

B: BÜYÜKÇAKIR

IZMIR KATIP CELEBI UNIVERSITY

2020

**IZMIR KATIP CELEBI UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED
SCIENCES**

**MACHINE LEARNING AND SIGNAL PROCESSING ON
RECOGNIZING EPILEPTIC SEIZURE PATTERNS**

M.Sc. THESIS

Barkın BÜYÜKÇAKIR

Department of Electrical and Electronics Engineering

Thesis Advisor:

Prof. Dr. Adnan KAYA

FEBRUARY 2020

**IZMIR KATIP CELEBI UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED
SCIENCES**

**MACHINE LEARNING AND SIGNAL PROCESSING ON
RECOGNIZING EPILEPTIC SEIZURE PATTERNS**

M.Sc. THESIS

Barkın BÜYÜKÇAKIR

Y180207004

Department of Electrical and Electronics Engineering

Thesis Advisor: Prof. Dr. Adnan KAYA

FEBRUARY 2020

**İZMİR KATİP CELEBİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**EPİLEPTİK KRİZ ÖRÜNTÜLERİNİN TANINMASINDA
MAKİNE ÖĞRENMESİ VE SİNYAL İŞLEME**

YÜKSEK LİSANS TEZİ

Barkın BÜYÜKÇAKIR

Y180207004

Elektrik ve Elektronik Mühendisliği Ana Bilim Dalı

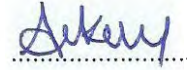
Tez Danışmanı: Prof. Dr. Adnan KAYA

ŞUBAT 2020

Barkın BÜYÜKÇAKIR., a M.Sc. student of **IKCU Graduate School Of Natural And Applied Sciences**, successfully defended the thesis entitled “**Machine Learning and Signal Processing on Recognizing Epileptic Seizure Patterns**”, which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

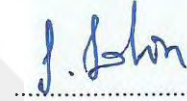
Thesis Advisor :

Prof. Dr. Adnan KAYA
İzmir Kâtip Çelebi University



Jury Members :

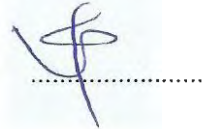
Assoc. Prof. Savaş ŞAHİN
İzmir Kâtip Çelebi University




Asst. Prof. Özgün YÜCEL
Ege University



Asst. Prof. Volkan KILIÇ
İzmir Kâtip Çelebi University



Assoc. Prof. Devrim ÜNAY
İzmir Kâtip Çelebi University



Date of Defense : 13.03.2020

FOREWORD

First and foremost, I would like to thank Dr. Ali Yener MUTLU for his guidance during his time as my advisor, and his continuous support during the entirety of my studies. I am immeasurably grateful for having the opportunity to learn from him. It is thanks to him that I chose put my passion for machine learning to work and arrived where I am now, and his encouement will always be shaping my career and life in the future.

Secondly, I would like to thank my advisor, Prof. Dr. Adnan KAYA, for the great deal of understanding that he showed and the support he provided even under unusual circumstances, which allowed me to complete my studies with utter peace of mind.

I also would like to thank Dr. Özgün YÜCEL, for showing me professional guidance and always having an open door for me when he did not have to. I thank Dr. Savaş ŞAHİN for the wisdom he shared and his hilarious tales.

My fellow researcher Furkan ELMAZ deserves a special thanks, as had we not shared a laboratory together, I would never have the my current level of progress. All the laughter we had with tired looks on our faces and passion in our hearts will always stay with me.

February 2020

Barkın BÜYÜKÇAKIR

TABLE OF CONTENTS

	<u>Page</u>
FOREWORD	v
TABLE OF CONTENTS	vi
LIST OF TABLES	viii
LIST OF FIGURES	ix
ABBREVIATIONS	xi
ABSTRACT	xii
ÖZET	xiii
1. INTRODUCTION	14
1.1 Motivation	14
1.2 Epileptic Seizure Detection.....	15
1.3 Epileptic Seizure Prediction.....	16
1.4 Organization	17
2. SEIZURE DETECTION	18
2.1 Related Work.....	18
2.2 Data Acquisition.....	19
2.3 Signal Decomposition	20
2.3.1 Empirical mode decomposition.....	20
2.3.2 Multivariate empirical mode decomposition	22
2.3.3 Hilbert vibration decomposition	23
2.4 Feature Extraction	26
2.4.1 Mean power frequency.....	26
2.4.2 Skewness	27
2.4.3 Kurtosis	27
2.4.4 Variance	28
2.4.5 Peak power frequency	28
2.4.6 Spectral Shannon entropy	29
2.4.7 Spectral Renyi entropy.....	29
2.5 Classifiers	29
2.5.1 Multilayer perceptron.....	30
2.5.2 Support vector machine.....	33
2.5.3 Random forest	34
2.5.3.1 Decision tree.....	34
2.5.3.2 Usage of Decision Tree Forests.....	36
2.5.4 K-nearest neighbors	36
2.5.5 10-fold cross-validation	38
2.6 Results	38
2.6.1 Performance metrics.....	38
2.6.2 Performance evaluation.....	39
2.6.3 Discussion	39
3. SEIZURE PREDICTION	42
3.1 Related Work.....	43
3.2 Data Acquisition.....	44
3.3 Decomposition Process	45
3.4 Feature Extraction	46

3.4.1	Sample entropy.....	46
3.4.2	Spectral power.....	47
3.4.3	Higuchi's fractal dimension.....	48
3.5	Convolutional Neural Network Architecture.....	48
3.6	Post-processing for Alerts.....	51
3.7	Results.....	53
4.	CONCLUSION.....	60
	REFERENCES.....	62
	APPENDIX.....	66
	CURRICULUM VITAE.....	75



LIST OF TABLES

	<u>Page</u>
Table 3.1 Classification sensitivity and specificity of the patients. True positives are counted as the correct prediction as preictal.	53
Table 3.2 Alarm sensitivity and false alarm rate of the patients. True positives are counted as the alarms raised within 120 minutes before a seizure.	57
Table 3.3 Performance comparison between the methods of this work and other similar studies in the literature.	59
Table A.1 MLP classification metrics of all classes.....	71
Table A.2 Linear SVM classification metrics of all classes.....	71
Table A.3 Quadratic SVM classification metrics of all classes	72
Table A.4 RBF SVM classification metrics of all classes	72
Table A.5 DT classification metrics of all classes	73
Table A.6 RF classification metrics of all classes.....	73
Table A.7 Cosine kNN classification metrics of all classes.....	74
Table A.8 Weighted kNN classification metrics of all classes	74

LIST OF FIGURES

	<u>Page</u>
Figure 2.1 The distribution of sections shown on several EEG recordings from the dataset. Differently colored sections signify different stages of an epileptic seizure. The purple rectangle depicts the aim of the classification of this section.....	19
Figure 2.2 A depiction of IMFs extracted from a 23.6 s period of an EEG signal of the dataset.....	22
Figure 2.3 The trained model of MLP classifier. The colored lines illustrate the weights. Blue stands for negative weight values, while red signifies a positive weight.....	32
Figure 3.1 The distribution of sections shown on several EEG recordings from the dataset. Differently colored sections signify different stages of an epileptic seizure. The purple rectangle depicts the aim of the classification of this section.....	43
Figure 3.2 A 1-minute section from the EEG recordings of chb01 as represented on PhysioNet.....	45
Figure 3.3 Common CNN architecture	49
Figure 3.4 The features extracted from the first HVD subcomponent of chb01. The red points signify the windows predicted as preictal by the CNN. The green lines indicate the seizure onsets. Blue lines indicate the point where an alarm is raised. (a) Alarms raised on chb01 with no silent period. It is seen to become frequent during preictal periods. (b) Alarms raised on chb01 with the silent period. The silent period can be seen to stop the alarm spam while also covering all three seizures for this patient.....	52
Figure 3.5 The features extracted from the first HVD subcomponent of (a) chb01 and (b) chb23. Green lines indicate the seizure onsets, and magenta sections indicate the signal points labeled as preictal.....	55
Figure 3.6 The classification sensitivity heatmaps from different channels of (a) <i>chb01</i> , (b) <i>chb08</i> and (c) <i>chb16</i> . All 7 HVD subcomponents have been used for this classification process.....	56
Figure 3.7 The alerts and classifier predictions for <i>chb11</i> illustrated on the features of the first subcomponent. Red points signify the preictal predictions by the classifier. The green vertical lines indicate seizure onsets and the blue vertical lines indicate alarms raised.....	58
Figure A.1 10-fold cross-validation accuracies of the classification process with MLP classifier	66
Figure A.2 10-fold cross validation accuracies of classification process with linear kernel SVM classifier.....	67

Figure A.3 10-fold cross validation accuracies of classification process with quadratic kernel SVM classifier.....	67
Figure A.4 10-fold cross validation accuracies of classification process with RBF kernel SVM classifier.....	68
Figure A.5 10-fold cross validation accuracies of the classification process with DT classifier	68
Figure A.6 10-fold cross validation accuracies of classification process with RF classifier	69
Figure A.7 10-fold cross validation accuracies of classification process Cosine distance kNN classifier	69
Figure A.8 10-fold cross validation accuracies of classification process with weighted Euclidean distance kNN classifier.....	70



ABBREVIATIONS

EEG	: Electroencephalography
MLP	: Multilayer Perceptron
kNN	: k-Nearest Neighbors
SVM	: Support Vector Machine
DT	: Decision Tree
RF	: Random Forest
CNN	: Convolutional Neural Network
HVD	: Hilbert Vibration Decomposition
EMD	: Empirical Mode Decomposition
MEMD	: Multivariate Empirical Mode Decomposition
IMF	: Intrinsic Mode Function
AED	: Anti-Epileptic Drug
MPF	: Mean Power Frequency
PPF	: Peak Power Frequency
SEN	: Shannon Entropy
REN	: Renyi Entropy
SampEn	: Sample Entropy

THESIS TITLE

ABSTRACT

Machine learning methods thrive in cases where there is a present but obscure relation between the inputs and outputs of a mechanism. Epilepsy, one of the most common brain disorders, is one such case. Leveraging the descriptive power signal processing techniques along with the predictive capabilities of machine learning algorithms is, therefore, a suitable approach to detect and predict epileptic seizures from EEG recordings. This work presents two separate but compatible frameworks in order to detect and predict seizures. The first framework consists of feature extraction and classification of EEG signals decomposed with the Hilbert vibration decomposition in order to detect seizure activity with several classifiers. Also, the performances of the HVD method and other conventional decomposition techniques are compared. The second framework builds upon the methodology of the first one by shifting the classification target to pre-seizure periods to detect preictal activity and raise alarms using a convolutional neural network and a novel post-processing algorithm. The findings of both frameworks indicate the suitability of signal decomposition and feature extraction with machine learning algorithms in the context. The first framework can reliably detect epileptic seizures with classification accuracies reaching 100%. The second framework is able to predict and alert, with a mean sensitivity of approximately 90% and false alarm rates as low as 0.02/h, therefore outperforming other frameworks proposed in the literature, while demonstrating the effectiveness of the HVD method along with machine learning algorithms.

