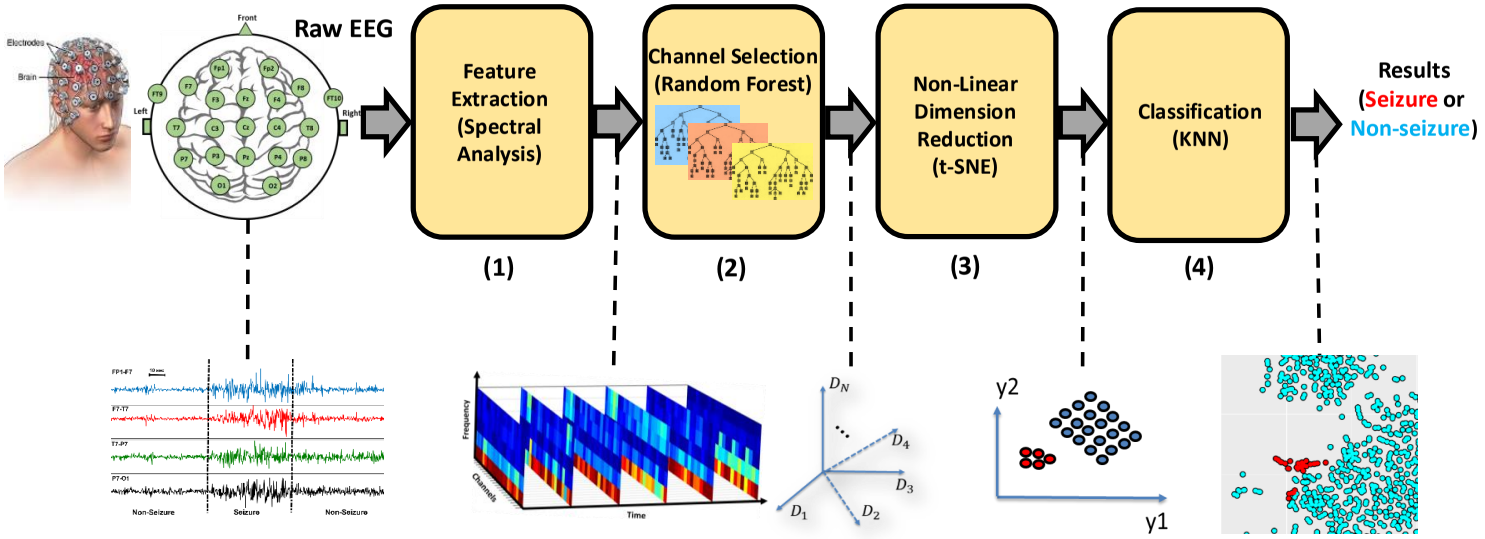


Figure 2.1. Overall architecture of our proposed method.

2.2 Feature Extraction

Feature Extraction is a critical step in EEG-based seizure detection since it extracts seizure related characteristics. Researchers categorize EEG signals into unique bandwidths as shown in Table 2.1. Both normal and abnormal brain functions are listed by bandwidth [58]. Note that seizures affect most of the EEG frequencies. Consequently, frequency features are widely used for EEG-based epileptic seizure detection [15].

In our method, we first segmented the EEG signals into 10 seconds windows. Next, we calculated the power spectral density of all EEG channels for each window using Fourier transforms. The results are divided into the frequency bands shown in Table 2.1, i.e., $w = \{\delta, \theta, \alpha, \beta, \gamma\}$.

We calculate the Power Spectral Density (PSD) of a specific channel c recorded with frequency f_s as follows [13]:

$$P^c(f) = \frac{1}{f_s N}\left|\sum_{n=0}^{N-1} x_n^c e^{-j2\pi f n}\right|^2 \qquad \frac{-f_s}{2} < f < \frac{f_s}{2} \tag{2.1}$$

where x_n^c represents the n^{th} sample of channel c (out of total N samples). We define P_w^c to calculate the normalized PSD of channel c in frequency band $w = [w_1, w_2]$ as:

$$P_w^c = \frac{\sum_{f=w_1}^{f=w_2} P^c(f)}{\sum_{f=0}^{f=\frac{f_s}{2}} P^c(f)} \tag{2.2}$$

where w_1 and w_2 represent the starting and ending points of the frequency band. In order to take full advantage of the power spectrum, the amount of power in each band is divided by the overall spectral power of the window (i.e. $f_s = 256$ Hz). The full channel feature space includes five normalized in-band power spectral density features for each channel. For example, a patient with 23 EEG channels provides a high-dimensional space of 115 features [8][14].

Table 2.1. EEG bandwidths and frequencies [58].

Bandwidth	Frequency [Hz]	Normal Functions	Abnormal Functions
Delta (δ)	0.1-4	hyperventilation, encephalopathy, seizures	
Theta (θ)	4-8	drowsiness, encephalopathy, seizures	
Alpha (α)	8-12	closing the eyes, inhabitation	coma, seizures
Beta (β)	12-30	effect of medication, drowsiness	drug overdose, seizures
Gamma (γ)	30-70	voluntary motor movement, learning	seizures

2.3 Limited Channel Selection

2.3.1 Motivation & Challenges

Limited-channel EEG epileptic seizure detection offers three advantages. First, it reduces the computational complexity of seizure detection, leading to a faster run time and lower power consumption, thereby making seizure detection models faster and more cost-effective (e.g. for inexpensive embedded systems). Second, for some patients, it increases the detection accuracy by avoiding redundancy of non-focal/unnecessary channels. Finally, reduction from 23 to 1-3 channels makes wearable EEG monitoring (e.g. for daily use) practical. In other words, channel selection is a prerequisite for implementing wearable seizure detection systems. In this section, we focus on a methodology for selecting the best few channels which reduces the high dimensional feature space and yet provides accurate seizure information.

Most machine learning algorithms provide low accuracy when the number of irrelevant features is notably high [35]. Therefore, it is quite important to select a subset of features for the best classification results. Selecting the best feature subset that optimizes evaluation criteria is an NP-hard problem [41]. The simplest way – an exhaustive search of all combinations of M features ($\sum_{i=1}^{M} \binom{M}{i} = 2^M$) – is not computationally feasible for high dimensional feature sets. An alternate method is estimating the best feature subsets in a practical run time. The way that feature selection integrates search methods with the structure of classification techniques divides feature selection techniques into two main classes: scheme-dependent techniques (a.k.a wrapper techniques) and scheme-independent techniques (a.k.a filter techniques).

1) **Scheme-Dependent Techniques**: In these techniques, every feature subset candidate is evaluated using a classifier and the best feature subset is selected based on classification accuracy. Such techniques have two main drawbacks. First, the selected features depend on the classifier, which increases the chance of overfitting. Second, wrapper techniques are not efficient for large number of features, when the classification model is complex [3].

2) **Scheme-Independent Techniques**: These techniques filter out the irrelevant features regardless of the classification model. They independently evaluate the effectiveness of candidate features using various criteria. In most of the cases, features are sorted based of their relevance using a search algorithm and irrelevant features are filtered out. Filter techniques are computationally fast and simply scalable since they are classification independent [3].

2.3.2 Random Forest Based Technique

In this work, we propose the use of the random forest algorithm [17] to select the best EEG channels. Random forest algorithm is proved to be robust in applications where there are large numbers of irrelevant features [36]. The goal is to find the k most informative channels out of all P channels. Algorithm 1 shows all the steps of our proposed technique in detail. First, we generate a random forest with N_T trees using all M spectral features extracted from all original P channels ($M = 5 \times P$). $mtry$ defines the number of features randomly nominated at each node split during tree generation. When growing an individual decision tree, before each split, first we selected $mtry$ of the M original features ($mtry \leq M$) at random as candidates for splitting. Next, we split that node at the feature which minimizes the sum of impurities ([17]) associated with children of node-split. We repeated these steps for all the nodes in each decision tree. Note that by random selection, all features have an equal chance to be candidates in each node-split of each decision tree. However, only one feature will be selected for a node-split on that node, i.e. the feature which minimizes the sum of the child node impurities. If a feature is irrelevant, even if it gets selected randomly for node splitting, it will not be selected in the final round because it cannot minimize the sum of child nodes impurities compared to relevant features. We set $mtry$ to the square root of the number of features ($mtry = \sqrt{M}$) as suggested by researchers [17]. Next, we calculate δ_i, the contribution of the i^{th} spectral feature to the random forest by counting how many times each spectral feature appears in the entire forest. Then, we sum the contribution of

all five spectral features in the p^{th} channel to find the total contribution of that channel, μ_p. Next, we sort all the channels based on the total contribution of their spectral channels and call it our list of sorted channels $\bar{\mu}$. Finally, we select the best k channels (largest RF contribution) and call it our list of k selected channels, L.

As an example, Figure 2.2 shows the different EEG channel contributions in the random forest (RF contribution) for Patient 1. For this patient, channels 17, 6, and 8 provide the highest contribution and are selected for limited channel seizure detection with $k = 3$. Incidentally, neurophysiologists suggest three channels for limited channels seizure detection since this is often the number covering the epileptic focus [25]. The rational behind our proposed channel selection method is that random forest randomly generates many trees and we expect that relevant and informative features appear in those trees more frequently than redundant features do.

In order to show the effectiveness of the model, we compared our proposed feature selection model with the exhaustive feature selections for one example (patient #1). For this purpose, we measured the performance of seizure detection using all the $\binom{23}{3} = 1771$ possible solutions of selecting 3 channels out of 23 channels. Figure 2.3 shows the performance of all the solutions exhaustively in terms of F1 measure. Since the overwhelming majority of the instances belong to the non-seizure class. Therefore, performance metrics (i.e. sensitivity, precision, and F1), which are not affected by the number of correctly classified non-seizure instances class are selected to be more reliable parameters. The experiment is done for Patient 1. Please note that exhaustive feature selection is an NP-hard problem and the running time increases exponentially with increasing number of features. Therefore, exhaustive search is not practiced for the datasets with high number of features. Figure 2.4 shows the performance of all the exhaustive points in terms of precision and sensitivity. Each black point represents one of the possible iterations of 3 limited channels. The red point shows the performance of the limited channels that is selected by proposed (random forest) method. Also note

that we performed the channel selection in two steps. First, we select the most important (discriminative) features in the entire dataset. Then, we combined the importance of features per channel to see the effectiveness of each channel. Using the two-stage channel selection has the advantage of observing the importance of in-channel features (spectral domain features per channel).

Algorithm 1 Channel Selection

Inputs : (i) Number of trees in forest (N_T),
(ii) Number of limited channels (k),
(iii) Number of node split features ($mtry$),
(iv) Full channels feature set ($F = \{f_1, ..., f_M\}$).
Output : List of selected channels (L)
begin
  $\{Tr_1, ..., Tr_N\}$=*CreateRandomForest* $(F, N_T, mtry)$ **for** $1 \leq n \leq N_T$ **do** *// All trees in forest*
    for $1 \leq i \leq M$ **do** *// All features*
      if $f_i \in Tr_n$ **then** *// Update feature contribution*
        $\delta_i = \delta_i + 1$
  for $1 \leq p \leq P$ **do** *// All channels*
    $\mu_p = \sum_{q=1}^{5} \delta_{5p+q}$
  Sort μ *in ascending order* $\bar{\mu} = \{\bar{\mu}_1, ..., \bar{\mu}_P\}$ $L = \{\bar{\mu}_1, ..., \bar{\mu}_k\}$

2.4 Dimension reduction

Using our technique, the total number of channels is significantly reduced (e.g. $k = 3$ vs. originally 18-23 channels). Consequently, the number of spectral features is also reduced (e.g. 15 vs. originally 90-115 features). However, 15 is still a high dimensional feature space suffering from the curse of dimensionality. The main problem is the *empty space phenomenon*, which means data points become more and more sparse as the feature space dimension increases. This phenomenon does not allow classifiers to properly extract information hidden in the data [37]. In order to tackle this problem, we use data embedding, which is a low dimensional representation of high dimensional data points.

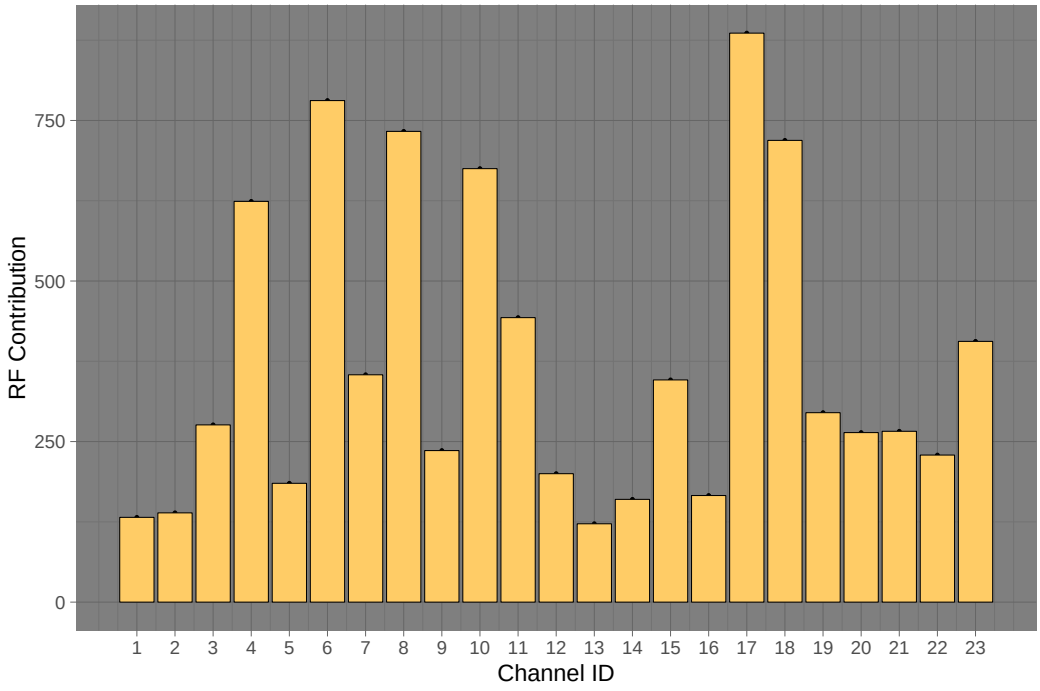


Figure 2.2. Contribution of channels in a Random Decision Forest with 1000 trees. Taken from patient 1's data.