TEZ BAŐLIĐI

ÖZET

Makine öğrenimi yöntemleri, bir mekanizmanın girdileri ve çıktıları arasında mevcut ancak belirsiz bir ilişkinin olduğu durumlarda öne çıkmaktadır. En yaygın beyin bozukluklarından biri olan epilepsi de böyle bir durumdur. Bu nedenle, makine öğrenme algoritmalarının öngörücü yetenekleri ile birlikte sinyal işleme tekniklerinin açıklayıcı gücünden yararlanmak, EEG kayıtlarından epileptik nöbetleri saptamak ve tahmin etmek için uygun bir yaklaşımdır. Bu çalışma, nöbetleri tespit etmek ve tahmin etmek için iki ayrı ancak uyumlu çerçeve sunar. İlk çerçeve, birkaç sınıflandırıcı ile nöbet aktivitesini tespit etmek için Hilbert titreşim ayrışması ile ayrıştırılan EEG sinyallerinin özellik çıkarılması ve sınıflandırılmasından oluşur. Ayrıca, HVD yönteminin performansları ve diğer geleneksel ayrışma teknikleri karşılaştırılmaktadır. İkinci çerçeve, bir evrişimli sinir ağı ve bir post-proses algoritması kullanarak, preiktal aktiviteyi tespit etmek ve alarmları yükseltmek için sınıflandırma hedefini nöbet öncesi dönemlere kaydırarak ilkinin metodolojisini geliştirmektedir. Her iki çerçevenin bulguları, bağlamda makine öğrenme algoritmaları ile sinyal ayrışmasının ve özellik çıkarmanın uygunluğunu göstermektedir. İlk çerçeve, %100'e ulaşan sınıflandırma doğrulukları ile epileptik nöbetleri güvenilir bir şekilde tespit edebilmektedir. İkinci çerçeve, yaklaşık %90'lık bir ortalama duyarlılık ve 0.02/saate kadar düşen yanlış alarm oranları ile öngörme ve uyarma yapabilir, bu nedenle, HVD yönteminin makine öğrenme algoritmaları ile birlikte etkinliğini gösterirken, literatürde önerilen diğer çerçevelerden daha yüksek performans gösterir.

1. INTRODUCTION

1.1 Motivation

Epilepsy is a brain disorder that is characterized by the occurrence of sudden abnormal brain activity, which stems from excessive activation or synchronization of neurons in a part or the entirety of the brain [1]. More than 50 million people are estimated to suffer from epilepsy worldwide, which is approximately 8 in 1000 people, and around half of this population experiences active epileptic seizures [2]. These seizures can affect the patients in ways that can range from minor behavioral changes to complete loss of consciousness and muscular control [3]. As a result, epileptic patients experience seizure-related discomfort to varying degrees in their daily lives.

Epileptic patients were shown to be at higher risk of accidents and injuries due to the seizure-related incidents alone [4], as well as having elevated rates of depression [5]. Children with epilepsy are reported to sustain more injuries compared to healthy children and, therefore, to require parental supervision [6]. Risks are also more significant with epileptic patients during pregnancy. In pregnancies where the mother has epilepsy, are more likely to result in undesired conditions in which miscarriage or post-partum complications occur [7].

It is, therefore, evident that epileptic seizures cause severe difficulties to the patients and their social circles. While anti-epileptic drug (AED) administration is frequently applied to newly diagnosed epilepsy patients, recent medical literature specifies the success rate of complete seizure control with AEDs to be around 64%, unchanged since 2000 [8, 9, 10]. Out of the remaining patient pool, those with drug-resistant epilepsy, around 25%, or around 70% of patients who undergo epilepsy surgery, achieve long-term seizure control after the procedure [11]. These ratios are promising in the sense that approximately 90% of epilepsy patients are able to enjoy seizure freedom with medical intervention. On the other hand, this information also means

that around 10% of patients that experience active epileptic seizures remain unresponsive to treatment, a population of approximately 2.5 million people.

These numbers reveal the need for alternative precautionary methods of seizure control, one of which is the widely studied subject of epileptic seizure detection. This working field focuses on the detection of the preictal and ictal stages, usually by monitoring the electroencephalography (EEG) recordings of the epileptic brain [12]. With the EEG records providing high temporal resolution usually with sampling rates between 250 to 2000 Hz, the characteristics of epilepsy can be captured within recordings, therefore making EEG the preferred method of epileptic signal recording.

With this incentive and the predictive capabilities of machine learning algorithms, the analysis on the field of epileptic seizure detection and prediction from EEG signals stands to yield a lot.

1.2 Epileptic Seizure Detection

Epileptic seizure detection via EEG signal decomposition has been widely studied in recent years. A work applied a discrete wavelet transform to decompose EEG signals and achieved an accuracy of 97% using a binary classifier. In [8], empirical mode decomposition (EMD), which performs well in extracting sub-components from nonlinear and nonstationary signals such as the EEG [9], is employed for decomposing regular, ictal and interictal EEG recordings. The extracted sub-components were then classified using a support vector machine (SVM) with a classification accuracy above 85%, where statistical features such as the mean and standard deviation are utilized. Another work proposed using the EMD for decomposing EEG signals to differentiate between ictal and seizure-free oscillations, which are then classified by the SVM [10]. This binary classification perspective yielded above 88% accuracy using different kernel functions for the SVM classifier. A work with a neural network classifier, using variance, skewness, and kurtosis as features, employed EMD, and achieved 100% accuracy for the multi-class problem of epileptic EEG classification. Another signal decomposition method, the Hilbert vibration decomposition (HVD), was also proposed for the decomposition of nonlinear and nonstationary signals [11] and has been used mainly for machine fault diagnosis [12], baseline wander removal of

electrocardiogram signals [13] and system identification [14]. HVD has been applied to EEG signals with the goal of detecting epileptic seizures using a least-squares SVM classifier, where delta, theta and alpha bands of EEG are utilized, and 97.6% classification accuracy is achieved [15]. Although both the EMD and the HVD are frequently used signal decomposition methods, there is no widely known comparison of the two methods on the subject of EEG signal decomposition. This work aims to demonstrate the success of the more recent HVD working end to end with machine learning classifiers in terms of seizure detection capabilities.

1.3 Epileptic Seizure Prediction

There have been many studies in the literature that approach the problem of epileptic seizure prediction from EEG recordings. Iasemidis & Sackellares showed that EEG recordings on the human cortex show a chaos-order-chaos pattern while transitioning from interictal to pre-ictal and to post-ictal states [13]. They were able to quantify the chaos of the system by estimating the largest Lyapunov exponents over time in the EEG recordings. Later, Mormann et al. investigated the differences in the phase synchronization of EEG signals during interictal and pre-ictal intervals [14]. Using mean phase coherence as the synchronization metric, they discovered a steady increase in mean phase coherence leading to seizure onset. Iasemidis et al. used their previous findings in order to predict seizures based on the convergence of the largest Lyapunov exponents, achieving 83% sensitivity with a false prediction rate of 0.17/h [15]. Gigola et al. applied a wavelet-based method to estimate the accumulated energy in EEG signals from epileptic patients [16]. Their framework predicted 12 out of 13 seizures. Schelter et al. used tested the statistical significance of seizure prediction techniques and have reported a mean sensitivity of 70% and a false prediction rate of 0.15/h [17]. Chisci et al. approached the problem by auto-regressive modeling EEG signals and performed classification using a support vector machine where the classification features were auto-regressive coefficients [18]. Their work reports 100% sensitivity with false alarm rates as low as 0/h. Song et al. also followed a classification path in their work, using an extreme learning machine as the classifier and sample entropy-based features, achieving 86.47% sensitivity and 83.80% specificity [19]. Parvez & Paul used the phase correlation between the current and reference EEG signals in order

to identify the preictal state [20]. This approach yielded 91.95% prediction accuracy. Yang et al. applied the classification perspective once again, using permutation entropy extracted in a sliding window from EEG recordings. Employing the support vector machine once again, they obtained an average sensitivity of 94% with a false prediction rate of 0.111/h [21]. Yuan et al. have employed the Bayesian linear discriminant analysis (BLDA) on intracranial epileptic EEG recordings using the diffusion distance metric in order to determine preictal EEG periods, and have achieved a sensitivity of 85.11% along with a false prediction rate of 0.08/h [22]. These studies clearly show that epileptic seizures are preceded by changes in the human brain that can be captured in EEG recordings and can be evaluated with specific measures.

1.4 Organization

This work is organized as follows:

- Section 2 describes the framework of seizure detection with machine learning and signal decomposition while sharing findings and providing performance comparisons of decomposition methods and machine learning classifiers.
- Section 3 describes the methodology followed in the construction and application of the epileptic seizure prediction. The section also describes the data preprocessing and the convolutional neural network architecture. The findings of the prediction framework are presented in this section.

2. SEIZURE DETECTION

In this section, a framework is proposed in order to extract features from healthy, interictal and ictal EEG signals decomposed via EMD and Hilbert vibration decomposition (HVD), and then to classify these signals with classifiers such as multilayer perceptron (MLP) [6], support vector machine (SVM)[7] and random forest (RF) [8]. The chosen classes in this section aim the differentiate the area of EEG signals marked, shown in Figure 2.1, **Error! Reference source not found.** from other sections. Then, the performances of all the decomposition method mentioned above in detecting epileptic seizures are compared on the sample dataset from Bonn University [9].

2.1 Related Work

It is generally challenging to differentiate epileptic seizures by manual observation from EEG signals due to several factors, such as the length of recordings and the ratio of ictal sections to other sections, depicted in Figure 2.1. Furthermore, Fourier based signal processing methods are unable to sufficiently analyze EEG signals as they are nonlinear and nonstationary by nature [10]. Therefore, methods such as empirical mode decomposition (EMD), or more suitably to the nature of EEG signals, the multivariate empirical mode decomposition (MEMD) are exploited when working on epileptic EEG signals to detect epileptic seizures [11,12]. Another signal decomposition method, the Hilbert vibration decomposition (HVD), was also proposed for the decomposition of nonlinear and nonstationary signals [13] and has been used mainly for machine fault diagnosis [14], baseline wander removal of electrocardiogram signals [15] and system identification [16].