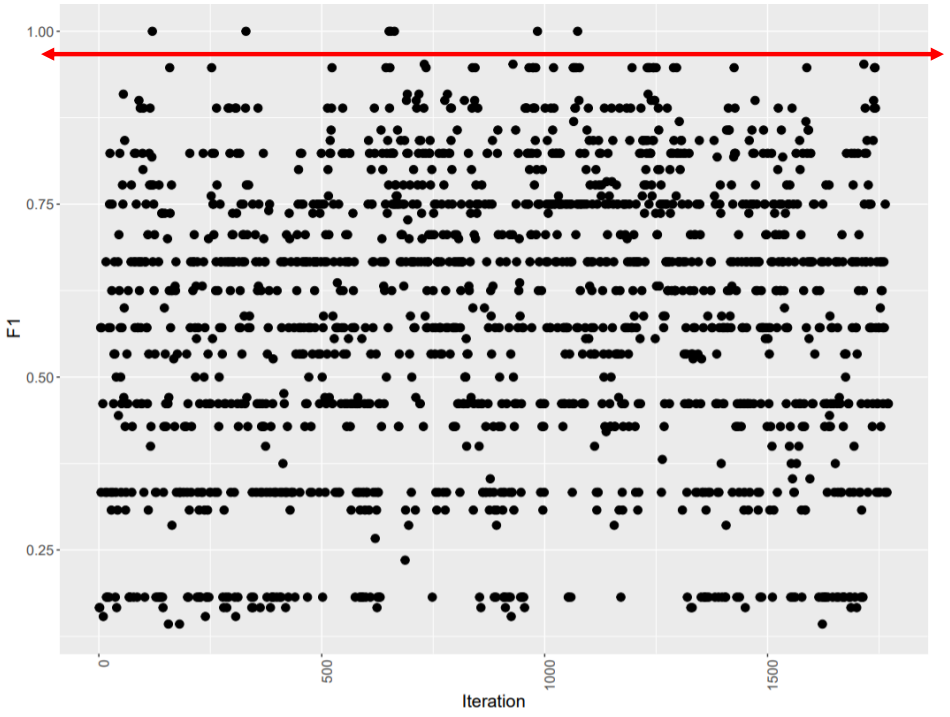


Figure 2.3. F1 measure accuracy of different iterations of exhaustive channel selection. The points corresponding to the proposed channel selection are above the red line.

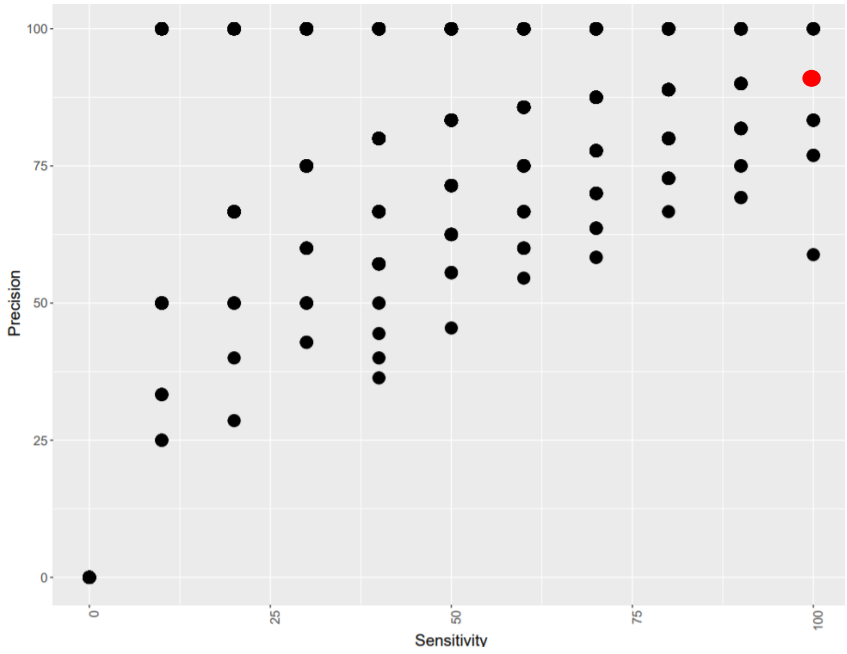
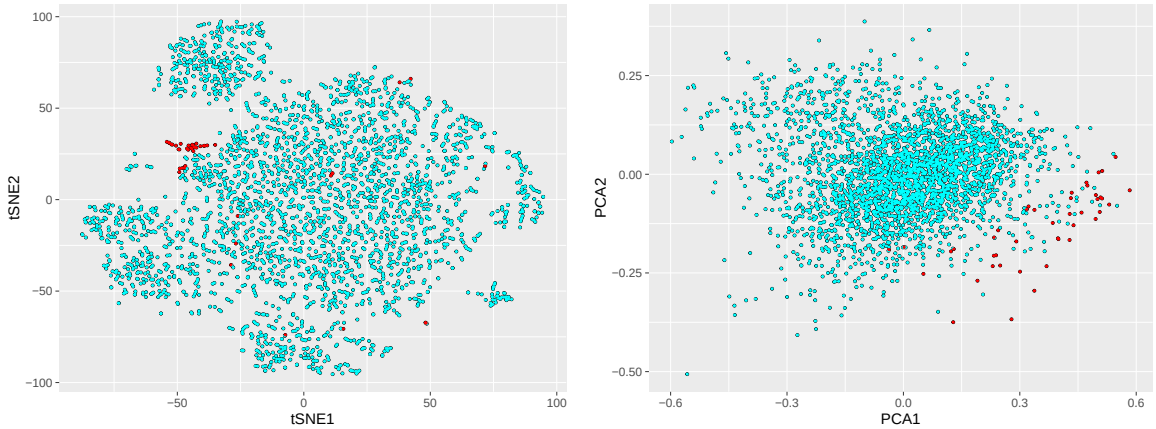


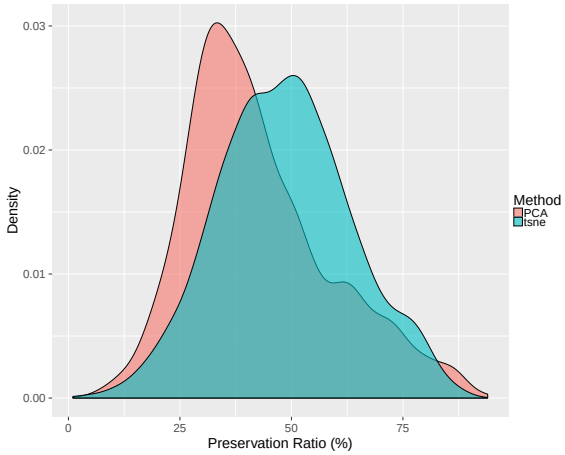
Figure 2.4. Precision and sensitivity of different iterations of exhaustive channel selection (black points) versus proposed channel selection using random forest (red point). Note that there are total of 1171 points here but many of them overlap each other.

Recently, robust dimension reduction techniques are proposed to capture low dimensional structures in a high dimensional data [51][50]. Dimension reduction techniques are categorized as linear or non-linear based on the way they map data points into lower dimensions. Linear techniques like PCA project the data points by preserving the maximum variance or minimum reconstruction error of data points in a linear way [32] . However, these linear approaches cannot preserve the distance relationship of instances from a system that generates datasets with non-linear structures or features. On the contrary, non-linear dimension reduction methods are able to capture pair-wise distances between data points in high dimensional spaces [12]. We were inspired by t-distributed Stochastic Neighbor Embedding (t-SNE), a state-of-the-art non-linear dimension reduction technique, to represent features of all three selected channels in two dimensions [60]. Traditionally, t-SNE has been used in high dimensional single-cell analysis [4] due to its ability to deal with large (e.g. 13-31) dimensions.



(a)

(b)



(c)

Figure 2.5. Patient 1 preservation ratio analysis (a) visualizing the seizure (red) and non-seizure (blue) data points data points in two dimensional feature space using t-SNE. (b) visualizing the seizure (red dots) and non-seizure (blue dots) data points in two dimensional feature space using PCA. (c) distribution of preservation ratio by t-SNE and PCA during data mapping into 2D feature space.

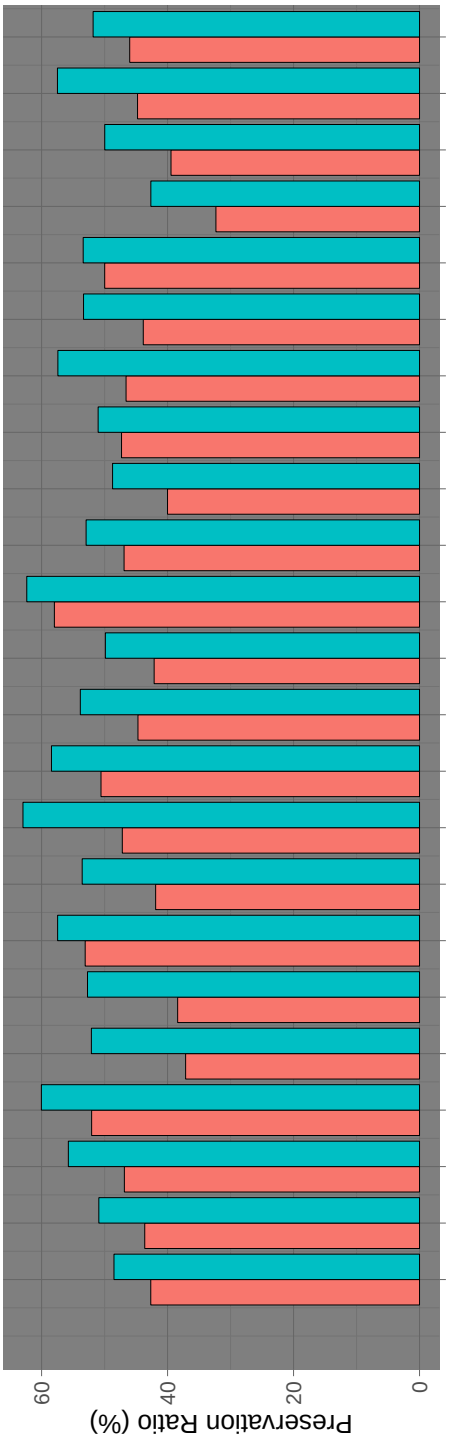


Figure 2.6. Per subject averaged nearest neighbor preservation ratio for 2D t-SNE (blue) and 2D PCA (pink)

2.4.1 t-SNE

t-SNE employs a pairwise differentiation to all the data points and efficiently minimizes the high-dimension and low-dimension data joint distributions. More importantly, the initial distance relationships among data points are reproduced during data embedding and are thus presented in the transition to low dimension. Let us define $X = [x_1, x_2, ..., x_{N_D}]^T$ as our initial high-dimensional dataset including N_D data points where x_i represents M-dimensional (M-feature) $x_i = [x_{i,1}, ..., x_{i,M}]$ instances. The final goal is estimating a two dimensional $y_i = [y_{i,1}, y_{i,2}]$ representation in a way that pair-wise distances are preserved. Note that the two new dimensions are completely different from any of the initial dimensions and are defined to represent data points in low dimension such that the key similarity/dissimilarity in high-dimension space is preserved.

$$p_{i,j} = \frac{p_{i|j} + p_{j|i}}{2N_D} \tag{2.3}$$

Metric $p_{i,j}$ represents a pair-wise distance probability in original space. The objective is to extract $p_{i,j}$ from each pair in a way that if x_i and x_j are close together, $p_{i,j}$ has a higher value. The distance relationship between data points x_i and x_j is modeled by the conditional probability $p_{i|j}$. This relationship is modeled using a t-student distribution as follows:

$$p_{i|j} = \frac{e^{-d(x_i,x_j)^2/2\sigma_i^2}}{\sum_{k \neq i} e^{-d(x_i,x_k)^2/2\sigma_k^2}} \qquad p_{i|i} = 0 \tag{2.4}$$

where σ_i represents the Gaussian distribution bandwidth centered at x_i [60]. $d(x_i, x_j)$ denotes the Euclidean distance between two data points x_i and x_j in their original feature space. $p_{i,j}$ is calculated from $p_{i|j}$ and $p_{j|i}$ as follows:

The original joint distribution set $P = \{p_{1,1}, p_{1,2}, ..., p_{N_D,N_D}\}$ contains all the pair-wise probability $p_{i,j}$ values. Similarly, $q_{i,j}$ represents the distance relationship between two data points x_i, x_j in low-dimensional feature space. Parameters $q_{i,j}$ are calculated as follows:

$$q_{i,j} = \frac{(1 + ||y_i - y_j||^2)^{-1}}{\sum_{k \neq l} (1 + ||y_k - y_l||^2)^{-1}} \qquad q_{i,i} = 0 \tag{2.5}$$

All the individual $q_{i,j}$ values make the set of joint probability distributions in low dimension $Q = \{q_{1,1}, q_{1,2}, ..., q_{N_D,N_D}\}$. Note that $||.||$ symbol indicates the Euclidean distance. The Kullback-Leibler divergence (D_{KL}) model is used to calculate information lost while estimating Q using P [23]. We optimize D_{KL} to minimize the difference between Q and P:

$$D_{KL}(P|Q) = \sum_{i,j}(p_{i,j} \times log\frac{p_{i,j}}{q_{i,j}}) \tag{2.6}$$

The optimal D_{KL} is calculated by the Gradient descent optimization (Equation 2.7) [60]:

$$\frac{\partial D_{KL}}{\partial y_i} = 4\sum_{j}(p_{i,j} - q_{i,j})(y_i - y_j)(1 + ||y_i - y_j||^2)^{-1} \tag{2.7}$$

This iterative technique is able to find the optimum value $y_i = [y_{i,1}, y_{i,2}]$ per data point in the new feature space, which minimizes the divergence between high and low dimensional data.

2.4.2 Preservation Ratio

Quality assessment of dimension reduction helps us understand the degree of accuracy and the extent of information loss caused by mapping high dimensional data points into a low dimension feature space [38]. We ran a neighborhood preservation ratio experiment [19] to evaluate the performance of t-SNE as a nonlinear dimension reduction technique in comparison with a classic linear dimension reduction method like PCA. Consider K_i^H and K_i^L as our list of λ nearest neighbors for each data point ($1 \leq i \leq N_S$) in high and low dimensional feature spaces, respectively. The preservation ratio PR is calculated as the percentage of common neighbors in low and high dimensions over the total number of neighbors. The preservation ratio is calculated by averaging the preservation of each individual data point:

$$PR = \frac{100}{N_S} \times \sum_{i=1}^{N} \frac{|K_i^L \cap K_i^H|}{|K_i^H|} \tag{2.8}$$

where we chose $\lambda = |K_i^H| = |K_i^L| = 100$ as suggested in [19] to be a reasonable number of nearest neighbors for preservation ratio analysis. Note that the perfect value for preservation ratio (PR) is 100, indicating that all the nearest neighbors are identical in both high and low dimensions. In order to show the effectiveness of our proposed data-embedding technique, we performed the per subject neighborhood preservation analysis for both t-SNE and PCA. We extracted the preservation ratio after mapping the original feature space into two dimensional feature spaces.