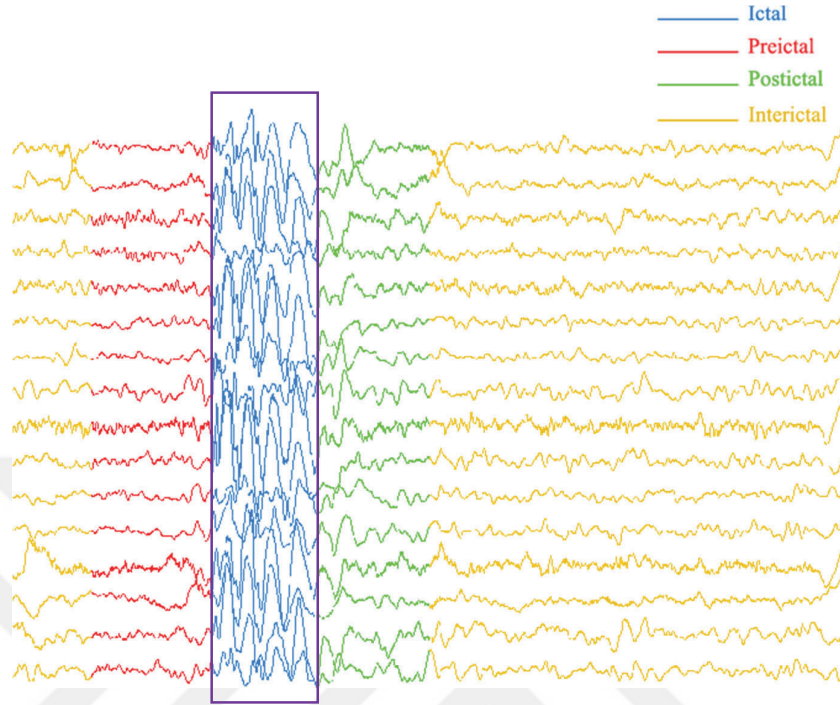


Figure 2.1 The distribution of sections shown on several EEG recordings from the dataset. Differently colored sections signify different stages of an epileptic seizure. The purple rectangle depicts the aim of the classification of this section.

2.2 Data Acquisition

In this work, the publicly available dataset provided by the University of Bonn was used [9]. The EEG signals in this dataset are presented in five sets as Z, O, N, F, and S, each consisting of 100 signals, where Z and O contain signals from healthy patients, N and F contain interictal signals from epilepsy patients, and S contains ictal signals. All data have been previously undergone visual inspection to eliminate segments contaminated by artifacts caused by muscle activity and eye movement while constructing the dataset. The chosen data were then split into 23.6 seconds segments, where all segments conform to the stationarity condition. The EEG signals were sampled with 173.61 Hz and were band-pass filtered where the passband ranges from 0.53 to 40 Hz. Within the scope of this work, only Z, F, and S data sets are used as standard, interictal, and ictal recordings. Therefore, only pre-processing for the normalization of voltage values of EEG recordings, as shown in Equation 2.1, in order to minimize computing time is performed.

$$\hat{x}(t) = \frac{x(t) - x_{min}}{x_{max} - x_{min}} \quad (2.1)$$

where $x(t)$ and $\hat{x}(t)$ are the original and normalized recordings, respectively, and x_{min} and x_{max} represent the minimum and the maximum values within $x(t)$. Using this method of normalization, all data were scaled to the fixed range of [0,1].

2.3 Signal Decomposition

One advantage of the signal decomposition – or a necessity thereof – is the ability to adaptively overcome the issues of a version of the Heisenberg uncertainty principle. Remembering the Heisenberg uncertainty principle at this point is critical. In signal decomposition, the precision that can be achieved in generating a time-frequency representation is constrained on the basis of this definition [17]. The concept is to either correctly allocate the frequency, but then the corresponding time frame is not well defined, or vice versa. The decomposition methods enable the estimation of

2.3.1 Empirical mode decomposition

Empirical mode decomposition (EMD) is a data-driven signal decomposition technique that implements a sifting algorithm in order to adaptively disintegrate the processed signal to its AM/FM modulated subcomponents [11]. These components, called intrinsic mode functions (IMFs), are the oscillation modes within the original signal and, by definition, have their number of local extrema and zero crossings differ by one at most, where their upper and lower envelopes average to approximately zero. This property of the IMFs allows for the analysis of long recordings of EEG without the need to worry about the nonlinearity and non-stationarity of the signals, as IMFs present a quasi-stationary behavior, that is, they can be assumed stationary in short periods due to the slow-changing instantaneous frequency and instantaneous amplitudes [18]. With the EMD, the target signal can be described as in Equation 2.2,

$$x(t) = \sum_{i=1}^M C_i(t) + r(t) \quad (2.2)$$

where the original signal $x(t)$ is described as the sum of the IMFs $C_i(t), i = 1, \dots, M$ and the $r(t)$ is the residual, i.e., the final component of the signal that has no more AM/FM subcomponents in it.

The algorithm of the EMD is of the iterative kind. A sifting process allows for an undetermined number of energy-wise decreasing IMFs to be extracted. The steps followed in the calculation of IMFs are given in Algorithm 2.1 **Error! Reference source not found..**

Algorithm 2.1: The EMD process	
1.	$\tilde{x}(t) = x(t)$
2.	Find all local extrema of $\tilde{x}(t)$.
3.	Interpolate through all minima and maxima of $\tilde{x}(t)$, respectively.
4.	Determine the lower envelope $e_{\min}(t)$ and upper envelope $e_{\max}(t)$.
5.	Subtract the average of two envelopes from $\tilde{x}(t)$ as: $d(t) = \tilde{x}(t) - 1/2 (e_{\min}(t) + e_{\max}(t))$
6.	If $d(t)$ satisfies the IMF conditions: <ul style="list-style-type: none"> - Let $r(t) = x(t) - d(t)$. - Go to step 7. Else: <ul style="list-style-type: none"> - Let $\tilde{x}(t) = d(t)$. - Go to step 2.
7.	If the energy of $r(t) <$ Energy threshold: <ul style="list-style-type: none"> - Stop. Else: <ul style="list-style-type: none"> - Go to step 1.

In this work, the EMD algorithm is ceased so that each signal in the dataset would yield exactly seven IMFs, as depicted in Figure 2.2. Therefore, the uncertainty of input dimensions to the classifiers (MLP, RF, SVM), is eliminated.

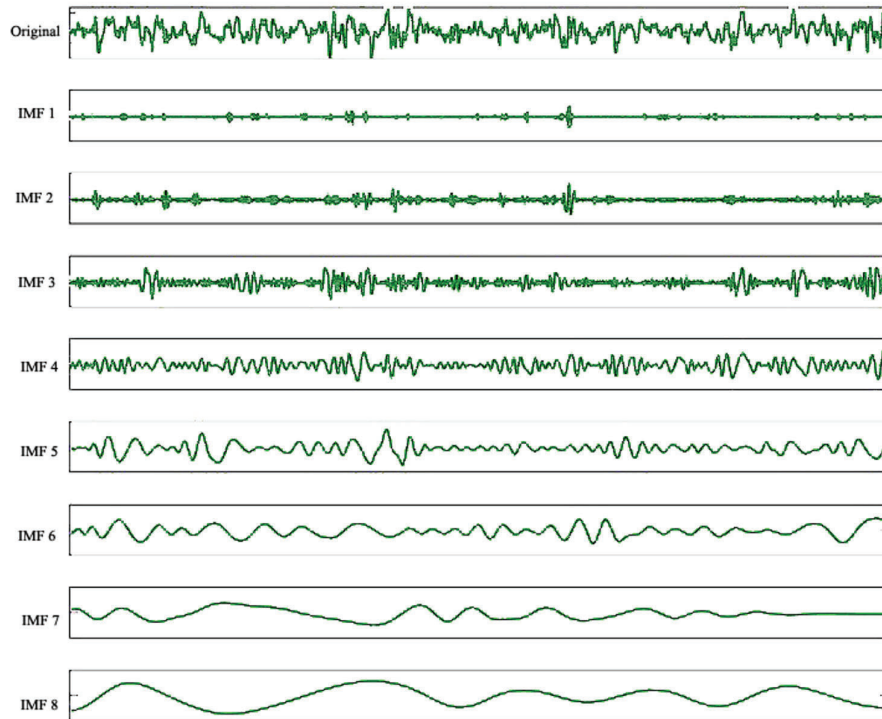


Figure 2.2 A depiction of IMFs extracted from a 23.6 s period of an EEG signal of the dataset.

The IMFs extracted from all signals from the dataset are then used for the calculation of the features with the process explained in Section 2.4 of this work.

2.3.2 Multivariate empirical mode decomposition

One of the main points of the EMD algorithm is the determination of the local mean of the initial signal, which is critically dependent on the location of the extrema. Though, this is non-trivial for the extension to multivariate signals.

The multivariate empirical mode decomposition (MEMD), a variation of the original method, approached this problem of extrema location by estimating the right set of n -dimensional vectors through the employment of low-difference points resulting from quasi-Monte Carlo methods [19]. The MEMD is, by its multi-channel suitability, naturally convenient with EEG signals, and has been previously employed on them in various studies [20,21].

In the MEMD method, the local mean of the multiple signals is found by taking multiple projections of the original signal. The direction vectors of these projections are found by the sampling of a hypersphere.

The algorithmic process of IMF calculation with the MEMD method is given in Algorithm 2.2.

Algorithm 2.2: The MEMD process

1. Create a set of points by sampling from an $n - 1$ dimensional sphere surface.
 2. For all k 's, locate the projection $p^{\theta_k}(t)_{k=1}^K$ of the original signal $\{\bar{x}(t)\}_{t=1}^T$ on the direction vector \bar{s}_{θ_n} .
 3. Find the times $\{t_i^{\theta_k}\}$ corresponding to the maximum of the projections.
 4. Interpolate through $[t_i^{\theta_k}, x(t_i^{\theta_k})]$ to find multiple envelope curves $e^{\theta_k}(t)_{k=1}^K$.
 5. For K direction vectors, calculate the average of envelopes as:

$$m(t) = \frac{1}{K} \sum_{k=1}^K e^{\theta_k}(t).$$
 6. Calculate $d(t) = x(t) - m(t)$.
 7. If $d(t)$ satisfies the IMF conditions:
 - Let $r(t) = x(t) - m(t)$
 - Go to step 7.
 - Else:
 - Let $x(t) = d(t)$.
 - Go to step 1.
 8. If the energy of $r(t) < \text{Energy threshold}$:
 - Stop.
 - Else:
 - Go to step 1.
-

As in the EMD implementation, the number of IMFs is once again set to exactly seven, therefore the classifier architectures are kept stable with a set number of inputs. The IMFs extracted from all signals from the dataset are then used for the calculation of the features with the process explained in Section 2.4 of this work.