Figure 2.5 (c) compares the distribution of the preservation ratio by t-SNE and PCA during data mapping into 2D feature space for Patient 1. It shows that t-SNE (blue curve) provides a higher preservation values than does PCA (pink curve). Figures 2.5 (a) and (b) visualize the data points in two dimensional feature space using PCA and t-SNE, respectively. As is shown, seizure data points are distributed all around the non-seizure data points using PCA dimension reduction. On the other hand, seizure data points are well-grouped at a corner using t-SNE dimension reduction. This is because t-SNE preserves the local neighborhood information much better than does PCA. Thus, non-linear dimension reduction techniques for analysis of EEG data for epileptic patients works much better because the data has a non-linear structure. As Figure 2.6 shows, t-SNE provides a higher preservation ratio value than PCA for all the subjects. Although t-SNE needs a more complex model than PCA, it achieves a high accuracy by preserving local information. A second advantage of using t-SNE is that we will be able to visualize the data (i.e. the seizure/non-seizure windows). This visualization can improve physician's understanding of long-term data (e.g. progress/length of seizures episodes) by a quick and effective visual inspection of 2D color-coded images like Figure 2.5 (c).

2.5 Classification

For each patient, we selected 8 hours of known nonseizure data that was at least 30 minutes from the nearest seizure. For this group of patients, seizure duration varies from 10 to 120 seconds. We segmented EEG signals into 10 second windows to accommodate the shortest seizures. We refer to data points (instances) as segmented windows, which are used for spectral feature extraction. After channel selection and dimension reduction, each data point is represented by only two features (2D t-SNE). We use a k-nearest neighbor classification model (KNN) to differentiate seizure and non-seizure events in the new 2D feature space.

EEG signal response to seizure is different from patient to patient. So, we train the classifier in a patient-specific style using each patient's own data. This personalized approach increases the seizure detection performance for each patient. Personalization can be implemented by using flexible electrode positioning. Currently, some companies (e.g., Epitel [31]) are developing single channel EEG nodes for real life data collection. Note that the proposed method uses full channel EEG data, collected in a clinical setting (e.g in Epilepsy Monitoring Unit), for channel selection. Since channel selection is completed prior to fitting the patient with his/her personalized wearable device, the random forest algorithm will not run on the wearable device. For each patient, our method requires a group of training seizures for channel selection However, in some cases, there may not be sufficient data during the patients stay at the EMU because, for example, the patient has seizures only once or twice a month. In this case, channel selection will be done based on the patient's limited data.

2.6 Experimental Result

2.6.1 EEG Dataset

Twenty three epileptic patients' EEG manifestations from the open access scalp EEG database is used in this work. The dataset was recorded from twenty three pediatric patients with

intractable seizures at the Childrens Hospital Boston [54]. The dataset is available at the PhysioNet website: www.physionet.org/pn6/chbmit [30]. Data collection was done while patients were off anti-seizure medications to capture EEG responses to seizure and thereby determine which patients were candidates for surgery. The data is collected based on the international 10-20 system of EEG electrodes. The 10 and 20 indicate that the distances between adjacent electrodes are either 10% or 20% of the total front to back or right to left distance of the skull. The EEG waveforms were collected in a differential montage in which each signal represents the difference between two adjacent electrodes in the 10-20 standard system as shown by the connecting lines in Figure 2.7. For example, C3-P3 represents the electrical potential between electrodes C3 and P3. EEG was sampled at 256 Hz and provides 16-bit resolution. The beginning and end of each seizure event was annotated by experts and is used as ground truth in this work. Table 2.2 shows the Patients' information including age, gender, number of seizures, duration of seizure EEG, and number of channels used for data collection.

2.6.2 Study Design and Performance Evaluation

In this study, we use cross-validation analysis to show how accurately our proposed limited channel selection method detects seizures in practice. We divided the entire dataset into complementary subsets, applied the analysis to each subset and validated the results on the other subset. In order to decrease variability, we repeated cross-validation analysis multiple times using different partitions, and averaged the results of the validation phase over all the individual cross-validation analyses.

In order to report the performance of our proposed seizure detection technique we used sensitivity, precision, False Positive Rate (FPR) per hour and F-measure parameters. Sensitivity is the number of correctly detected seizure instances over all the instances detected as seizure. Precision is the number of correctly detected seizures over all the seizure instances.

Table 2.2. Clinical Study Information (MIT dataset).

Patient ID.	Age	Gender	No. Channels	No. Seizures	Seizure (MM:SS)
1	11	F	23	7	7:10
2	11	M	23	3	2:50
3	14	F	23	7	5:40
4	22	M	23	4	4:00
5	7	F	23	5	9:00
6	1.5	F	21	10	2:00
7	14.5	F	21	3	5:10
8	3.5	M	23	5	15:10
9	10	F	21	4	4:00
10	3	M	21	7	6:50
11	12	F	23	3	13:20
12	2	F	23	27	14:50
13	3	F	18	12	8:10
14	9	M	23	8	2:30
15	16	F	22	20	27:20
16	7	F	23	8	1:20
17	12	F	23	3	4:40
18	18	F	23	6	4:50
19	19	F	23	3	3:40
20	6	F	23	8	3:30
21	13	F	23	4	3:10
22	9	F	23	3	3:10
23	6	F	21	7	6:40

FPR per hour is the number of non-seizure instances incorrectly detected as seizure over the total duration of non-seizure EEG. F-measure is a harmonic mean of precision and sensitivity metrics [47]. The seizure dataset is grossly unbalanced since each patient had seizures for a relatively short time. So, the overwhelming majority of the instances belong to the non-seizure class. Therefore, performance metrics (i.e. sensitivity, precision, FPR per hour and F-measure), which are not affected by the number of correctly classified non-seizure instances class are expected to be more reliable parameters [11].

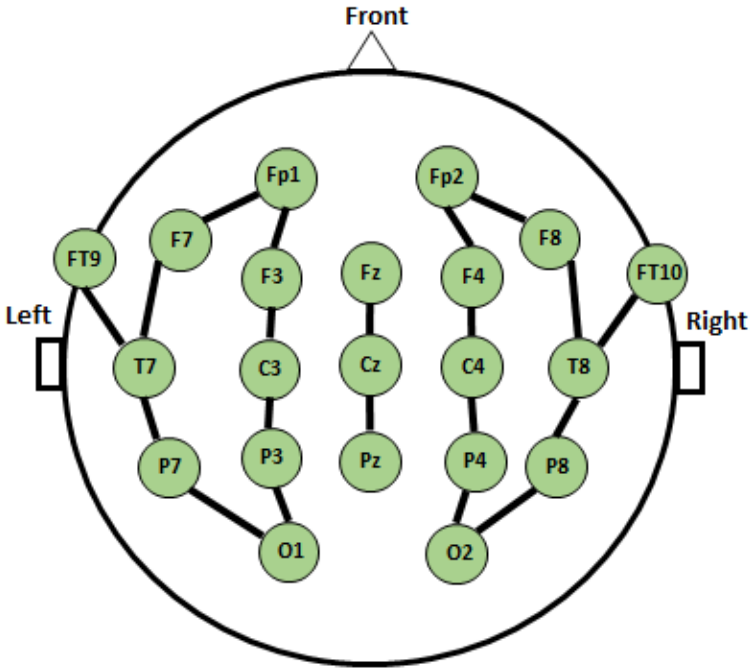


Figure 2.7. International 10-20 system of EEG electrode replacement.

2.6.3 Discussion

Figure 2.8 illustrates how effectively different channel selection techniques separate seizure from non-seizure data for Patient 1. This 2D representation is obtained by applying t-SNE dimension reduction to channels selected by different channel selection techniques. Figure 2.8 shows that seizure data points (red circles) are well-separated using our proposed random forest based channel selection. Figure 2.9 shows the difference in sensitivity ($\Delta S = S_{full} - S_{k=3}$) between full channel and limited channel seizure detection for different channel selection techniques. Figure 2.10 compares the F-measure of seizure detection among different channel selection techniques. Note that the top, middle, and bottom horizontal lines in each boxplot represent the first quartile, median, and third quartile of corresponding F-measure scores among all the patients, respectively. Moreover, the vertical lines at the two ends of each boxplots show the range of F-measure score variation among all the patients.

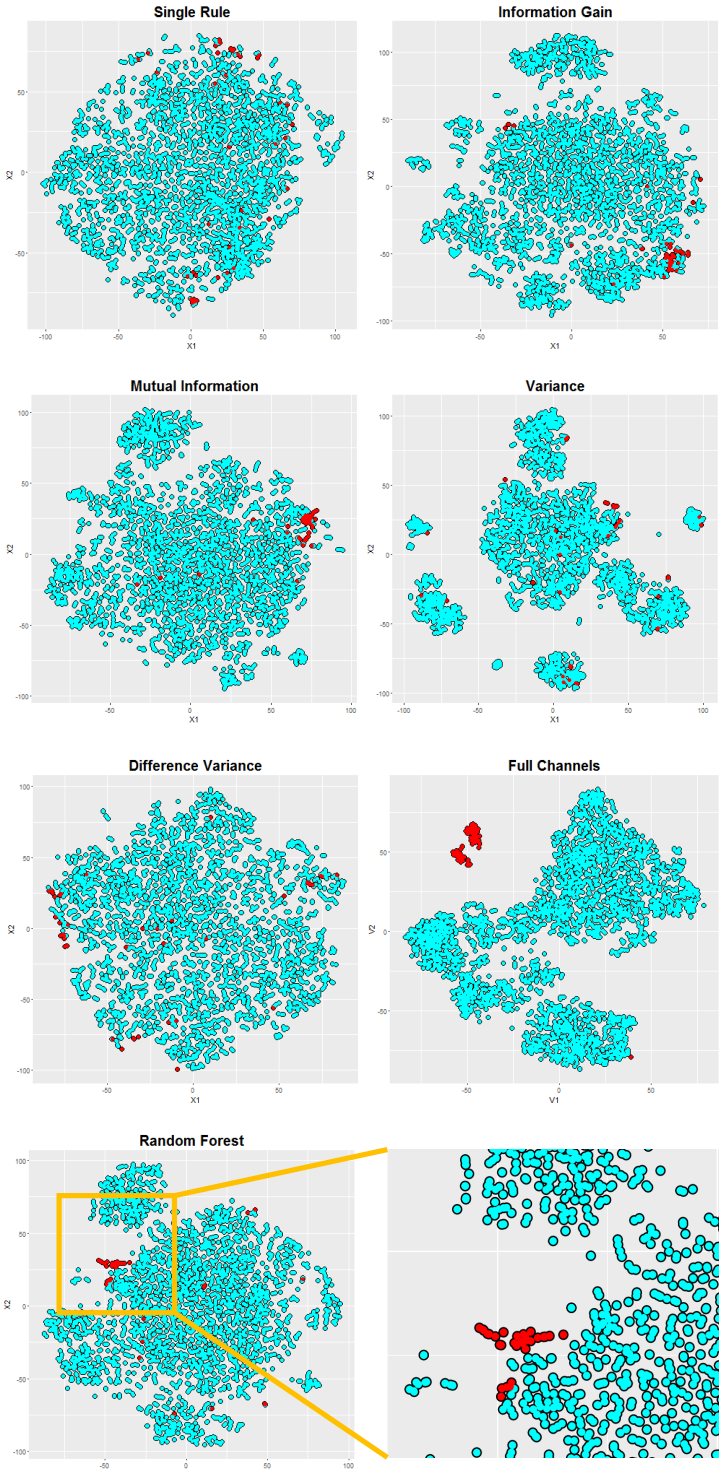
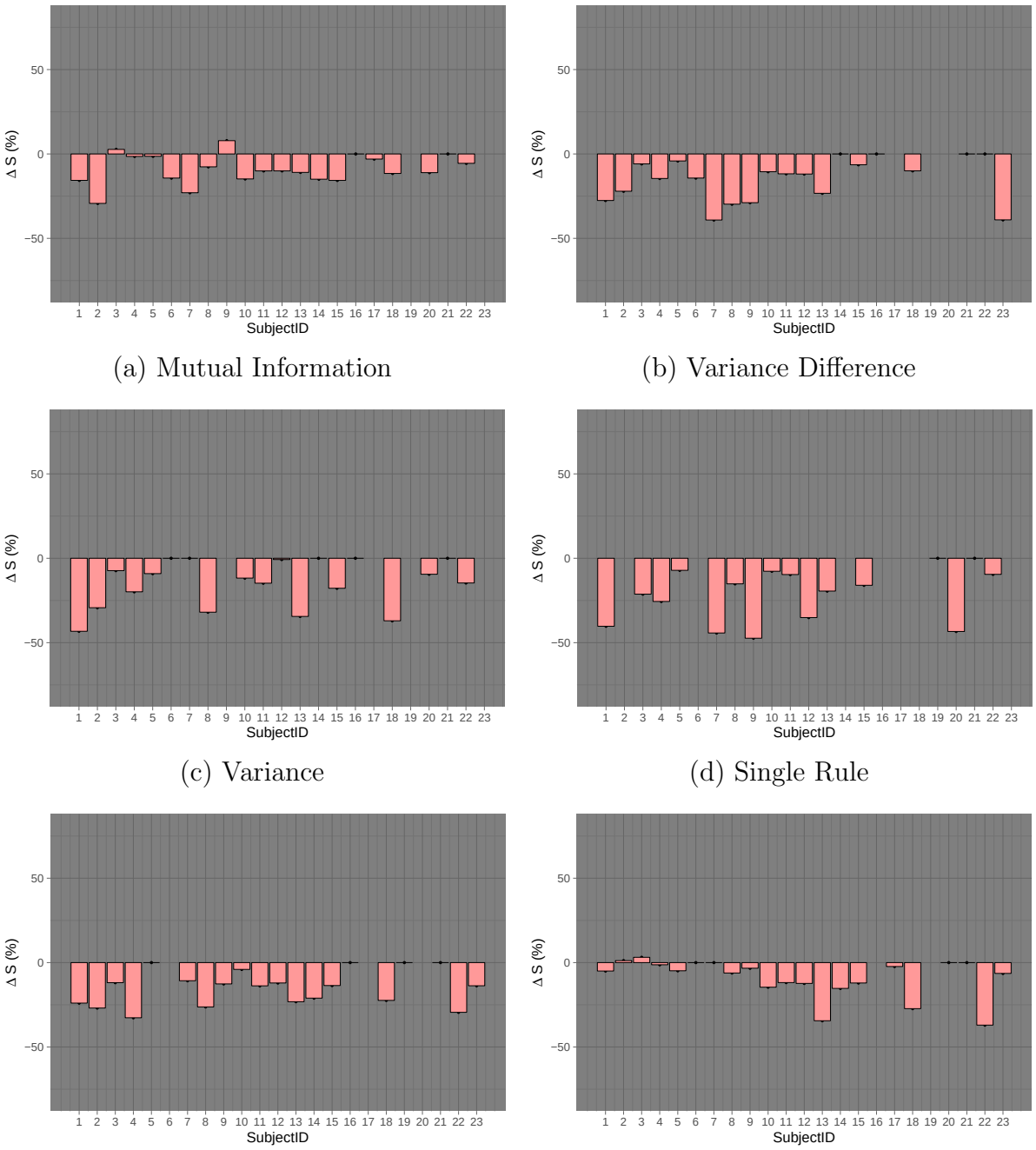


Figure 2.8. 2D t-SNE representation of seizure and non-seizure events for Patient 1 using different channel selection techniques: single rule [61], information gain [20], mutual information [5], variance [48], variance difference [25], and proposed random forest method.



(a) Mutual Information

(b) Variance Difference

(c) Variance

(d) Single Rule

(e) Information Gain

(f) Random Forest

Figure 2.9. A comparison between sensitivity of random forest based channel selection technique and other techniques.

Table 2.6.3 represents the average accuracy of epileptic seizure detection calculated by 10-fold cross validation with 10 times reputation. It shows that the performance of random forest channel selection and t-SNE (this work) for seizure detection is better than that of other methods. The proposed two-stage, low dimension EEG based epileptic seizure detection provides better performance. Note that the results reported in Table 2.6.3 are based on our own implementation of works reported in [61], [48], [25], [20], and [5] in R environment [49].

The last column of Table 2.6.3 shows the accuracy obtained using dimension reduction only. The results compare favorably with the results obtained using both RF channel selection and dimension reduction for two key reasons. First, not all EEG channels are relevant to the epileptic seizure. Although we missed some brain activity information using a limited number of channels, the missing information is irrelevant or redundant in most cases and it does not affect our ability to detect a patient's seizures. Second, applying a linear classification (e.g. KNN) to raw high dimensional data may not capture the similarity of seizure data points very well since the notion of distance is very different in high dimensions. Although there is always information loss in any reduction technique, t-SNE successfully preserves the local neighborhood information between instance pairs. This characteristic of t-SNE enables it to preserve accuracy during the dimension reduction process.

2.7 Summary

Full channel EEG analysis constrains the performance of seizure detection since it provides extremely high dimensional feature spaces. Confounding effects of irrelevant EEG channels can decrease the performance of seizure detection techniques. Generating a large number of random decision trees using features coming from all the EEG channels is an efficient way to select a small number of EEG channels that can most effectively detect the patient's seizures. Relevant and informative features appear in those trees more frequently than do redundant features. We choose the channels with the highest contribution in the entire forest for use

Table 2.3. Average Sensitivity, Precision, FPR per hour and F-measure for different channel selection techniques.

Method	**Single Rule [61]**	**Variance [48]**	**Variance Difference [25]**	**Information Gain [20]**	**Mutual Information [5]**	**Random Forest [This work]**	**All**
Sensitivity	53.43	60.71	63.75	63.79	73.04	80.87	89.80
Precision	23.23	28.52	28.28	36.67	40.47	47.45	48.98
FPR per hour	3.5	3.3	3.2	2.9	2.8	2.5	2.2
F-Measure	30.3	37.01	36.78	45.07	50.34	56.23	61.5

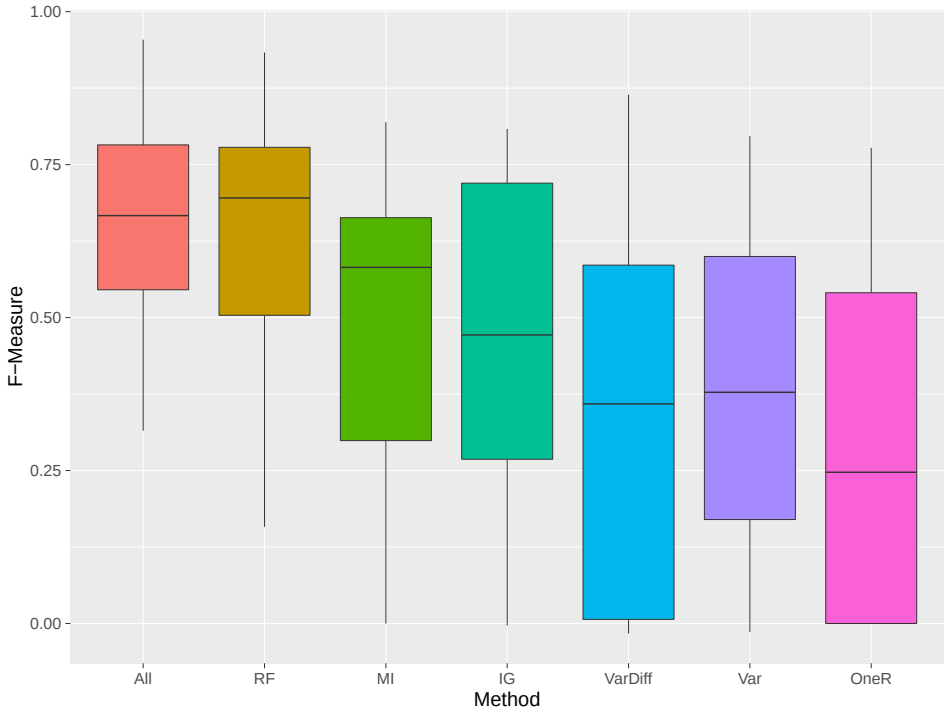


Figure 2.10. Comparison of channel selection accuracy (F-measure).

in limited channel seizure detection. After channels selection, the feature space is still large since each channel provides multiple frequency domain features. Classifiers cannot properly extract information hidden in the high dimensional feature space. Non-linear dimension reduction techniques are good candidates for tackling this problem. We grouped seizure data points using t-SNE, a non-linear dimension reduction technique, because it preserves the data's neighborhood information. The combination of our channel selection approach and our nonlinear dimension reduction technique provides excellent seizure detection accuracy as it effectively represents seizure and nonseizure data points as separable groups.

CHAPTER 3

IMBALANCE LEARNING[1]

Around 1% of world's population suffer from epileptic seizures which can lead to injuries and even unexpected death. We used EEG signals, which are proven to be the best indicators of seizures. However, the limited availability of seizure events in the EEG data makes it difficult for the automatic classifiers in general to accurately classify seizure events. To improve this, we propose two imbalance learning approach to improve accuracy of highly imbalanced seizure dataset. First, since each patient provides a different response to the seizure, we personalize the neural network classification models in terms of training data and model parameters. Second, we used the notion of clustering to group the non-seizure class into different clusters and consider the seizure class with each cluster separately. Both suggested imbalance learning techniques improve the accuracy of seizure detection.

3.1 Imbalance Learning Using Customized Neural Network

3.1.1 Introduction

To confirm the diagnosis of epilepsy in patients, ElectroEncephaloGraphy (EEG) is used to record electrical activity in the brain. EEG measures voltage fluctuations resulting from ionic current within the neurons of the brain. Routine, non-invasive EEG requires placement of electrodes on the scalp of the patient under observation. Epileptiform abnormalities like spikes and sharp waves in the EEG support a propensity for epileptic seizures [43].

Often, even when the EEG recording is prolonged, EEG monitoring produces only a few captured seizure events during the patient stay in the Epilepsy Monitoring Unit (EMU).

[1]Birjandtalab, J., Jarmale, V. N., Nourani, M., Harvey, J. (2018, October). Imbalance Learning Using Neural Networks for Seizure Detection. In 2018 IEEE Biomedical Circuits and Systems Conference (BioCAS) (pp. 1-4).

Birjandtalab, J., James, M., Nourani, M., Harvey, J. (2018, October). Learning from Non-Seizure Clusters for EEG Analytics. In 2018 IEEE Biomedical Circuits and Systems Conference (BioCAS) (pp. 1-4).

Therefore, the EEG data available to the Machine Learning (ML) algorithms, that attempt to learn from the data, is dominantly one *majority* (non-seizure) class. So researcher's main challenge is to extract as much learning as possible from the *minority* (seizure) class. A conventional method to achieve this would be to over-sample or under-sample data points of minority and majority classes, respectively. A more effective way, that we advocate in this work, would be to empower the learning part of popular ML algorithms to learn and respond to the imbalance, without changing anything in the data.

In this chapter, we study the performance of Artificial Neural Networks (ANNs) while training on the data extracted from the power spectral density of EEG signals. If an ANN is trained on data where one of the classes is dominant compared to other classes, its classification capabilities will not be very accurate. We address this issue by changing the cost function, compare results between the old and the new cost functions and also juxtapose the results when the system learns from imbalanced EEG data. Notice that we are not doing anything to change (e.g. replicate) the data. Instead, we change the way we learn from the imbalanced data. This is fundamentally different from oversampling and under-sampling which inherently change the data that neural networks learn from.

3.1.2 Methodology

The general view of our proposed methodology is shown in Figure 3.1. A full-channel (often 18 to 24 electrodes) scalp EEG dataset is used as input of the model. We used a spectral analysis to extract features from time domain EEG signals. The details of feature extraction is explained in Section 2.2. Then, we modified the cost function of the artificial neural network to address problem of imbalance dataset. Finally, we classify seizure and non-seizure events for each patient [10].

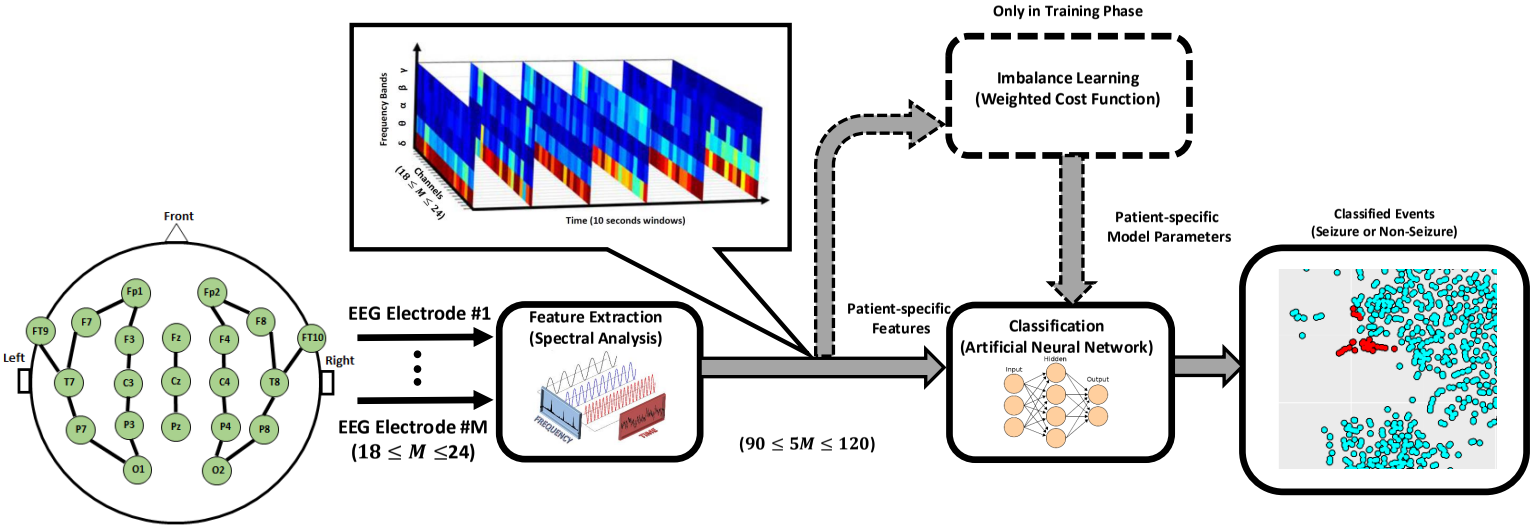


Figure 3.1. The general view of our proposed methodology.

Artificial Neural Networks

ANNs are inspired from the networks in brains, which learn from examples and without the requirement of task specific programming. ANNs consist of layers of artificial neurons which are generally densely connected with each other. The connections between neurons in ANNs are equivalent to the synapses between neurons in brain. When neural networks have more than one hidden layer, i.e. more than three including the input and output layers, they are often called Deep Neural Networks (DNNs) [40].

Considering the input layer to be the 0^{th} layer of the DNN, the L hidden layers of a DNN to be numbered from $1, \ldots, L$, and the output layer of DNN to the $(L+1)^{th}$ layer, we have the following equation [40]:

$$\boldsymbol{v}_l = f(\boldsymbol{z}_l) = f(\boldsymbol{W_l}\boldsymbol{v}_{l-1} + \boldsymbol{b}_l) \tag{3.1}$$

where $\boldsymbol{v}_l$ is the output vector of layer l, $\boldsymbol{z}_l$ is the activation vector of the l^{th} layer with predetermined number of neurons, $\boldsymbol{W}_l \in \mathbb{R}^{\mathrm{N_l} \times \mathrm{N_{l-1}}}$ is the weight matrix and $\boldsymbol{b}_l \in \mathbb{R}^{\mathrm{N_l}}$ is the bias vector. Function f here is the activation function in layer l of the ANN. The Softmax activation function is used in the output layer of an ANN classifier to determine which of the classes does the data-point under examination belong to [40].

3.1.3 ANN architecture for imbalance learning

ANN Architecture

For the number of neurons in the hidden layer in our neural network, we have used a popular thumb rule for the number of neurons in the hidden layer, which is, to be somewhere between the number of neurons in the input layer and the output layer [16], i.e. $N_{hidden} = \left\lceil \frac{N_{input}+N_{output}}{2} \right\rceil$. So, in our experiments, we have 46 to 61 $((90+2)/2 to (120+2)/2)$ hidden

ReLU (Rectified Linear Unit) neurons in the single hidden layer of the network depending upon the number of input features available for each patients' data.

We have chosen to have only one hidden layer since adding just one hidden layer turns a linear ANN (a linear ANN is one with no hidden layer) to a non-linear ANN [6], which means, with just as little as one hidden layer, neural networks are capable of being a universal approximator [22].

While training our ANNs, one of the most important steps is to back-propagate the cost (error) accrued across N training data points in the batch. It is defined as a function of the value predicted by the neural network and the target value, i.e. the value the ANN is trying to learn and is generally denoted as J, called the cross-entropy cost function. When we have K classes, it is defined in the following manner [40]:

$$J = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} t_{i,k} log(\hat{y}_{i,k}) \tag{3.2}$$

In a binary classification problem, for N training data points, the average cross-entropy cost function is defined as:

$$J = -\frac{1}{N}\sum_{i=1}^{N}\Big[t_i log(\hat{y}_i) + (1-t_i) log(1-\hat{y}_i)\Big] \tag{3.3}$$

where, for training example i, t_i is the target class and $\hat{y}_i$ is the value predicted by the ANN.