2.3.3 Hilbert vibration decomposition

In this work, the Hilbert Vibration Decomposition method proposed by Feldmann for the extraction of the AM/FM modulated subcomponents from nonstationary signals [13] is employed for feature extraction. The method relies on the calculation of the analytic signal shown in Equation 2.3. Moreover, the frequently used method of synchronous demodulation for the estimation of the signal envelope, therefore avoiding the data-driven approach such as the one adopted by empirical mode

decomposition (EMD), which is another widely used decomposition technique for nonstationary signals. Furthermore, as shown in [22,23], the HVD method is more suitable to narrow-band signals with frequency bands packed together, such as the EEG signal with conventional bands stated as delta (<4 Hz), theta (4-8 Hz), alpha (8-15 Hz) and beta (16-31 Hz). Another desirable aspect of the HVD is the ability to change the low-pass filter cutoff frequency in order to change the frequency resolution of the extracted monocomponents. Therefore it is possible to get monocomponents that oscillate at characteristic EEG frequency bands [13].

$$\Psi(t) = s(t) + j\tilde{s}(t) = A(t)e^{j\phi(t)} \quad (2.3)$$

In Equation 2.3, $\Psi(t)$ is the analytic signal, $\phi(t)$ stands for the instantaneous phase, $d\phi(t)/d(t) = \omega(t)$ for the instantaneous frequency and $A(t)$ for the instantaneous amplitudes, and $\tilde{s}(t)$ (Equation 2.4) represents the Hilbert transform of the original signal $s(t)$

$$\tilde{s}(t) = \frac{1}{\pi} P.V. \int_{-\infty}^{\infty} \frac{s(\tau)}{t - \tau} d\tau \quad (2.4)$$

where P.V. is the Cauchy principal value to avoid discontinuity. The Hilbert transform acts as a filter that shifts the phase of $s(t)$ by $-\pi/2$. The multicomponent signals are expressed as a structure of multiple monocomponent vibrations, as can be seen in Equation 2.5. The HVD method aims to estimate $\omega_k, a_k, \forall k$ that are the instantaneous frequency and amplitude of each monocomponent.

$$s(t) = \sum_k s_k(t) = \sum_k a_k(t) \cos(\int \omega_k(t) dt) \quad (2.5)$$

It has been shown in [13] that the instantaneous amplitude $A(t)$ and frequency $\omega(t)$ of the analytic signal shown in Equation 2.6,

$$A(t)e^{j\phi(t)} = A(t)e^{j\int \omega(t)dt} \quad (2.6)$$

both consist of two parts, a slow oscillating part and a faster-oscillating part around it. If certain conditions are satisfied, the fast-changing part of $\omega(t)$ averages to zero, leaving only the largest energy part to determine the instantaneous frequency value. This central condition allows the low-pass filtering in order to average the instantaneous frequency in order to obtain the slow oscillating component. For envelope detection, the HVD method employs synchronous demodulation, which estimates the quadrature and in-phase projections of the component based on a reference frequency value, and computes the signal envelope as the sum of squares of the previously estimated projections.

The in-phase and quadrature projections of the l^{th} monocomponent signal within $s(t)$ with the reference frequency $\omega_r(t)$ are represented as in Equation 2.7 and Equation 2.8, respectively.

$$\begin{aligned} s_{l=r}(t) &= \sum [A_l(t) \cos(\int \omega_l(t)dt + \phi_l(t))] \cos(\omega_r(t)dt) \\ &= \frac{1}{2} A_l(t) [\cos(\phi_l(t)) + \cos(\int ((\omega_l(t) + \omega_r(t))dt + \phi_l(t))] \end{aligned} \quad (2.7)$$

$$\tilde{s}_{l=r}(t) = \frac{1}{2} A_l(t) [\sin(\phi_l(t)) - \sin(\int (\omega_l(t) + \omega_r(t))dt + \phi_l(t))] \quad (2.8)$$

where the instantaneous amplitude, frequency, and phase of the l^{th} monocomponent signal are $A(t)$, $\omega_l(t)$ and $\phi_l(t)$ respectively.

The fast oscillating components within the projections shown in Equations 2.7 and 2.8, which have $\omega_l \neq \omega_r(t)$, are eliminated by employing a low-pass filter with a suitable cut-off frequency. So, the phase -and therefore, frequency- and the amplitude of the monocomponent signal can be estimated. The iterative approach of the HVD method can be listed in the following steps.

Algorithm 2.3: The HVD Process

1. Estimate the mean instantaneous frequency of the largest energy vibration by making use of the analytic signal and low-pass filter in accordance with this frequency.
 2. Calculate the envelope by setting the instantaneous frequency of the reference component with the largest energy as the instantaneous frequency computed in the previous step.
 3. Subtract the largest energy component $s_k(t) = a_k(t)\cos(\int \omega_k(t)dt)$ from the signal.
-
-

In order to extract seven monocomponents, each 23.6 s portion is decomposed with the HVD method with a 4 Hz low-pass filter cutoff frequency. The decomposition is applied to all 17 EEG channels, producing $17 \times 7 = 119$ subcomponents per patient per window. All these monocomponents are used for the calculation of features described in Section 2.4 in order to construct a comprehensive dataset to evaluate the classifiers on.

2.4 Feature Extraction

In this work, the selected classifiers were tested using several different features extracted from each IMF provided by the EMD and the MEMD, and each monocomponent signal provided by the HVD. The features described below were selected to be as simple as possible with the goal of minimizing computational complexity.

All features were calculated using all of the monocomponents extracted from all of the signals in the dataset, as described in Section 2.2, meaning the following features were calculated using all subcomponent signals with the mentioned decomposition methods.

2.4.1 Mean power frequency

Mean power frequency (MPF) is defined as a weighted sum of signal power over the frequency axis and indicates which frequency component has the most substantial

dominance in the signal's power spectra [24]. Mean power frequency is described as in Equation 2.9,

$$MPF = \frac{\sum_{i=1}^N f_i p_i}{\sum_{i=1}^N p_i} \quad (2.9)$$

where f_i and p_i are the frequency value and the power of the i^{th} point in the Fourier transform of the signal, whereas M is the length of the Fourier transform.

2.4.2 Skewness

Skewness is a measure of asymmetry in a distribution [25]. The dataset is symmetric if its right side is identical to its left side of the center point. Note that the skewness of the data itself was calculated rather than the probability distribution of the time series.

$$S = \sqrt{N} \frac{\sum_{t=1}^N (X(t) - \bar{X})^3}{\left(\sum_{t=1}^N (X(t) - \bar{X})^2\right)^{\frac{3}{2}}} \quad (2.10)$$

In Equation 2.10, X_i is the i^{th} data point in the time series and \bar{X} is the average of the entire time series.

2.4.3 Kurtosis

Kurtosis is defined as the normalized form of the fourth central moment of a distribution [26]. Intuitively, it is described as the measure of the sharpness of the peak of a distribution. Kurtosis represents the "peakedness" of the EEG signal, and it may hold valuable information related to brain activity. Kurtosis is defined by the formulation shown in Equation 2.11,

$$\beta_2 = \frac{\mu_4}{\mu_2^2} \quad (2.11)$$

where μ_i represents the i^{th} central population moment of the distribution about the mean.

2.4.4 Variance

Variance is a measure of how spread out the data points are in a population, being the average of the squared distances from each point to the mean [27]. In the scope of this work, the variance is considered to be an indicator of an epileptic seizure as the EEG signals may demonstrate increased standard deviation during and in between seizures, which is simply the square root of the variance. The variance is given by Equation 2.12,

$$\sigma^2 = \frac{1}{N} \sum_{t=1}^N (X(t) - \bar{X})^2 \quad (2.12)$$

where X_i is the i^{th} data point within the time series and $\bar{X} = \mu$ is the mean of the time series.

2.4.5 Peak power frequency

Simply being the frequency of maximum power, peak power frequency (PPF), in the scope of this work, can be considered as a more specific variant of MPF, again conveying information about which frequency band of the EEG signal carries the most power. PPF is signified in Equation 2.13.

$$PPF = \operatorname{argmax}_{i:f_i \in \mathcal{F}} p_i \quad (2.13)$$

In Equation 2.13, p_i stands for the power of the signal. $\operatorname{argmax}_{i:f_i \in \mathcal{F}}$ on the other hand, means the frequency in the range of Fourier transform frequencies, that maximizes p_i .

2.4.6 Spectral Shannon entropy

Shannon entropy (SEN) is defined, on average, as the minimum number of yes / no questions in base-two, required to identify a sampled signal (Equation 2.14). SEN can be used as a measure of diversity [28]. In the context of this work, Shannon entropy was calculated by the Fourier transform of the signal.

$$SEN = - \sum_{i=1}^N p_i \log p_i \quad (2.14)$$

$p_i, i = 1, 2, \dots, N$ is the power value at the i^{th} point of the Fourier transform. In this respect, this feature has been evaluated as spectral Shannon entropy.

2.4.7 Spectral Renyi entropy

Renyi entropy (REN, Equation 2.15) is again a measure of the irregularity of an observed system or the information obtained by observing the system in question [29]. Renyi entropy is a generalized form of Shannon entropy and is reduced to Shannon entropy for $\alpha = 1$.

$$REN = \frac{1}{1 - \alpha} \log \sum_{i=1}^N p_i^\alpha \quad (2.15)$$

where $p_i, i = 1, 2, \dots, N$ is the power value at the i^{th} data point in the Fourier transform of the signal.

2.5 Classifiers

All classifiers were presented to the features expressed in Section 2.4. The features were organized as all the features, plus the subcomponent number (as in the order of extraction) of all the subcomponents. The organization of the classifier input matrix \mathbf{X} in Equation 2.16 .

$$\mathbf{X} = \begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \\ A_6 \\ A_7 \end{bmatrix} \quad (2.16)$$

where,

$$A_i = \begin{bmatrix} MPF_i \\ Skewness_i \\ Kurtosis_i \\ Variance_i \\ PPF_i \\ SEN_i \\ REN_i \end{bmatrix}, i = 1, 2, \dots, 7$$

Also, one more type of input matrix was constructed, which contains only one type of feature as shown in Equation

$$\mathbf{B} = \begin{bmatrix} F^1 \\ F^2 \\ F^3 \\ F^4 \\ F^5 \\ F^6 \\ F^7 \end{bmatrix} \quad (2.17)$$

where F^i is the feature F , which stands for one type of feature shown in Section 2.4, and i is the subcomponent number in the order of extraction. This enables for the observation of each feature and its predictive capability in the context of epileptic seizure detection.

2.5.1 Multilayer perceptron

MLP is a machine learning approach that is influenced by the human brain and its information processing system [30]. Because of its proven success in predicting both

continuous and discrete variables, it has found widespread use across various types of problems in different disciplines [31]. MLP consists of neurons, which are the primary processing elements, and neuron clusters, which are called layers. Neurons receive input from neurons in the previous layer, process information with activation feature, and send feedback to neurons in the next layer [32]. This transmitting process begins with the input layer and continues until the neurons in the output layer produce output(s).