Imbalance Learning

Note carefully that in justifying Eqn. (4) and (5) researchers assumed that the classes are not imbalanced, i.e. both classes in the training data get almost equal opportunity to be learnt. However, in many applications this is not true. For example, the real world EEG data, has more non-seizure points than seizure points. The imbalance ratio can be as high as 400, meaning, the number of non-seizure (majority class) points is 400 times the number of the

seizure points (minority) in the data. So, we modify the cross entropy cost function which attempts to level the opportunity for the minority class to train with the majority class. One of the way to do that is to punish the misclassifications of the minority class harsher than the misclassifications in the majority class. To quantify this, we define Imbalance Ration (IR):

$$IR = \frac{|M|}{|m|} \tag{3.4}$$

where $|M|$ and $|m|$ are number of data points (samples) in majority and minority classes in the training dataset, respectively.

We multiply the minority class misclassification error by IR to give it proportional weight during training [64]. After making these considerations, the weighted cross entropy cost equation J_w becomes:

$$J_w = -\frac{1}{N}\sum_{i=1}^{N}\{t_i log(\hat{y}_i) + IR[(1-t_i)log(1-\hat{y}_i)]\} \tag{3.5}$$

By doing this, we are magnifying the error when the ANN misclassifies a datapoint of the minority class, forcing the neural network to back-propagate the error to ensure that the neural network adapts itself and ultimately gets the prediction of minority class correctly.

3.1.4 Experimental Results

In our work, we investigated the EEG dataset of 23 patients collected at the Childrens Hospital Boston (also known as MIT/Physionet dataset) [54]. The details of this dataset is explained in section 2.6.1.

We chose F-measure as the key metric to evaluate the training of the ANN. Note that, F-measure (i.e. $F_{measure} = \frac{2\times tp}{(2\times tp)+fn+fp}$) mixes true positives ($tp$), false positives ($fp$), and false negatives ($fn$). It does not intentionally take into the account the number of true negatives (tn), which, in our case, means we do not consider the number of non-seizure points classified correctly. This is better in imbalance learning investigation as our main focus here

is to classify the seizure points correctly. Table 3.1 lists the number of seizure points, the Imbalance Ratio (IR) and the average F-measure values when two cost functions F_{J_w} and F_J are used and the difference ΔF for each patient.

The values in Table 3.1 are the results of the following procedure carried out on each patient's dataset. We first shuffle the dataset and then divide it into 10 stratified folds (preserving the class imbalance ratio of the original dataset in each fold of the dataset). After performing 10-fold Cross Validation (CV) on ANN employing J, we obtain an average F-measure value F_J which goes into Table 3.1, for that patient. Similarly, we obtain the average F-measure value F_{J_w} using our improved cost function J_w. Finally, we calculate the difference $\Delta F = F_{J_w} - F_J$.

In Table 3.2, we consider the first six patients with the highest IR. For each of these patients' data, we separated the labeled seizure and non-seizure points. 100%, 50% or 33% of the seizure points were randomly sampled and all the non-seizure points were taken, shuffled to form a dataset. Just like the aforementioned procedure, we performed 10-fold CV on each of the three datasets per patient (100%, 50% and 33% of seizure points respectively per patient) to obtain an average F-measure value per dataset as listed in the table. For subjects 16 and 21, under the F_J column, we see values of average F-measure being 0.0 under high IR (33%) scenario. An average of 0.0 indicates that for every test fold, the F-measure observed was 0.0; which means, there were no true positive points predicted by the ANN. This happens because the number of seizure points to learn from, for the ANN, were completely insufficient.

Note that in 11 out of 23 cases we have $\Delta F < 0$ indicating that cost function J slightly produces better results. However, average of these cases is $\overline{\Delta F} = -0.05$. On the other hand, other 12 cases with $\Delta F > 0$ produce $\overline{\Delta F} = +0.082$. The average of all 23 cases produced $\overline{\Delta F} = +0.02$. All together, J_w still is more effective in dealing with imbalance EEG data.

Observing Table 3.2, we see that as we superficially reduce the number of seizure points, the F-measure performance of ANNs keeps going down, as expected. This is true for both J_w

Table 3.1. Comparison of average F-measure for all subjects

Subject	sz. pts.	IR	F_{J_w}	F_J	ΔF
16	8	360.00	0.70	0.40	0.30
6	12	240.00	1.00	0.80	0.20
14	15	192.00	0.80	0.83	- 0.03
2	17	169.41	0.90	0.77	0.13
21	19	151.58	0.67	0.53	0.14
22	19	151.58	0.87	0.82	0.05
20	21	137.14	0.94	0.88	0.06
19	22	130.91	0.90	0.91	- 0.01
9	26	110.77	0.90	0.93	- 0.03
17	28	102.86	0.66	0.78	- 0.12
18	29	99.31	0.79	0.76	0.03
7	31	92.90	0.75	0.80	- 0.05
3	34	84.71	0.90	0.92	- 0.02
5	35	82.29	0.87	0.90	- 0.03
4	36	80.00	0.85	0.83	0.02
23	40	72.00	0.97	0.96	0.01
10	41	70.24	0.99	0.97	0.02
1	43	66.98	0.96	0.97	- 0.01
13	49	58.78	0.84	0.93	- 0.09
11	80	36.00	0.99	0.99	0.00
12	89	32.36	0.69	0.84	- 0.15
8	91	31.65	0.87	0.84	0.03
15	180	16.00	0.89	0.90	- 0.01
Average	**42**	**68.64**	**0.86**	**0.84**	**0.02**

and J. However, we see that, even in case of limited availability of data to learn, $F_{J_w} > F_J$, sometimes giving significant values of ΔF. Except for subject 14, all cases show $\Delta F > 0.05$, in fact, by a large margin. This explains that using J_w can be important when we are looking at a patient with a very few seizure points. An average of these six patients also show significant improvement of $\overline{\Delta F}$ i.e. 0.13, 0.17 and 0.19 when 100%, 50% and 33% of the seizure-points are taken.

The classification accuracy is expected to be highly subjective. Yet, our method operates very well. A remarkable example of this is of subject 6 where just 12 points are enough to

Table 3.2. Top 6 patients for 100%, 50% and 33% of original IR

Subject	sz. pts. %	sz. pts #.	IR	F_{J_w}	F_J	ΔF
16	100	8	360.00	0.70	0.40	0.30
	50	4	720.00	0.25	0.0	0.25
	33	2	1440.00	0.11	0.0	0.11
6	100	12	240.00	1.00	0.80	0.20
	50	6	480.00	0.57	0.29	0.28
	33	4	720.00	0.31	0.01	0.30
14	100	15	192.00	0.80	0.83	- 0.03
	50	7	411.43	0.49	0.46	0.03
	33	5	576.00	0.35	0.31	0.04
2	100	17	169.42	0.90	0.77	0.13
	50	8	360.00	0.64	0.43	0.21
	33	5	576.00	0.39	0.06	0.33
21	100	19	151.58	0.67	0.53	0.13
	50	9	320.00	0.37	0.33	0.04
	33	6	480.00	0.14	0.0	0.14
22	100	19	151.58	0.87	0.82	0.05
	50	9	320.00	0.77	0.61	0.16
	33	6	480.00	0.46	0.23	0.23
Average	100	**15**	**192**	**0.82**	**0.69**	**0.13**
	50	**7**	**411.42**	**0.52**	**0.35**	**0.17**
	33	**5**	**576**	**0.29**	**0.10**	**0.19**

obtain an average F-measure of 1.00 (i.e. perfect classification). On the other hand, we have an example of subject 15, where we have 180 seizure points and would expect similar results (average F-measure nearing 1), but turns out the average F-measure in this case is 0.89.

3.2 Cluster based Imbalance Learning

EEG data collected in EMU is highly imbalanced and accuracy of automatic epileptic seizure detection is naturally low. Our aim is to increase the accuracy by reducing the imbalance ratio of seizure and non-seizure classes. We hypothesis that the non-seizure class itself includes various daily brain activities and then the data points are distributed as clusters in this class. In training phase, we propose a technique to cluster the majority (non-seizure) class

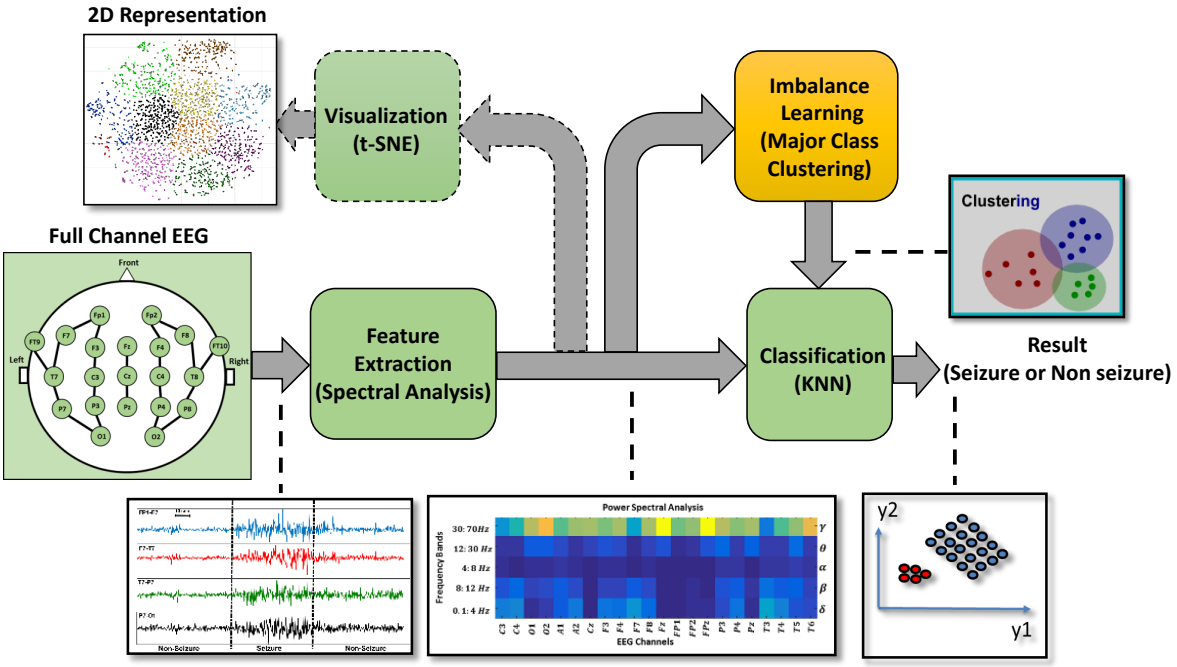


Figure 3.2. Overall view of our proposed methodology

into k clusters. Then, we train k KNN classifiers using each of the k non-seizure clusters plus seizure class. In the testing phase, we classify an incoming sample using this model and the non-seizure cluster closest to the incoming sample. We employed a state-of-the-art visualization technique to illustrate clusters of majority non-seizure class in two dimensions. The results, applied to MIT EEG dataset, show that our technique provides a higher average F-Measure accuracy.

3.2.1 Methodology

General view of our proposed model is shown in Figure 3.2. Inputs of this model are multi-channel scalp EEG signals. A frequency domain feature extraction technique is used to convert time domain EEG signals to meaningful features. The details of feature extraction is explained in Section 2.2. Finally, a patient-specific clustering based imbalance reduction

technique followed by K-Nearest Neighbour (KNN) classification technique detects seizure and non-seizure events. We use t-SNE for visualizing the clustered data.

Clustering Based Imbalance Learning

This is a critical data prepossessing technique to reduce class imbalance by clustering the majority class. The steps of the proposed algorithm are [9]:

- **Majority-Class Clustering:** The non-seizure data is clustered into k (k is subject dependent determined practically) clusters using K-means clustering. A copy of the remaining seizure data is assigned to each of the obtained clusters. The resultant structure would be k groups of data, each containing one cluster of non-Seizure data appended with all of the seizure data.

- **Multi-Model Training:** The input for training would be each of the k datasets, each of them being a mixture of non-seizure and seizure data. Our objective for preparing k models is to use them in the testing phase.

- **Model Selection (During Test):** To improve accuracy of classification, we would have to test each samples in the testing set against the appropriate model. To find the best model to test a data-point, we find the *closest cluster* to each sample in the test set by calculating the Euclidean distance to the center of each cluster and select the corresponding model. Algorithm 1 summarizes the steps of the proposed clustering based imbalance learning technique.

Classification

After clustering the majority (non-seizure) class, we generate multiple KNN classifiers with an aim to classify seizure and non-seizure events. EEG time series provide a different seizure pattern which is highly patient dependent. This is the reason that we applied personalization

(patient-specific approach) by using the patient's own data for both clustering and classification phases. Table 3.3 shows the original Imbalance Ratio (IR) of each patient in MIT dataset [54]. For each patient, we use 8 hours of known non-seizure EEG data that was at least 30 minutes from the nearest seizure. For this group of patients, duration of seizure events varies from 10 to 120 seconds. Finally, we applied patient dependent 10-fold cross-validation to validate our suggested technique.

Visualization

The visualization of this data is done for the purpose of validation and development. It shows how different clusters would form in the non-seizure data and for the depiction of seizure data among them. As our data is in high dimension (90-115), t-Distributed Stochastic Neighbor Embedding (t-SNE) for dimensionality reduction that is particularly well suited for the 2D visualization of high-dimensional datasets is applied to the whole data [60]. Figure 3.3 shows an example of non-seizure and seizure data points in 2D for a specific patient. The new dimensions (V1 and V2 in Figure 3.3) are different from any of the original dimensions. In fact, they represent two artificially built dimensions, which are modeled by distribution of similar/dissimilar samples in high-dimensional feature space [12]. The non-seizure data is clustered using K-means clustering with k=10. All seizure data are assigned to a separate cluster. In the following subsection, we briefly discuss the t-SNE. The detailed explaintion of t-SNE is mentioned in Section 2.4.1.

Algorithm 2 Algorithm 1: Clustering Based Imbalance Learning

Data: Highly imbalanced EEG data
Result: Classified events
Divide the data into *TRAIN* and *TEST*
Separate Non-Seizure data N and Seizure data S in *TRAIN*
K-means Clustering(N) into $N_1, N_2, ...N_k$
while for each of $N_1, N_2, ...N_k$ in TRAIN **do**
$C_i \leftarrow N_i \bigcup S$
$Model[i] \leftarrow KNN(C_i)$
end
for each record in *TEST*
for each cluster
$d[j]$ ←Find distance of data point from cluster centers
end
$minIndex \leftarrow Min(d[j])$
Predict class for the record using $Model[minIndex]$
end
Calculate F-measure

3.2.2 Experimental Result

The EEG data of twenty three epileptic patients from the open access scalp EEG signals is used in this work which are recorded at the Childrens Hospital Boston [54]. The details of this dataset is explained in section 2.6.1.