A layer between the input and output layers is called a hidden layer, the number of hidden layers and the number of neurons on each hidden layer, and the activation functions are hyper-parameters that must be calculated beforehand. The output formula for a single hidden layer and a single output feed-forward neural network can be defined as in Equations 2.18 and 2.19,

$$\hat{y} = \sigma_2 \left(\sum_{i=1}^N \mathbf{w}_i^{(2)} \sigma_1(\mathbf{X}) + \mathbf{b}^{(2)} \right) \quad (2.18)$$

$$\mathbf{X} = \sum_{j=1}^M \mathbf{x}_j \mathbf{w}_{xj}^{(1)} + \mathbf{b}^{(1)} \quad (2.19)$$

where, \hat{y} is the prediction vector of the MLP model, N is the number of samples in the data set, M is the number of features in the data set. \mathbf{x}_j is the j^{th} feature vector, $\mathbf{w}^{(2)}$ are the weights between the hidden layer and the output layer, $\mathbf{w}^{(1)}$ are the weights of inputs connected to the hidden layer σ_2 is the activation function of the output layer, σ_1 is the activation function of the neurons in the hidden layer, $\mathbf{b}^{(2)}$ and $\mathbf{b}^{(1)}$ are the bias vectors in the output layer and hidden layer, respectively. In this work's MLP models, two architectures were employed, one for using each feature separately, and one for all of them combined. The former model structure contains 8 input nodes, who hidden layers with 64 nodes each, and one output node, while the latter preserves the same structure with the exception of having 56 input nodes. The sigmoid activation function is used in all hidden and output layers. Training of a neural network is performed by adjusting the weights between each connection, which minimizes the difference between predictions and the actual output. Widely used backpropagation

and batch gradient descent algorithms are employed for the training phase [33]. Figure 2.3 illustrates a trained multilayer perceptron model.

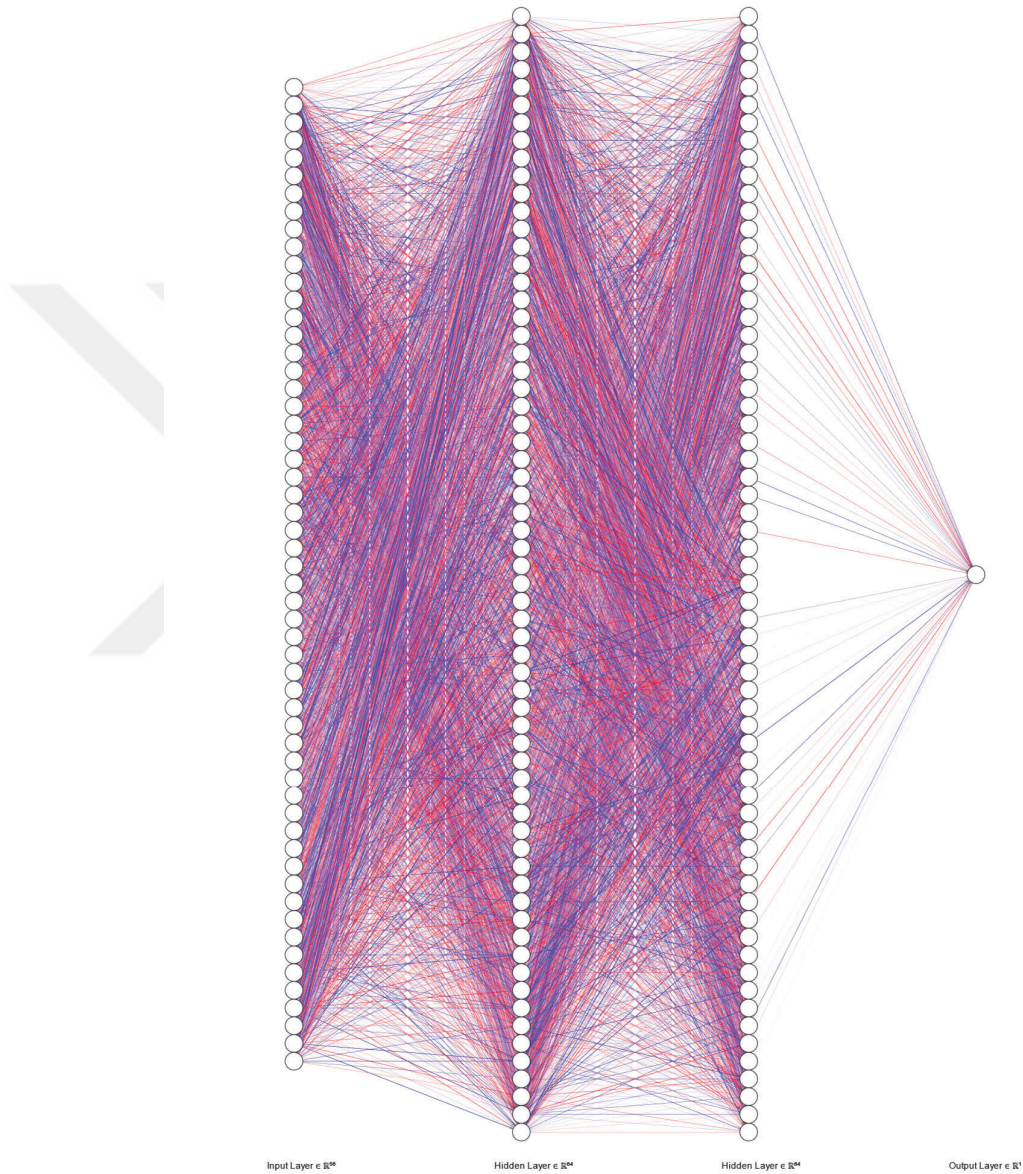


Figure 2.3 The trained model of MLP classifier. The colored lines illustrate the weights. Blue stands for negative weight values, while red signifies a positive weight.

2.5.2 Support vector machine

A support vector machine (SVM) is a supervised learning type classifier that draws an optimal hyperplane to separate two classes [34]. The SVM has been widely used in recent years in many fields.

The SVM applies classification by solving a constrained optimization problem in order to draw a hyperplane that has the largest distance to the extreme samples in each class. This constrained optimization problem is generally solved by employing the Lagrange multiplier method. This aspect of the SVM enables fast and straightforward computation of the model. However, the base method of the SVM can only draw a linear hyperplane, meaning it may not perform well on nonlinearly separable data [35]. This problem is overcome by introducing kernel functions. These functions map the input data to a higher dimensional space where the transformed data is linearly separable. In other words, instead of separating the data in its original form, the SVM can classify the function of input elements in order to perform more desirably. The SVM uses the decision function in Equation 2.20.

$$f(\mathbf{x}) = \text{sign}[\boldsymbol{\lambda}^T g(\mathbf{x}) + b] \quad (2.20)$$

where $g(\mathbf{x})$ is the kernel function that maps an N -dimensional input data \mathbf{x} to a higher-dimensional space, $\boldsymbol{\lambda}$ is an N -dimensional vector of weights, and b is the bias term. $\boldsymbol{\lambda}$ and b are computed by solving the constrained optimization problem

$$\min_{\{\boldsymbol{\lambda}, b, \mathbf{e}\}} J(\boldsymbol{\lambda}, b, \mathbf{e}) = \frac{\boldsymbol{\lambda}^T \boldsymbol{\lambda}}{2} + \frac{\gamma}{2} \sum_{i=1}^N |e_i|^2 \quad (2.21)$$

Subject to constraints,

$$y_i [\boldsymbol{\lambda}^T g(\mathbf{x}) + b] = 1 - e_i, \quad i = 1, 2, \dots, M \quad (2.22)$$

where $\{x_i, y_i\}_{i=1}^M$ are the M training pairs, $y_i = \pm 1$ stands for the class that x_i belongs to and $e = [e_1, e_2, \dots, e_M]$. Implementing the Lagrange multiplier method, the objective function becomes as in Equation 2.23.

$$\mathcal{L}(\lambda, b, e; \alpha) = J(\lambda, b, e) \sum_{i=1}^M \alpha_i [y_i [\lambda^T g(x)_i + b] - 1 + e_i] \quad (2.23)$$

The decision function then becomes,

$$f(x) = \text{sign} \left[\sum_{i=1}^M \alpha_i y_i \kappa(x, x_i) + b \right] \quad (2.24)$$

where $\kappa(x, x_i)$ is the more common representation of a kernel function. The kernel functions that are implemented in this work are the linear function kernel (Equation 2.25), quadratic function kernel (Equation 2.26), Gaussian radial basis function kernel (Equation 2.27).

$$\kappa(u, v) = (\bar{u} \cdot \bar{v} + 1) \quad (2.25)$$

$$\kappa(u, v) = \beta_0 + \beta_1 u^T u + \beta_2 v^T v + \beta_3 u \cdot v \quad (2.26)$$

$$k(u, v) = e^{-\left(\frac{\|u-v\|^2}{2\sigma^2}\right)} \quad (2.27)$$

2.5.3 Random forest

In order to clearly understand the random forest (RF) classification, it is essential first to clarify its weak learner, the decision tree.

2.5.3.1 Decision tree

Decision tree (DT) classifiers are trees where each node represents another feature of the data, and each branch represents a decision rule [36]. The DT classifier is widely

used because of its straightforward approach to classification and regression and also the low computational demand even in higher dimensional data, from which a single feature is tested at a time. In this work, the Classification and Regression Trees (CART) algorithm for DT model construction is adopted, where the Gini Index is used as a metric [37].

For any subset S_q of training dataset S , the Gini index is given by Equation 2.28.

$$G(S_q) = 1 - \sum_{k=1}^K (p_{k,q})^2 \quad (2.28)$$

where k is the class label for the data $p_{k,q}$ is the conditional probability of class k at node q . From Equation 2.28, it can be seen that the Gini index reaches the minimum 0 when all observations fall in a single class, and the maximum is achieved when observations are equally distributed among classes. Furthermore, the weighted Gini index, or the Gini index of a subset S_q resulting from the partition of data A_L^i can be defined as

$$wG(S_q, A_L^i) = p_{L,q}(A_L^i)G(L_q(A_L^i)) + (1 - p_{L,q}(A_L^i))G(R_q(A_L^i)) \quad (2.29)$$

where Gini indices of sets $L_q(A_L^i)$ and $R_q(A_L^i)$ are calculated as in Equation 2.28. The splitting function is then defined as the difference between the Gini index and the weighted Gini index, and this function is used analogously to information gain (Equation 2.30).

$$g(S_q, A_L^i) = G(S_q) - wG(S_q, A_L^i) \quad (2.30)$$

The maximization of this function, called Gini gain, provides optimal decision boundaries. It is therefore desirable to choose the partition $\bar{A}_{L,q}$ as in Equation 2.31.

$$\bar{A}_{L,q} = \max_{A_L^i \in V_i} \{g(S_q, A_L^i)\} \quad (2.31)$$

2.5.3.2 Usage of Decision Tree Forests

The random forest (RF) classifier is a combination of decision tree classifiers where each classifier is generated using a random vector separately from the input vector, and each tree casts a vote for the most popular class to classify an input vector [38]. In this work, the random forest is generated for the $N \times M$ -dimensional training data \mathbf{x} , generating a bootstrap sample of the training data, $\bar{\mathbf{x}}$, by choosing N random observations and $k \leq M$ features from the original data for each DT or learner in the ensemble. Each learner is then built via using the Gini index of the bootstrap sample $\bar{\mathbf{x}}$.

The significant advantage of RF classification is that the RF classifier, as opposed to the DT, is not prone to overfitting [8]. The RF classifier can reduce the error due to variance and therefore provide higher classification performance on test data or previously unobserved data.