Evaluation

For performance measurement, we calculate the confusion matrices for each fold. From this the F-measure is calculated using the following formula:

$$F_m = \frac{2TP}{2TP + FN + FP} \tag{3.6}$$

Here, TP = True Positive, FP = False Positive, FN = False Negative, TN = True Negative. F-measure shows a harmonic mean of precision and sensitivity numbers. Even after applying imbalance learning techniques, seizure dataset is highly imbalanced because each patient

have seizures for only short amount of time. So, most instances still belong to the majority class. This is the reason that we select F-measure as an appropriate metric for reporting the accuracy. This metric is not affected by the number of correctly classified non-seizure class are are considered more reliable metrics in this work.

Discussion

We performed both clustering and classification tasks using R-Studio software tool [57]. We compared the seizure detection accuracy for different number of clusters. Since there is no side information about number of clusters in the majority class, we set number of clusters based on two scenarios. In the first scenario, we assumed that there are $k = 10$ clusters in the non-seizure group. This number $k = 10$ is only an estimation on the number of typical daily brain activities. In the second scenario, we ran the experiment for different number of clusters in range between 1 to 10 clusters and found the best k comparing the classification accuracy. Table 3.3 shows the F_m without clustering (3rd column) and with clustering non-seizure class for each patient (both $k = 10$ and best k) in 4th and 6th columns, respectively. Our proposed clustering based imbalance learning technique has improved the accuracy of seizure detection from 0.58 to 0.74 (27% improvement) on average. The rationale is that by clustering the non-seizure class, we potentially find different brain activities. So, the chance of correctly classifying a seizure event against clusters of brain activities is higher than classifying it against treating all non-seizure activities the same way.

3.3 Summary

While, EEG signal is known as the best indicator of epileptic seizures, there are classification challenges. Specifically, the highly imbalanced nature of EEG signal which provides only few seizure events compared to a large number of non-seizure events. We addressed this

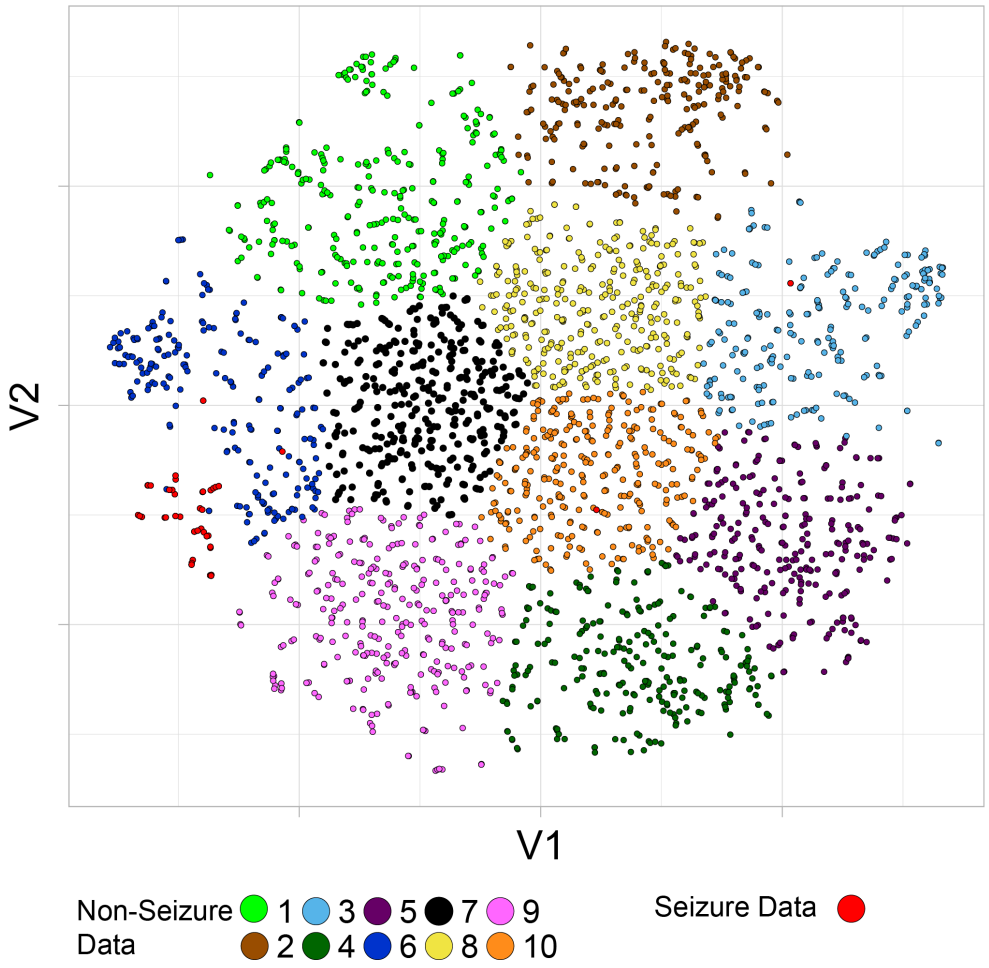


Figure 3.3. 2D visualization of seizure (red dots) and non-seizure (colorful dots) points.

issue by proposing a customized ANN with personalized cost function that learns from the imbalanced data and achieves high accuracy in automated seizure detection.

Real world EEG data is highly imbalanced in a way that they suffer from lack of sufficient seizure events. Since the non-seizure class contains various daily brain activities, we proposed to cluster non-seizure class into k clusters in the training phase. Then, we train k KNN classification models. In testing, we classify a new data point using the classifier which its corresponding non-seizure cluster is the closest to the new data point. We used an efficient

Table 3.3. Experimental Results

Sub	**IR**	F_m **(KNN)**	F_m **(KNN+k=10 Clusters)**	**Best k**	F_m **(KNN+best k Clusters)**
sub1	66.98	0.83	0.85	4	0.89
sub2	169.42	0.60	0.64	4	0.71
sub3	84.71	0.71	0.84	1	0.89
sub4	80.00	0.53	0.56	2	0.73
sub5	82.29	0.78	0.78	2	0.82
sub6	240.00	0.47	0.47	1	0.60
sub7	92.91	0.27	0.53	1	0.64
sub8	31.65	0.79	0.82	1	0.83
sub9	110.77	0.53	0.74	2	0.89
sub10	70.25	0.92	0.92	1	0.93
sub11	36.00	0.96	0.96	4	0.96
sub12	32.36	0.67	0.71	1	0.74
sub13	58.78	0.76	0.79	1	0.84
sub14	192.00	0.34	0.47	1	0.55
sub15	16.00	0.82	0.85	1	0.86
Sub16	180.00	0.40	0.67	8	0.73
sub17	102.86	0.56	0.62	3	0.69
sub18	99.32	0.33	0.43	1	0.43
sub19	130.91	0.40	0.40	7	0.62
sub20	137.15	0.39	0.48	7	0.62
sub21	75.79	0.07	0.17	8	0.35
sub22	151.58	0.54	0.62	2	0.83
sub23	72.00	0.69	0.75	3	0.85
Avg	**100.56**	**0.58**	**0.65**	$\simeq$ **3**	**0.74**

visualization technique to validate and visualize clusters of majority non-seizure class in 2D. The results shows that dividing the non-seizure class into clusters reduces the imbalance ratio and improves the seizure detection accuracy.

CHAPTER 4

IMPACT OF PERSONALIZATION ON EPILEPTIC SEIZURE PREDICTION[1]

The main contribution of this chapter is a personalization method which systematically selects the algorithm's parameters based on patient's individual data. The conventional seizure prediction techniques use a fixed set of parameters (like window size and time-to-seizure of preictal data). In this work, we report how personalizing the preictal data parameters improves the quality of seizure prediction. Experimental results show that using a personalized small set of parameters increases the F-measure accuracy of seizure prediction.

4.1 Introduction

Research has been carried out in the field of seizure detection and prediction for a long time in an attempt to improve the quality of life of epilepsy patients. Earlier, most of them involved experts studying seizure patterns with the help of EEG and other data available and formulating algorithms which would determine if a patient is going to experience a seizure or not. Osorio et al. published one of the early algorithms that achieved sensitivity and specificity over their entire dataset and allowed prediction of clinical onset of a seizure by a mean of 15.5 seconds in 92% of seizures [45]. They clearly stated in their paper they have not relied on training as a tool to develop, test and improve algorithmic performance. Their algorithm had the ability to quantify signal changes rather than comparing to determine whether or not the detected changes correspond to a seizure. The paper recommended adapting it to individuals, seizure types or individual data channels can possibly improve the speed of detection and the length of prediction times [45].

[1]Birjandtalab, Javad, V. Jarmale, and Mehrdad Nourani. "Impact of Personalization on Epileptic Seizure Prediction." 2019 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI). IEEE, 2019.

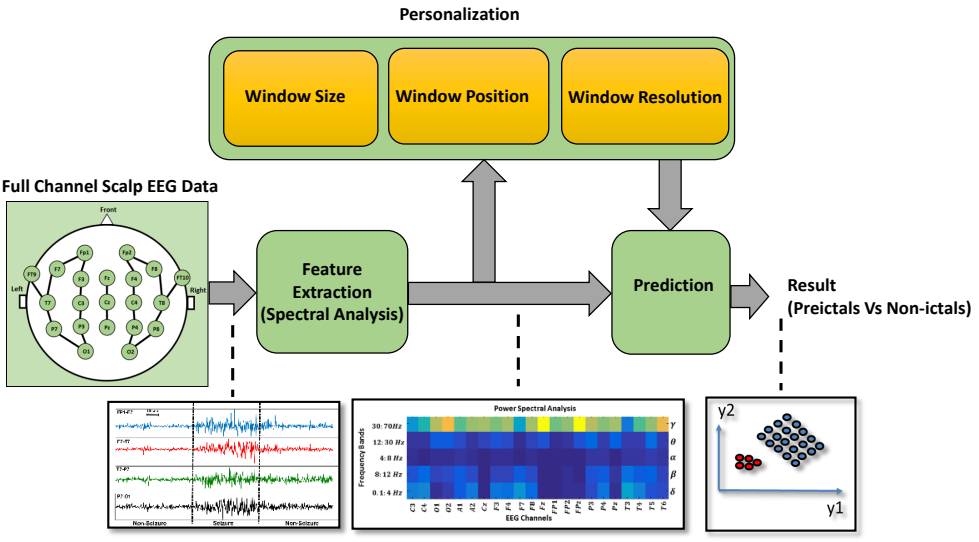


Figure 4.1. The general view of our proposed methodology.

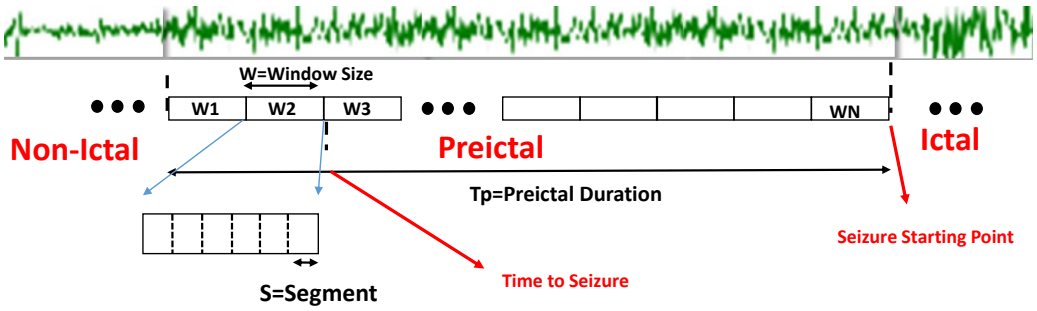


Figure 4.2. Preictal Window Selection.

4.1.1 Main Contribution

The main contribution of this chapter is a personalization method which systematically selects the seizure prediction parameters for each patient. When it comes to predicting the seizure, the main question is how to select the data prior to the seizure. The conventional seizure prediction techniques use a fixed set of parameters for pre-seizure data like window size and time-to-seizure. In this work, we show how personalizing the pre-seizure data parameter can improve the quality of seizure prediction. Our idea involves the concept of sliding windows of various window sizes to squeeze in the shortest monitoring time to accurately call an impending seizure. The general view of our proposed methodology is shown in Figure 4.1.

A full-channel (often 18 to 24 electrodes) scalp EEG dataset is used as input of the model. We used a spectral analysis to extract features from time domain EEG signals.The details of feature extraction are explained in Section 2.2. Then, we employed a window selection step to find the best portion of data for that particular patient to use for prediction. Finally, we classify preictal and non-ictal events for each patient.

Algorithm 3 Personalizing seizure prediction parameters

Input (i): Set of preictal window size $W_1, W_2, ...W_N$
Input (ii): Set of time-to-seizure $T_1, T_2, ...T_M$
Output: Personalized preictal window size W_p and time-to-seizure T_p
Separate non-ictal data X and initial preictal data P
For each window size $W_1, W_2, ...W_N$
For each time-to-seizure $T_1, T_2, ...T_M$
$P_{i,j} \leftarrow$ Extract preictal data using W_i and T_j parameters
$C_{i,j} \leftarrow P_{i,j} \bigcup X$
Train ANN classifier using $C_{i,j}$ and report F-measure $F_{i,j}$
$\{i_{max}, j_{max}\} \leftarrow$ indices which maximize $F_{i,j}$
end
end
$W_P \leftarrow W_{i_{max}}$
$T_P \leftarrow T_{j_{max}}$

4.2 Methodology

4.2.1 Personalization

We personalize the window size and time-to-seizure of preictal data. For this purpose, we initially consider five minutes (300 seconds) of data before each seizure episode as preictal window. Similar to Netoff, et. al, work [42], we have also chosen initial time period of five minutes before the occurrence of seizures as the preictal period for each patient. Then, we analyze smaller size windows, like 60, 90, 120, 180, 240 and 300 seconds (i.e, 1 minute to 5 minutes) to achieve the best performance on the data available for a particular patient. Algorithm 3 shows how personalized values of window size W_P and time-to-seizure T_P in

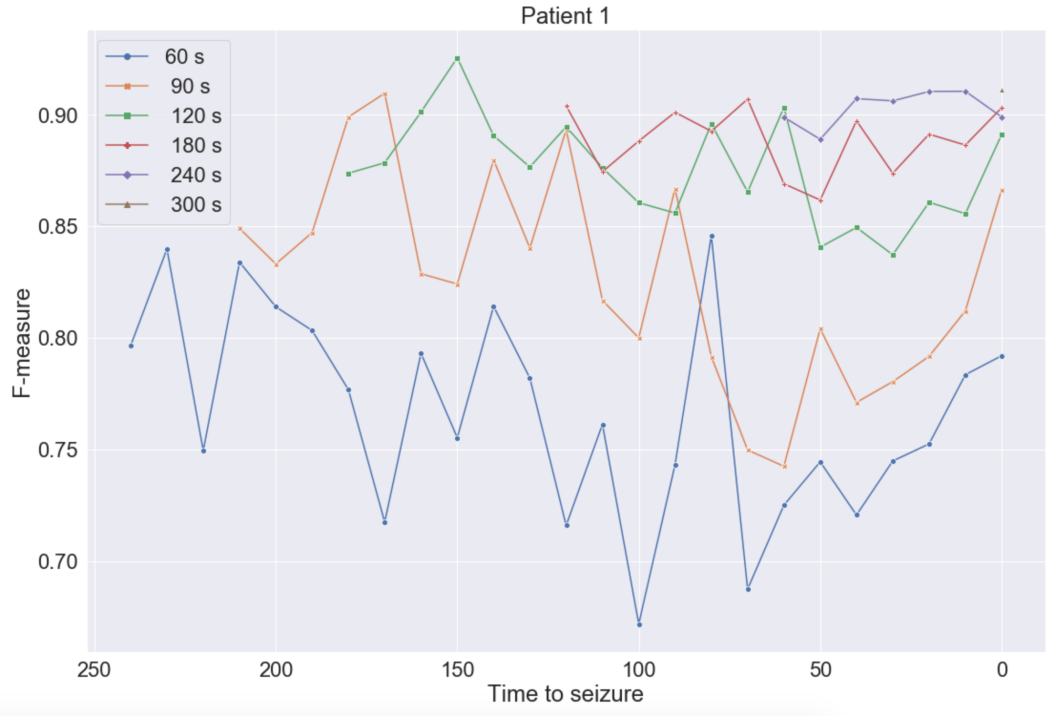


Figure 4.3. Accuracy of seizure prediction as parameter of time to seizure for different window sizes for patient 1 data.

preictal data are calculated. Figure 4.2 shows how we employ different window sizes and time-to-seizure for each patient. This process determines how long we should monitor, for each patient, before we can predict that the patient may have a seizure in a few minutes. For this, we have labeled the data points up to 5 minutes (300 seconds) before the actual seizure as preictal period data points and tried to learn the pattern of brain waves in the preictal period. So, we divided this preictal data before the seizure into various overlapping windows of size 60, 90, 120, 180, 240 and 300 seconds.

4.2.2 Prediction

We used Artificial Neural Network (ANN) configured for each patient using the data they recorded when put under observation in the Epileptic Monitoring Unit (EMU). The general architecture of the ANN is standard since the focus here is to discuss the technique of issuing

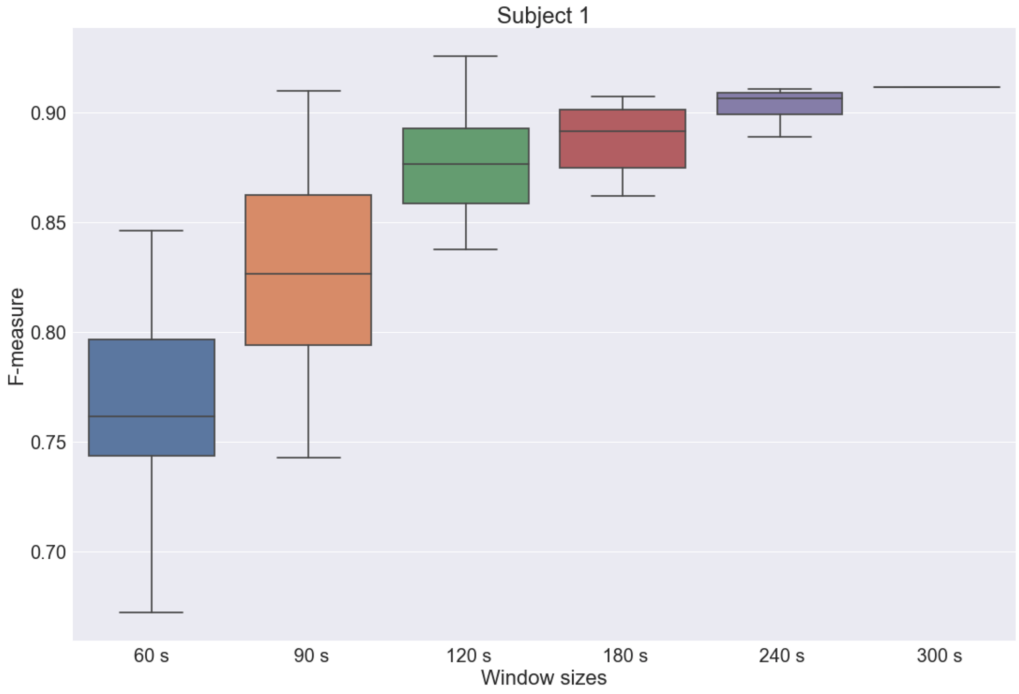


Figure 4.4. Distribution of F-measure recorded for patient 1

credible alerts with enough time for the patient/caregiver to respond. We have one hidden layer where the number of neurons are calculated using the popular thumb rule that the number of hidden neurons should be somewhere between number of neurons in the input layer and number of neurons in the output layer [16], i.e. $N_{hidden} = \left\lceil \frac{N_{input}+N_{output}}{2} \right\rceil$. With just one hidden layer, neural networks are capable of being universal approximators from a linear ANN [6][22].

4.3 Experimental Results

4.3.1 EEG Dataset

In our work, we analyzed the EEG dataset of 23 patients collected at the Childrens Hospital Boston (also known as MIT/Physionet dataset) [54]. In this dataset, each channel contains the difference of two adjacent electrodes collected in the sequential format. The EEG signals have a resolution of 16 bits and sampled with a sampling frequency of 256 Hz. The start and

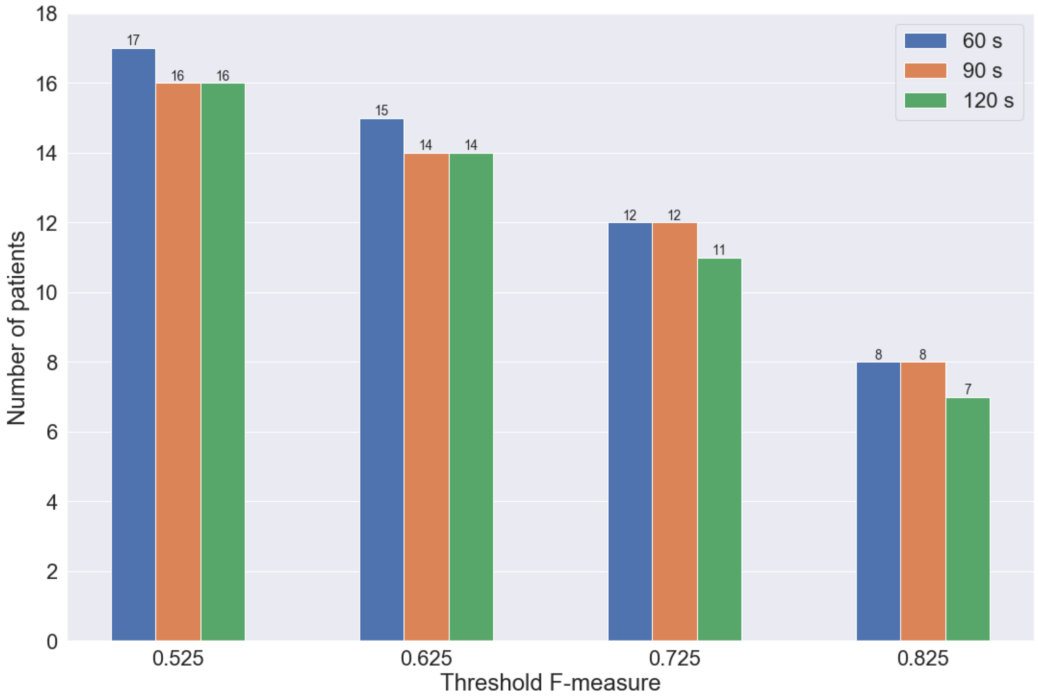


Figure 4.5. Number of patients for different accuracy target levels w.r.t different training window sizes

end points of each seizure event are annotated on a document along with EEG files which are used to indicate the ground truth.

We have marked the data points five minutes before the seizure as preictal period and approximately 8 hours of the data is non-ictal data points, i.e, data points which were recorded when the patient was not experiencing a seizure attack. We have used this to learn the patterns during the preictal and non-ictal period for each patient using the patient's data.

4.3.2 Discussion

To evaluate the performance of our models built on personalized preictal EEG data, we have chosen F-measure (the harmonic mean of the precision and recall) since it intentionally does not take into the account the number of true negatives, which, in our case, means we do not consider the number of non-ictal points correctly classified. For evaluation, we used a 4-fold cross validation technique. To see how the F-measure changed, as we learnt from different

windows, we have Figure 4.3. This figure illustrates the value of F-measure observed for a particular window size W and time-to-seizure T.

Figure 4.4 statistically summarized the F-measure values for the six window sizes, as observed in Figure 4.3 using a box and whisker plot. The plot shows the distribution of the values along with other statistics like maximums, minimums and quartile values as whiskers for each window size. Both Figures 4.3 and 4.4 are for patient 1 in our dataset. We carried out such analysis for each of the 23 patients' data. Clearly, the window size and time-to-seizure values affect the accuracy of seizure prediction. The box and whisker plot (Figure 4.4) illustrates two points. First, the size of window affects the accuracy of the prediction. Larger window sizes result in higher median values of accuracy in the box and whisker plot. Second, for a fixed window size, the time-to-seizure impacts the accuracy of the prediction. The inter-quartile range (IQR) of each box shows this effect. For example, a window size of 60 seconds (blue plot) has a difference of 17.3 percent between the highest (84.5 percent) and the lowest (67.2 percent) values.

To choose which window size that best serves the patient, we need to combine information of both Figure 4.3 and Figure 4.4. Looking at the box plot and the time-to-seizure plot, an informed decision must be made for each patient. Observing the values in the 120-seconds window (green) and the 90-second window (orange), we see that both record good F-measure values with significant time to spare. When the 120-seconds window detects the occurrence of seizure, we can alert patient/caregiver 150 seconds before seizure and the 90-seconds one similarly gives us 170 seconds. Since, the 120-seconds window gives us a higher value and the variance of those values is less than the 90-seconds one, we pick the 120-seconds window as the preferred window to predict seizures for this patient.

In order to show the impact of personalization on all the patients, we set some target level for prediction accuracy: 0.525, 0.625, 0.725, and 0.825. At minimum, we would like to alert the patient/caregiver at least a minute before the seizure occurs. If possible 120

seconds is even better as we are able to issue an alert a whole two minutes before the seizure. Therefore, we have tried values of 60 seconds, 90 seconds and 120 seconds as the thresholds for time-to-seizure values. We consider the F-measure value for each of the windows to determine the performance of the classifier (ANN) for that patient. This comparison is illustrated in Figure 4.5. It shows that for each accuracy target level, as we decrease the window size, we have a higher number of patients that satisfy that particular level threshold. This happens because the data closer to seizure could potentially have more information about the seizure.

4.4 Summary

Prediction of seizure episodes is extremely important since it can save the lives of many epileptic patients. Conventional seizure prediction techniques use a fixed set of parameters for pre-seizure data for all the patients. Since each patient responds differently to the seizures, this approach is not efficient. We proposed a personalized approach which finds the best set of pre-seizure parameters for each patient. These parameters include window size and time-to-seizure of pre-seizure data. This personalizing approach improves the accuracy of seizure prediction.

CHAPTER 5

CONCLUSION AND FUTURE DIRECTION

5.1 Summary of Findings

Monitoring seizures is an age-old struggle in the research community that seemed out of reach for decades. Persons who suffer from epileptic seizures are safer if accompanied by a caregiver when seizures strike. Consequently, there is a need for monitoring mechanisms capable of both detecting and predicting epileptic seizures in everyday life. This work presents three approaches for EEG-based analysis of epileptic seizures monitoring.

In the first algorithm, a hybrid low dimensional seizure detection methodology is developed to distinguish seizure and non-seizure samples among 23 epileptic patients. The proposed method shows a promising performance compared to the conventional techniques. The algorithm works in two steps: At first, the original full set of EEG channels are fed into a state-of-the-art channel selection model to select the most informative set of channels. Then, the spectral domain features of those selected features are transformed in low dimension. The computed information (in low dimension) is applied to a classifier model to detect seizure samples from non-seizure samples.

In the second algorithm, we advocate the use of the imbalance learning technique to make the seizure data more balanced in terms of percentage of seizure and non-seizure samples. We propose two novel ideas. First, we classify seizure class against each of non-seizure clusters. This approach increases the imbalance ratio and helps classifiers better understands the seizure pattern. Second, we customized the cost function of a neural network regarding the ratio of seizure class over the non-seizure class. The importance of seizure samples and non seizure samples are the same from the conventional neural networks perspective. In our approach, we modify the cost function of network in a way that gives a higher importance to correctly classifying seizure samples. Both proposed imbalance learning approaches improve that accuracy of seizure detection.

In most of seizure prediction scenarios, when researchers extract pre-seizure data from patients, they do not consider this fact that each patient responds differently to the occurrence of seizure. The conventional seizure prediction techniques use a fixed timing for pre-seizure data for all the patients. In the third part of this research, a personalized methodology is presented to predict seizure using a neural network model. We tested the proposed technique using two publicly available seizure datasets for 23 patients. The seizure prediction accuracy is improved when we consider personalized pre-seizure data.

5.2 Limitations

There are two limitations of our methods when utilized for EEG based epileptice seizure monitoring:

- **Insufficiency of EEG data**: Our methodology, like most of the techniques on seizure monitoring focuses on EEG data. Although there are meaningful information inside the EEG signal, many aspects of EEG-based neurological disease study are still unknown. In order to deeply understand the normal and abnormal brain activities like seizures, it is quite important to have access to more biological signals rather than EEG signals. For instance, Non-EEG biological signals like heart rate, Spo2, EDA, accelorometer and temperature are able to make significant differences and reveal new information which may not be reflected in EEG [21][7]. Use of multiple biosignals (e.g. EEG and non-EEG) can greatly decrease the false positive rate of seizure monitoring platforms.

- **Lack of long-term seizure data-sets**: Our seizure prediction technique relies on personalized seizure model. This is a limitation for seizure prediction problem. This issue limits our understanding about the mechanism leading to seizure occurrence. With increasing the number of subjects and using real world data of patients, there is a need for the new data collection and diary platforms that can save long term data to

capture all events and facilitate knowledge extraction. In fact, to have an accurate and robust monitoring both data from a large group of patient and long-term data for each patient are needed.

5.3 Future Works

Recently, several algorithms have been introduced to extract new features from EEG signals. Particularly, spectral domain features (i.e. of five unique frequency bands) which are considered as the most popular characteristics of EEG signals. So, The spectral feature that is used in this work can be expanded using other functions like Wavelet, short-FFT. Such techniques add new information to the spectral domain features by adding the time dimension to the features. This paves the way to comprehensively extract meaningful patterns from EEG signals. Moreover, these features are able to improve the accuracy of seizure prediction which requires more informative features comparing to seizure detection.