2.5.4 K-nearest neighbors

The k Nearest Neighbours (kNN) is a non-parametric classification algorithm that, for a given test data, calculates N distances between an N -dimensional input and the test data point, sorts these points and, with a majority vote or a weighted vote between the class labels of k nearest training data points, classifies the given data. The two user-defined parameters that primarily dictate the performance of the kNN classifier is the distance metric calculation and the parameter k , the number of neighbors to evaluate.

The kNN algorithm is advantageous in the sense that it is a simple algorithm that is relatively straight-forward, and that it does not make any assumptions about the underlying distribution of the data. However, it is not without problems. The kNN algorithm is computationally expensive and with too many data points, might be slow. Since it is also generalization-free, it stores the training data, resulting in high memory requirements.

The choice of k , the number for neighbors also greatly affect the algorithm performance. Choosing a too small value of k may make the classification process more susceptible to the effects of noise, and a large k makes the smoothing effects more dominant, causing underfitting.

There are also many choices available for distance metrics. Examples include the Euclidean distance, the Manhattan distance, and the Cosine distance, provided in Equations 2.32, 2.33, 2.34, respectively.

$$D^2 = \left(\sum_{i=1}^N (p_i - q_i)^2 \right)^{\frac{1}{2}} \quad (2.32)$$

$$D = \sum_{i=1}^N |p_i - q_i| \quad (2.33)$$

$$C = 1 - \frac{\sum_{i=1}^N p_i q_i}{\left(\sum_{i=1}^N |p_i|^2 \right)^{\frac{1}{2}} \left(\sum_{i=1}^N |q_i|^2 \right)^{\frac{1}{2}}} \quad (2.34)$$

There is also a weighted approach that can be followed, in which the k nearest neighbors contribute on the final decision in proportion to their distances from the test data point as

$$\text{Final Decision} = \frac{\sum_{i=1}^k (l_i d_i)}{\sum_{i=1}^k d_i} \quad (2.35)$$

where l_i and d_i are the class label and the distance from the test point to the i^{th} neighbor point. In this work, the weighted Euclidean distance and the cosine distance metrics in the kNN classifier were evaluated.

2.5.5 10-fold cross-validation

The classifier models described were tested by a 10-fold cross-validation method. The main purpose of this method is to get a preview of the overall performance of the trained model over a limited number of data. The results obtained with the 10-fold diameter verification method are less optimistic, but more realistic, than another popular method of training and test set separation [39]. 10-fold cross-validation steps are implemented as follows.

- The data set is divided into 10 groups.
- One of the groups is kept separate as a test set.
- The model is trained over the remaining 9 groups and an estimate is made on the test set.
- Estimates are stored and the model is reset.
- The process is repeated until all 10 groups are used as the test set.
- Using all predictions and all real classes, the performance of the model is measured.

The 10-fold cross-validation results obtained by classifying the features extracted with MEMD and HVD are given in Section 2.6.

2.6 Results

2.6.1 Performance metrics

During the performance evaluations, classification accuracy (A), Precision, Recall, and F1-Score, metrics are employed, which are given in Equations 2.36, 2.37, 2.38, 2.39, respectively.

$$A = \frac{TP + TN}{TP + TN + FP + FN} = \frac{\# \text{ of Correct Predictions}}{\# \text{ of Total Predictions}} \quad (2.36)$$

$$Precision = \frac{TP}{TP + FP} \quad (2.37)$$

$$Recall = \frac{TP}{TP + FN} \quad (2.38)$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.39)$$

where a TP (true positive) corresponds to the correct identification of a test sample from a particular class, Fn (false negative) means that the prediction is not the class that the test sample belongs to, which is the target class. TN (true negative) happens when predictions based on samples from classes that are not the target class, and they are classified as a non-target class. FP (false positive) happens when the sample is predicted wrongly as the target class [40].

2.6.2 Performance evaluation

The features kurtosis, skewness, variance, PPF, MPF, Shannon entropy, and Renyi entropy, extracted from the subcomponents of the EEG signals decomposed by the EMD and the HVD were fed to classifiers using 10-fold cross-validation method. In addition to these features, the classification process using binary classifiers were performed in a one-versus-one fashion. The evaluated classifiers were the MLP, the SVM with linear, quadratic and RBF kernels, the decision tree, the random forest, kNN with cosine distance, and weighted kNN with Euclidean distance analysis and the convolutional neural network. The classification accuracies, precision, recall, and F1 scores and the confusion matrices are shared in Appendix A.

2.6.3 Discussion

As can be observed from Figure A.1 through Figure A.8, the classification accuracy of the 10-fold cross-validation is higher for those features that are extracted from the subcomponents decomposed by the HVD (hereby referred to as the HVD features) compared to those extracted from the subcomponents provided by the EMD and the MEMD (hereby referred to as the EMD features and MEMD features).

For the 10-fold cross-validation, the HVD features provided higher accuracy compared to the EMD and MEMD features with the exceptions of the variance feature with the MLP, quadratic kernel SVM and random forest, where the EMD scored higher, and the REN feature with the DT, where the MEMD showed better performance. The

decision tree and the random forest achieved almost the same accuracy for the EMD, MEMD, and HVD, in all cases. In 10-fold cross-validation, the highest accuracies are achieved for the EMD features when the MLP was used, followed closely by the quadratic kernel SVM and the weighted kNN classifiers. The random forest classifier, being an ensemble of decision trees, provided higher classification accuracies compared to its weak learner, the decision tree. This was also the case for the HVD features, where the random forest classifier was able to add to the accuracy of the decision tree classifier. There was no single best-performing EMD feature across all classifiers in the 10-fold cross-validation results. The variance and REN feature provided the highest accuracies with no substantial difference with each classifier. The features variance and MPF consistently yielded less accuracy in MEMD than in EMD. The cases where all features were used together for classification also provided satisfying results across the board. It is also worth noting that the spectral SEN yielded lower accuracy than its generalized counterpart, the spectral REN, with all classifiers, the only exception being the weighted kNN classifier. The worst performing EMD feature was the skewness feature for all classifiers. This is also evident from the F1 scores of skewness for EMD and MEMD, as they are the lowest among the F1 scores of all features.

10-fold cross-validation using the HVD features provided satisfying results across the board. The usage of the HVD features enhanced the classification performance notably. Within the HVD features, there were several instances that of perfect classification accuracy, most frequently with the MPF and PPF features. These two frequency-related features achieved the perfect score invariably with all classifiers. This is most likely due to the intrinsic relationship of the HVD method's inner workings. The HVD is known to be more suitable to decompose narrow-band signals with frequencies are closer together [22]. This, along with the fact that the EEG is a narrow-band signal with its characteristic frequency bands lying in the 0-40 Hz range, makes the matchup of EEG signals and the HVD highly preferable. This natural suitability is even further indicated by the high classification accuracies and closer-to-ideal scores provided by all evaluated classifiers using the HVD features in the subject of epileptic seizure classification in comparison to the EMD and MEMD features. Not only the frequency-based features MPF and PPF, but also the spectral Renyi entropy

and variance features, and the combination of all features also provided accuracy greater than 90% and F1 metrics greater than 0.9 for all classes with each classifier. Aside from these features, the best classification performance was observed with the MLP classifier, closely followed by the SVM variants and the random forest. The worst-performing feature skewness yielded above 85% accuracy and above 0.8 F1 scores with this classifier.



3. SEIZURE PREDICTION

This section focuses on the detection of the preictal stage by monitoring the electroencephalography (EEG) recordings of the epileptic brain [41]. With the EEG records providing high temporal resolution usually with sampling rates between 250 to 2000 Hz, the characteristics of epilepsy can be captured within recordings, therefore making EEG the preferred method of epileptic signal recording.

In this section, a system that uses the Hilbert Vibration Decomposition (HVD) for signal decomposition, -which was shown to perform more reliably on seizure detection in Section 2- in order to extract features from epileptic surface EEG signals, and utilizes these features for a binary classification between preictal and non-preictal periods and provides alarms based on the predictions, is introduced. Another goal of this work is to evaluate the performance of traditional classification in the context of seizure prediction, and discuss whether a classification-only approach is suitable, or a postprocessing step such as alarm creation is beneficial in improving the overall performance of the system. The target class for the classification process is illustrated in Figure 3.1.

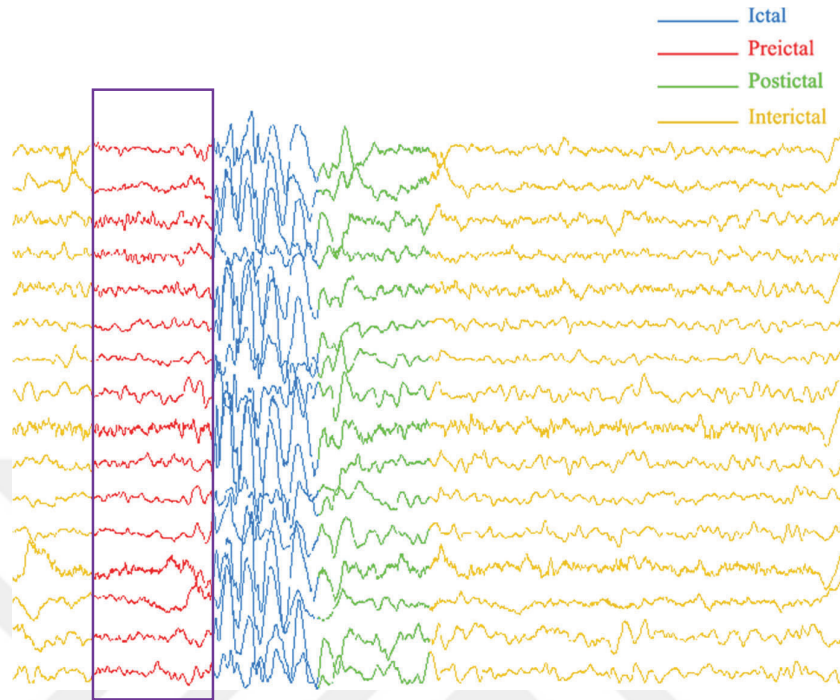


Figure 3.1 The distribution of sections shown on several EEG recordings from the dataset. Differently colored sections signify different stages of an epileptic seizure. The purple rectangle depicts the aim of the classification of this section.