Epileptic seizure data is collected in EMUs and the cost of data collection is relatively high. However, most of the seizure data-sets suffer from a very few number of seizure events. This limited number of seizure events make the monitoring process very difficult since the monitoring models are dominated by non-seizure events. In some cases, the imbalance ration is very high such that even imbalance learning techniques cannot improve the accuracy of seizure monitoring at all. This is the main limitation of applying imbalance learning in seizure data-sets.Researchers in academia and their collaborators at medical centers need to collect data from hundreds of epileptic patients to generate a large seizure dataset with more number of seizure episodes. Although we addressed the imbalance learning approaches in this work, there are other state-of-the-art imbalance learning techniques like auto-encoder networks and GANs (Generative Adversarial Networks) models which can reproduce seizure events. These two approaches use the notion of deep learning, can better get trained on the seizure event

patterns and produce similar seizure data which ultimately improve the accuracy of seizure monitoring.

REFERENCES

[1] Acharya, U. R., S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli (2018). Deep convolutional neural network for the automated detection and diagnosis of seizure using eeg signals. *Computers in biology and medicine 100*, 270–278.

[2] Adeli, H., S. Ghosh-Dastidar, and N. Dadmehr (2007). A wavelet-chaos methodology for analysis of eegs and eeg subbands to detect seizure and epilepsy. *IEEE Transactions on Biomedical Engineering 54*(2), 205–211.

[3] Alotaiby, T., F. E. A. El-Samie, S. A. Alshebeili, and I. Ahmad (2015). A review of channel selection algorithms for eeg signal processing. *EURASIP Journal on Advances in Signal Processing* (1), 1–21.

[4] Amir, E.-a. D., K. L. Davis, M. D. Tadmor, E. F. Simonds, J. H. Levine, S. C. Bendall, D. K. Shenfeld, S. Krishnaswamy, G. P. Nolan, and D. Pe'er (2013). visne enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nature biotechnology 31*(6), 545–552.

[5] Atoufi, B., C. Lucas, and A. Zakerolhosseini (2009). A survey of multi-channel prediction of eeg signal in different eeg state normal, preseizure, and seizure. In *Proceedings of the Seventh International Conference on Computer Science and Information Technologies*.

[6] Bigus, J. P. (1996). Data mining with neural networks: solving business problems from application development to decision support.

[7] Birjandtalab, J., D. Cogan, M. B. Pouyan, and M. Nourani (2016). A non-eeg biosignals dataset for assessment and visualization of neurological status. In *2016 IEEE International Workshop on Signal Processing Systems (SiPS)*, pp. 110–114. IEEE.

[8] Birjandtalab, J., M. Heydarzadeh, and M. Nourani (2017). Automated eeg-based epileptic seizure detection using deep neural networks. In *2017 IEEE International Conference on Healthcare Informatics (ICHI)*, pp. 552–555. IEEE.

[9] Birjandtalab, J., M. James, M. Nourani, and J. Harvey (2018). Learning from non-seizure clusters for eeg analytics. In *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pp. 1–4. IEEE.

[10] Birjandtalab, J., V. N. Jarmale, M. Nourani, and J. Harvey (2018). Imbalance learning using neural networks for seizure detection. In *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pp. 1–4. IEEE.

[11] Birjandtalab, J., M. B. Pouyan, D. Cogan, M. Nourani, and J. Harvey (2017). Automated seizure detection using limited-channel eeg and non-linear dimension reduction. *Computers in biology and medicine 82*, 49–58.

[12] Birjandtalab, J., M. B. Pouyan, and M. Nourani (2016a). Nonlinear dimension reduction for eeg-based epileptic seizure detection. In *Biomedical and Health Informatics (BHI), 2016 IEEE-EMBS International Conference on*, pp. 595–598. IEEE.

[13] Birjandtalab, J., M. B. Pouyan, and M. Nourani (2016b). Unsupervised eeg analysis for automated epileptic seizure detection. In *First International Workshop on Pattern Recognition*, pp. 100110M–100110M. International Society for Optics and Photonics.

[14] Birjandtalab, J., M. B. Pouyan, and M. Nourani (2016c). An unsupervised subject identification technique using eeg signals. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 816–819. IEEE.

[15] Blanke, O., G. Lantz, M. Seeck, L. Spinelli, R. G. de Peralta, G. Thut, T. Landis, and C. Michel (2000). Temporal and spatial determination of eeg-seizure onset in the frequency domain. *Clinical Neurophysiology 111*(5), 763–772.

[16] Blum, A. (1992). *Neural networks in C++: an object-oriented framework for building connectionist systems*, Volume 1. Wiley New York.

[17] Breiman, L. (2001). Random forests. *Machine learning 45*(1), 5–32.

[18] Brodie, M., S. Barry, G. Bamagous, J. Norrie, and P. Kwan (2012). Patterns of treatment response in newly diagnosed epilepsy. *Neurology 78*(20), 1548–1554.

[19] Cheng, Y., M. T. Wong, L. van der Maaten, and E. W. Newell (2016). Categorical analysis of human t cell heterogeneity with one-dimensional soli-expression by nonlinear stochastic embedding. *The Journal of Immunology 196*(2), 924–932.

[20] Cogan, D., J. Birjandtalab, M. Nourani, J. Harvey, and V. Nagaraddi (2016). Multi-biosignal analysis for epileptic seizure monitoring. *International Journal of Neural Systems*.

[21] Cogan, D., J. Birjandtalab, M. Nourani, J. Harvey, and V. Nagaraddi (2017). Multi-biosignal analysis for epileptic seizure monitoring. *International journal of neural systems 27*(01), 1650031.

[22] Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems 2*(4), 303–314.

[23] Dahlhaus, R. (1996). On the kullback-leibler information divergence of locally stationary processes. *Stochastic Processes and their Applications 62*(1), 139–168.

[24] de la Cal, E., J. R. Villar, P. Vergara, J. Sedano, and Á. Herrero (2017). A smote extension for balancing multivariate epilepsy-related time series datasets. In *International Joint Conference SOCO17-CISIS17-ICEUTE17 León, Spain, September 6–8, 2017, Proceeding*, pp. 439–448. Springer.

[25] Duun-Henriksen, J., T. W. Kjaer, R. E. Madsen, L. S. Remvig, C. E. Thomsen, and H. B. D. Sorensen (2012). Channel selection for automatic seizure detection. *Clinical Neurophysiology 123*(1), 84–92.

[26] Ettinger, A. B. and A. M. Kanner (2007). *Psychiatric issues in epilepsy: a practical guide to diagnosis and treatment*. Lippincott Williams & Wilkins.

[27] French, J. A. and T. A. Pedley (2008). Initial management of epilepsy. *New England Journal of Medicine 359*(2), 166–176.

[28] Galar, M., A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera (2012). 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.

[29] Ghosh-Dastidar, S., H. Adeli, and N. Dadmehr (2008). Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection. *IEEE Transactions on Biomedical Engineering 55*(2), 512–518.

[30] Goldberger, A. L., L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley (2000). Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. *Circulation 101*(23), 215–220.

[31] Hiscott, R. (2015). December 9, researchers develop wearable eeg device to record seizure activity. *Neurology Today*.

[32] Kambhatla, N. and T. K. Leen (1997). Dimension reduction by local pca. *Neural Computation 9*(7), 1493–1516.

[33] Kanner, A. and S. C. Schachter (2010). *Psychiatric controversies in epilepsy*. Elsevier.

[34] Khanmohammadi, S. and C.-A. Chou (2016). A simple distance based seizure onset detection algorithm using common spatial patterns. In *International Conference on Brain and Health Informatics*, pp. 233–242. Springer.

[35] Kohavi, R. and G. H. John (1997). Wrappers for feature subset selection. *Artificial intelligence 97*(1), 273–324.

[36] Kursa, M. B. (2014). Robustness of random forest-based gene selection methods. *BMC bioinformatics 15*(1), 1.

[37] Lee, J. A. and M. Verleysen (2007). *Nonlinear dimensionality reduction*. Springer Science & Business Media.

[38] Lee, J. A., M. Verleysen, et al. (2008). Quality assessment of nonlinear dimensionality reduction based on k-ary neighborhoods. In *FSDM*, pp. 21–35.

[39] Lin, W.-C., C.-F. Tsai, Y.-H. Hu, and J.-S. Jhang (2017). Clustering-based undersampling in class-imbalanced data. *Information Sciences 409-410*, 17 – 26.

[40] Ma, J., R. P. Sheridan, A. Liaw, G. E. Dahl, and V. Svetnik (2015). Deep neural nets as a method for quantitative structure–activity relationships. *Journal of chemical information and modeling 55*(2), 263–274.

[41] Narendra, P. M. and K. Fukunaga (1977). A branch and bound algorithm for feature subset selection. *Computers, IEEE Transactions on 100*(9), 917–922.

[42] Netoff, T., Y. Park, and K. Parhi (2009). Seizure prediction using cost-sensitive support vector machine. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pp. 3322–3325. IEEE.

[43] Niedermeyer, E. and F. L. da Silva (2005). *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins.

[44] Nigam, V. P. and D. Graupe (2004). A neural-network-based detection of epilepsy. *Neurological Research 26*(1), 55–60.

[45] Osorio, I., M. G. Frei, and S. B. Wilkinson (1998). Real-time automated detection and quantitative analysis of seizures and short-term prediction of clinical onset. *Epilepsia 39*(6), 615–627.

[46] Pack, A. M. (2012). Sudep: What are the risk factors? do seizures or antiepileptic drugs contribute to an increased risk? *Epilepsy Currents 12*(4), 131–132.

[47] Powers, D. M. (2011). Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation.

[48] Qaraqe, M., M. Ismail, Q. Abbasi, and E. Serpedin (2014). Channel selection and feature enhancement for improved epileptic seizure onset detector. In *Wireless Mobile Communication and Healthcare (Mobihealth), 2014 EAI 4th International Conference on*, pp. 258–262. IEEE.

[49] R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

[50] Rahmani, M. and G. Atia (2016a). Randomized robust subspace recovery and outlier detection for high dimensional data matrices. *IEEE Transactions on Signal Processing*.

[51] Rahmani, M. and G. Atia (2016b). A subspace learning approach for high dimensional matrix decomposition with efficient column/row sampling. In *Proceedings of The 33rd International Conference on Machine Learning*, pp. 1206–1214.

[52] Rajaei, H., M. Cabrerizo, P. Janwattanapong, A. Pinzon-Ardila, S. Gonzalez-Arias, and M. Adjouadi (2016). Connectivity maps of different types of epileptogenic patterns. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, pp. 1018–1021. IEEE.

[53] Saeedi, R., R. Fallahzadeh, P. Alinia, and H. Ghasemzadeh (2016). An energy-efficient computational model for uncertainty management in dynamically changing networked wearables. In *Proceedings of the 2016 International Symposium on Low Power Electronics and Design*, pp. 46–51. ACM.

[54] Shoeb, A. H. (2009). *Application of machine learning to epileptic seizure onset detection and treatment*. Ph. D. thesis, Massachusetts Institute of Technology.

[55] Shoeb, A. H. and J. V. Guttag (2010). Application of machine learning to epileptic seizure detection. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pp. 975–982.

[56] Song, Y. and P. Liò (2010). A new approach for epileptic seizure detection: sample entropy based feature extraction and extreme learning machine. *Journal of Biomedical Science and Engineering 3*(06), 556.

[57] Studio, R. (2012). Rstudio: integrated development environment for r. *RStudio Inc, Boston, Massachusetts*.

[58] Tatum, W. O. (2014). Ellen r. grass lecture: Extraordinary eeg. *The Neurodiagnostic Journal 54*(1), 3–21.

[59] Tzallas, A. T., M. G. Tsipouras, and D. I. Fotiadis (2007). Automatic seizure detection based on time-frequency analysis and artificial neural networks. *Computational Intelligence and Neuroscience 2007*.

[60] Van der Maaten, L. and G. Hinton (2008). Visualizing data using t-sne. *Journal of Machine Learning Research 9*(2579-2605), 85.

[61] Yıldız, M. and E. Bergil (2015). The investigation of channel selection effects on epileptic analysis of eeg signals.

[62] Yuan, Q., W. Zhou, S. Li, and D. Cai (2011). Epileptic eeg classification based on extreme learning machine and nonlinear features. *Epilepsy research 96*(1-2), 29–38.

[63] Zhou, Z.-H. and X.-Y. Liu (2006a). Training cost-sensitive neural networks with methods addressing the class imbalance problem. *IEEE Transactions on Knowledge and Data Engineering 18*(1), 63–77.

[64] Zhou, Z.-H. and X.-Y. Liu (2006b, Jan). Training cost-sensitive neural networks with methods addressing the class imbalance problem. *IEEE Transactions on Knowledge and Data Engineering 18*(1), 63–77.

BIOGRAPHICAL SKETCH

Javad Birjandtalab Golkhatmi was born in Tehran, Iran. Javad received a BSc degree in Electrical Engineering - Electronics in 2010 from Sharif University of Technology, Tehran, Iran and a MSc degree in Electrical Engineering - Circuits and Systems in 2013 from the University of Tehran, Tehran, Iran. He is currently a PhD candidate in Department of Electrical and Computer Engineering at University of Texas at Dallas. He is a member of the Quality of Life Technology (QoLT) laboratory. He is interested in data analytics and signal processing for health-care applications.

CURRICULUM VITAE

Javad Birjandtalab Golkhatmi

Jan. 2020

Contact Information:

Department of Electrical Engineering
The University of Texas at Dallas
800 W. Campbell Rd.
Richardson, TX 75080-3021, U.S.A.

Email: jxb132630@utdallas.edu

Educational History:

B.S., Electrical Engineering, Sharif University of Technology, 2010
M.S., Electrical Engineering, Tehran University, 2013

Employment History:

Network Engineer Intern, Ericsson Company, May 2019 – present
Teaching Assistant, The University of Texas at Dallas, September 2015 – 2019
Research Assistant, The University of Texas at Dallas, August 2013 – August 2015

Professional Recognitions and Honors:

Merit-based Ericsson Graduate Fellowships, Engineering and Computer Science, UTD 2018
Student Travel Grant Award, ICTALS 2017
Teaching Assistant Fellowship,Engineering and Computer Science, UTD 2015-2019
Research Assistant Fellowship,Engineering and Computer Science, UTD 2013-2015
Student Travel Grant Award from National Science Foundation, PerCom 2015
Best Paper Award, International Symposium on Telecommunications IST, 2012

Professional Memberships:

American Epilepsy Society (AES) Member
Biomedical Engineering Society (BMES) Student Member
Institute of Electrical and Electronics Engineers (IEEE) Student Member
IEEE Engineering in Medicine and Biology Society (EMBS) Student Member