3.1 Related Work

There have been many studies in the literature that approach the problem of epileptic seizure prediction from EEG recordings. Iasemidis & Sackellares showed that EEG recordings on the human cortex show a chaos-order-chaos pattern while transitioning from interictal to pre-ictal and to post-ictal states [42]. They were able to quantify the chaos of the system by estimating the largest Lyapunov exponents over time in the EEG recordings. Later, Mormann et al. investigated the differences in the phase synchronization of EEG signals during interictal and pre-ictal intervals [43]. Using mean phase coherence as the synchronization metric, they discovered a steady increase in mean phase coherence leading to seizure onset. Iasemidis et al. used their previous findings in order to predict seizures based on the convergence of the largest Lyapunov exponents, achieving 83% sensitivity with a false prediction rate of 0.17/h [44]. Gigola et al. applied a wavelet-based method to estimate the accumulated energy in EEG signals from epileptic patients [45]. Their framework predicted 12 out of 13 seizures. Schelter et al. used tested the statistical significance of seizure prediction techniques

and have reported a mean sensitivity of 70% and a false prediction rate of 0.15/h [46]. Chisci et al. approached the problem by auto-regressive modeling EEG signals and performed classification using a support vector machine where the classification features were auto-regressive coefficients [47]. Their work reports 100% sensitivity with false alarm rates as low as 0/h. Song et al. also followed a classification path in their work, using an extreme learning machine as the classifier and sample entropy-based features, achieving 86.47% sensitivity and 83.80% specificity [48]. Parvez & Paul used the phase correlation between the current and reference EEG signals in order to identify the preictal state [49]. This approach yielded 91.95% prediction accuracy. Yang et al. applied the classification perspective once again, using permutation entropy extracted in a sliding window from EEG recordings. Employing the support vector machine once again, they obtained an average sensitivity of 94% with a false prediction rate of 0.111/h [50]. Yuan et al. have employed the Bayesian linear discriminant analysis (BLDA) on intracranial epileptic EEG recordings using the diffusion distance metric in order to determine preictal EEG periods, and have achieved a sensitivity of 85.11% along with a false prediction rate of 0.08/h [51]. These studies clearly show that epileptic seizures are preceded by changes in the human brain that can be captured in EEG recordings and can be evaluated with certain measures.

3.2 Data Acquisition

The data used in this work is the surface EEG recording from pediatric patients collected in Children's Hospital Boston [52], provided publicly on PhysioNet [53]. Within this dataset, there are continuous and non-continuous recordings from 24 cases, where each case contains 9 to 42 multichannel EEG recordings with a total of 664 record files. These recordings are made in a 16-bit resolution at 256 Hz sampling frequency. While most of the recordings contain 23 channels, there are also cases with electrocardiography (ECG), or vagal nerve stimulus signals included. Figure 2 shows an example section from the recordings from the case named as *chb01*.

In this work, 10 cases from the dataset are used. These cases are randomly selected and can be listed with their labels in the original dataset as; *chb01*, *chb02*, *chb07*, *chb08*, *chb11*, *chb13*, *chb16*, *chb20*, *chb21*, and *chb23*. Each signal is band-stop

filtered with a stopband of 57-63 Hz and 117-123 Hz in order to eliminate line oscillations from the recordings. The channels used in this work are identical to the ones used in [54]. The labeling process is defined as assigning preictal labels to a 30 minute period before all seizures. All data outside of this period is labeled as non-preictal, creating a binary classification problem.

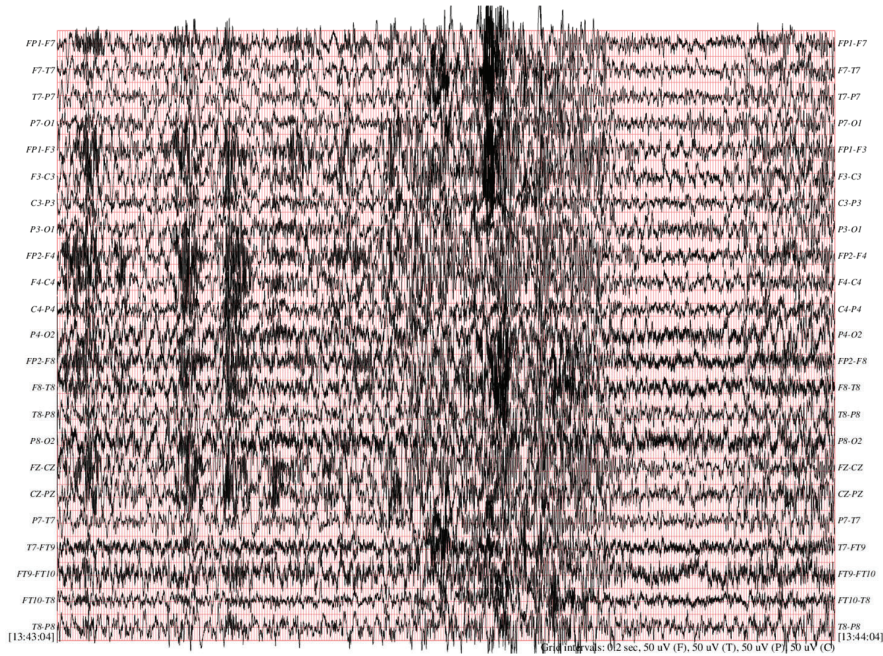


Figure 3.2 A 1-minute section from the EEG recordings of chb01 as represented on PhysioNet.

3.3 Decomposition Process

The HVD process in this section is applied in a window-based fashion. The sliding windows over the signals each capture a 16 s portion of the signal, with a 50% overlap is set. This portion is then decomposed with the HVD method with a 4 Hz low-pass filter cutoff frequency in order to extract 7 monocomponents. The decomposition is applied to all 17 EEG channels, producing $17 \times 7 = 119$ subcomponents per patient per window. All these monocomponents are used for the calculation of features described in Section 2.3 in order to construct a comprehensive dataset to evaluate the classifier on.

3.4 Feature Extraction

In order to construct process-ready data from the EEG recordings, a list of features is calculated using the monocomponent signals extracted from the recordings using the HVD method. In addition to the previously described features, skewness (Section 2.4.2), kurtosis (Section 2.4.3), and variance (Section 2.4.4), several new features are included in the framework to be processed. These new features are explained in the remainder of this section.

The features explained in this section are calculated on each subcomponent extracted in each window of EEG signals. These features are used in order to train the neural network classifier with the aim of distinguishing preictal EEG segments from non-preictal ones, in a binary classification sense.

3.4.1 Sample entropy

Sample entropy (SampEn) is a modification of approximate entropy (ApEn), proposed for the analysis of physiological time series [55]. As an entropy variant, SampEn is a measure of complexity and self-similarity. SampEn is defined as the negative natural logarithm of the conditional probability that two signal segments that are similar for m points are also similar at the following data point. In SampEn, exact matches within the signal are disregarded, removing bias within ApEn, where the measure changes under different conditions [56]. After SampEn was first proposed, there have been approaches in order to reduce the algorithmic complexity and the computation time of the SampEn value. The steps taken to calculate SampEn with $O(N^2)$ complexity, as shown in [48], are depicted in Algorithm 3.1.

Algorithm 3.1: SampEn calculation

```
Result: SampEn
for  $i \leftarrow 2$  to  $N - m$  do
     $B[i] \leftarrow 0$ 
for  $i \leftarrow 2$  to  $N - m - 1$  do
     $A[i] \leftarrow 0$ 
for  $i \leftarrow 2$  to  $N - m - 1$  do
    for  $j \leftarrow i + 1$  to  $N - m - 1$  do
        if  $\max_{k=1:m-1} (|x(i+k) - x(j+k)|) \leq r$  then
             $B[i] \leftarrow B[i] + 1$ 
             $B[j] \leftarrow B[j] + 1$ 
            if  $|x(i+m) - x(j+m)| \leq r$  then
                 $A[i] \leftarrow A[i] + 1$ 
                 $A[j] \leftarrow A[j] + 1$ 
     $i = N - m$ 
for  $j \leftarrow i + 1$  to  $N - m - 1$  do
    if  $\max_{k=1:m-1} (|x(i+k) - x(j+k)|) \leq r$  then
         $B[i] \leftarrow B[i] + 1$ 
         $B[j] \leftarrow B[j] + 1$ 
 $B^m \leftarrow 0$ 
for  $i \leftarrow 1$  to  $N - m$  do
     $B^m \leftarrow B^m + (B[i]/(N - m - 1))$ 
 $B^m \leftarrow B^m/(N - m)$ 
 $m \leftarrow m + 1$ 
 $A^m \leftarrow 0$ 
for  $i \leftarrow 1$  to  $N - m$  do
     $A^m \leftarrow A^m + (A[i]/(N - m - 1))$ 
 $A^m \leftarrow A^m/(N - m)$ 
 $SampEn \leftarrow \ln(A^m/B^m)$ 
```

3.4.2 Spectral power

The spectral power of EEG signals indicates the sum of the power spectral density (PSD) of the signals. The PSD is the power values of each frequency within the signal, and as such, it carries information regarding the status of the EEG signals. In the context of epilepsy, the spectral power of the characteristic EEG bands is usually investigated [57,58]. In this work, since the HVD method is used to decompose the EEG signals, merely the power of each subcomponent is calculated. The spectral power of each subcomponent is calculated as the logarithmic sum of the PSD.

$$p(n, m) = \log_{10} \sum PSD(x^{(n,m)}) \quad (3.1)$$

In Equation 3.1, $\chi^{(n,m)}$ stand for the n^{th} window of the m^{th} subcomponent. PSD is the power spectral density operator that estimates the PSD using Welch's method [59].

3.4.3 Higuchi's fractal dimension

The fractal dimension of a signal indicates the amount of self-similarity of a signal in the time domain, and by extension, the complexity of the signal [60]. This measure is between 1 and 2, where a dimension of 1 indicates a straight line and 2 indicates a 2-dimensional shape, such as a square. Higuchi's fractal dimension measure is a variant of the fractal dimension metric [61]. It has been shown to be more suitable for the analysis of physiological signals due to their nonlinear and nonstationary nature, compared to other fractal dimension variants [62]. In particular, analysis of signals capturing neuronal activity has been shown to benefit from the use of Higuchi's fractal dimension [63]. Higuchi's fractal dimension was calculated for all windows of all subcomponents extracted from the EEG signals via the HVD with the algorithm described in [62], with the maximal scale $k_{max} = 8$.

3.5 Convolutional Neural Network Architecture

Convolutional neural networks (CNNs) have been first proposed for, and ever since have been widely used within the area of image classification [64,65]. The convolution and pooling layers unique to the CNN provide the neural network architecture with the ability to extract their own features, which are then simply fed to fully connected neural network classifier. This particular ability has been inspired by the human visual system. In contrast to the 2-dimensional CNNs architecture used primarily for image classification, this work utilizes the 1-dimensional version of the CNNs architecture.

In CNNs, usually, three types of layers are present: convolution layers, pooling layers and fully connected (dense) layers. In convolution layers, many kernel matrices with the same number of dimensions slide over the input matrix with a determined step size, which results in convolution. These kernels are often randomly generated and layer adjusted with backpropagation methods. Although the convolution operation reduces the length of the input matrix, pooling layers reduce this length even further. Pooling layers scan over their inputs with certain window sizes and output the greatest value

in every step. This way the so-called most significant value is used and data sizes are further reduced. After a series of convolution and pooling layers, the traditional fully connected neural network classifier decides to which class the input belongs [66]. Fig. 1 illustrates the commonly used CNN architecture.

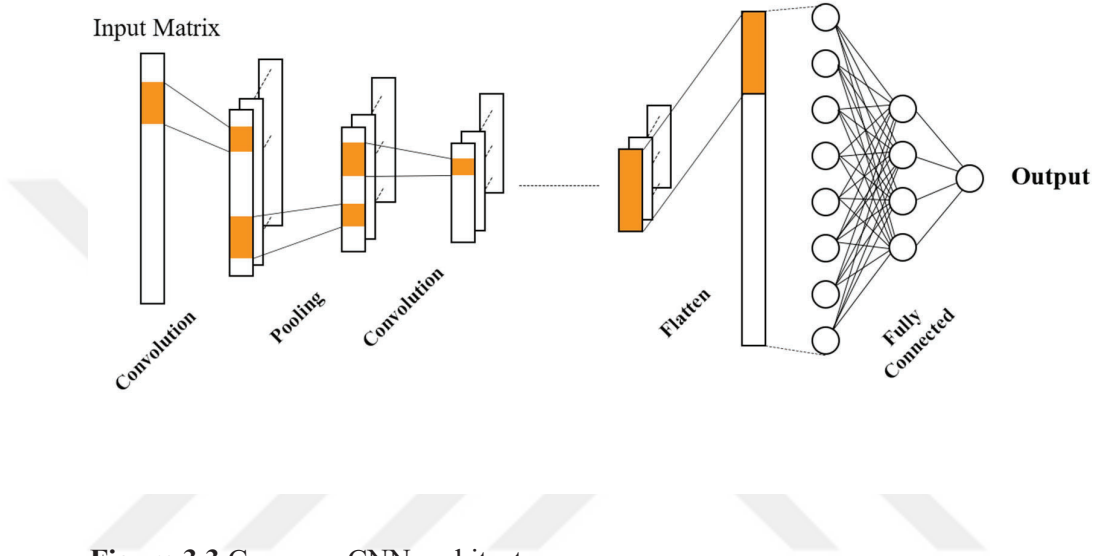


Figure 3.3 Common CNN architecture

In this work, the CNN is designed to contain five convolution layers, five pooling layers, and two fully connected layers. The inputs to the CNN classifier are provided as

$$\begin{bmatrix} \text{SubcompNo.} & \text{SubcompNo.} & \dots & \text{SubcompNo.} \\ \text{Variance}_1 & \text{Variance}_2 & \dots & \text{Variance}_N \\ \text{Skewness}_1 & \text{Skewness}_2 & \dots & \text{Skewness}_N \\ \text{Kurtosis}_1 & \text{Kurtosis}_2 & \dots & \text{Kurtosis}_N \\ \text{SampEn}_1 & \text{SampEn}_2 & \dots & \text{SampEn}_N \\ \text{SpectralPower}_1 & \text{SpectralPower}_2 & \dots & \text{SpectralPower}_N \\ \text{Higuchi}_1 & \text{Higuchi}_2 & \dots & \text{Higuchi}_N \end{bmatrix} \quad (3.2)$$

where N is the number of sub-signals, i.e., monocomponents. So, for the HVD, the input to the CNN consists of a 7×7 matrix since N is equal to 7.

Due to the large class imbalance problem between preictal and non-preictal classes, class weights are introduced into the classification process. The class weights are

chosen to balance the number of samples in each class from the perspective of the classifier. Weights are calculated as in Equation 3.3.

$$w_i = \frac{m_{i'}}{m_i + m_{i'}} \quad (3.3)$$

In Equation 3.3, w_i represents the class weight of class i . m_i stands for the number of samples belonging to class i , and $m_{i'}$ stands for the number of samples from outside of the class i . By using such weights, the CNN classifier is made to perceive the number of observations from both classes as equal.

The input matrices were fed to the CNN in batches of 20, and the weights, biases and kernel coefficients were adjusted over 50 epochs using backpropagation. At every layer except for the output layer, Rectified Linear Unit (ReLU) activation function is used. The output layer was assigned a Softmax activation function. The loss function, the measure of error between output labels and predictions, were chosen to be categorical cross-entropy, which is defined as in Equation 3.4.

$$L(\theta) = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C \mathbf{1}_{y_i \in C_c} \log p_{model}[y_i \in C_c] \quad (3.4)$$

Here, i stands for the number of training example and c for categories. The term $\mathbf{1}_{y_i \in C_c}$ is the indicator function of the i^{th} observation belonging to the c^{th} category. This model structure was tested on the dataset by the means of 10-fold cross validation. 10-fold cross validation has the primary purpose of giving a preview of sorts on how the model performs in general, extrapolating from a limited dataset [39].

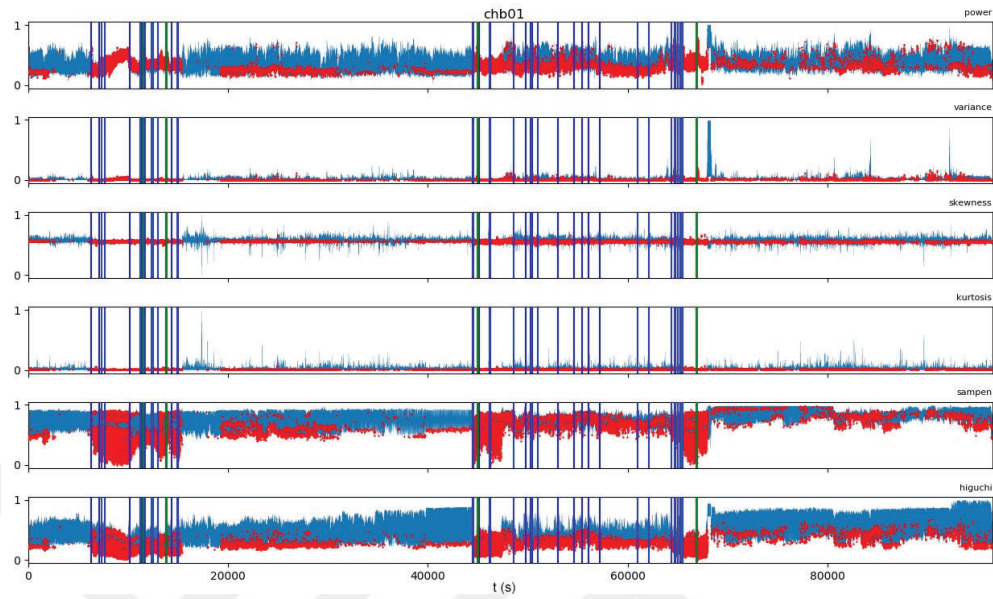
The classifier is cross validated on the data from all the patients, and the features are independent of the patient identity. Therefore it is ensured there is no way the classifier to learn which data belongs to which patient, thus making the process non-patient specific.

3.6 Post-processing for Alerts

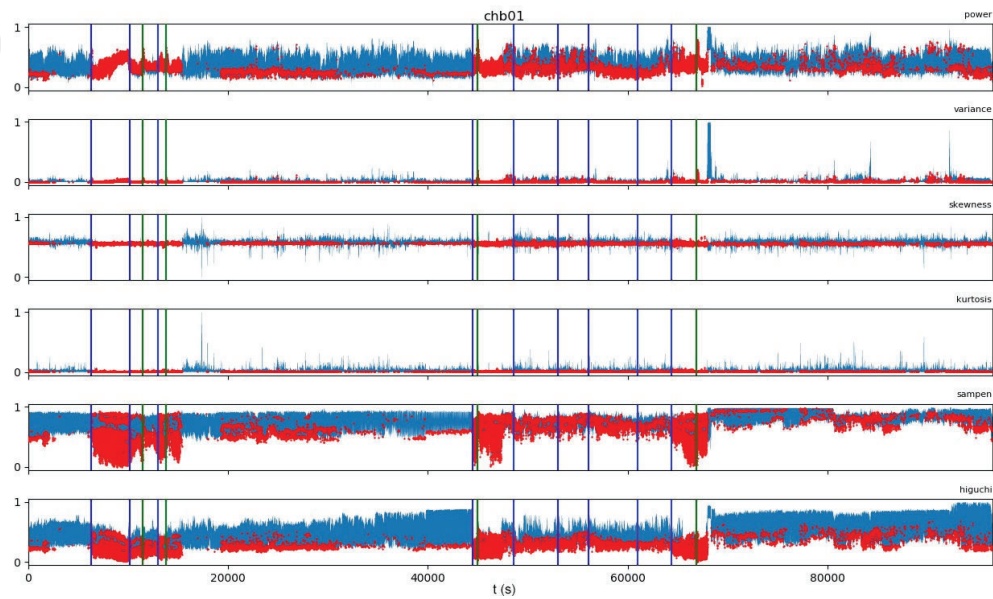
For sounding the alert of an incoming seizure, a sliding window on the predictions is used. Here, the window size is chosen as 30 predictions with no overlap, therefore, combined with the window length of the feature extraction stage, limiting the seizure prediction horizon to a minimum of $8 \times 30 = 240$ seconds, or 4 minutes.

The approach for the creation of the alerts is frequency-based. For every 30 predictions, the number of preictal predictions (1) counted and divided by 30 in order to limit the values to the range of $[0,1]$. After that, a threshold of 0.9 is set in order to decide when to raise the alarm. This threshold is chosen based on empirical observations of the frequency values seen in all cases. In other words, when 27 out of 30 predictions are preictal, an alert is given.

However, this approach sounds the alarm very frequently, especially within preictal periods. Therefore, a time limit is imposed on the algorithm where after an alarm is raised, it cannot be raised again. By this practice, the "spamming" of alerts is prevented, thus a real system that epileptic patients can use in their daily lives is approximated. This period of silence is chosen as 1200 seconds for this work, by observation of the intervals between seizures and the frequency of alerts. The alerts raised by the algorithm with and without the silent period are illustrated in Figure 3.4 on the features extracted from the first monocomponent of *chb01* recordings.



a)



b)

Figure 3.4 The features extracted from the first HVD subcomponent of chb01. The red points signify the windows predicted as preictal by the CNN. The green lines indicate the seizure onsets. Blue lines indicate the point where an alarm is raised. (a) Alarms raised on chb01 with no silent period. It is seen to become frequent during preictal periods. (b) Alarms raised on chb01 with the silent period. The silent period

can be seen to stop the alarm spam while also covering all three seizures for this patient.

3.7 Results

In order to measure the generalization performance of the proposed method across patients, the method was applied to all patients in the same manner, which is to extract features by sliding windows and decomposition via HVD, and 10-fold cross-validation of the CNN classifier. The sensitivity and the specificity metrics are calculated for every patient after acquiring predictions from cross-validation.

As can be seen from Table 3.1, the proposed non-patient specific method performed below the conventional thresholds. Especially in *chb11* and *chb19*, 0 sensitivity is observed, which means there were no correct preictal predictions by the CNN classifier. The only patient which the classifier can be said to perform well on is *chb01*, where reasonable sensitivity is achieved.

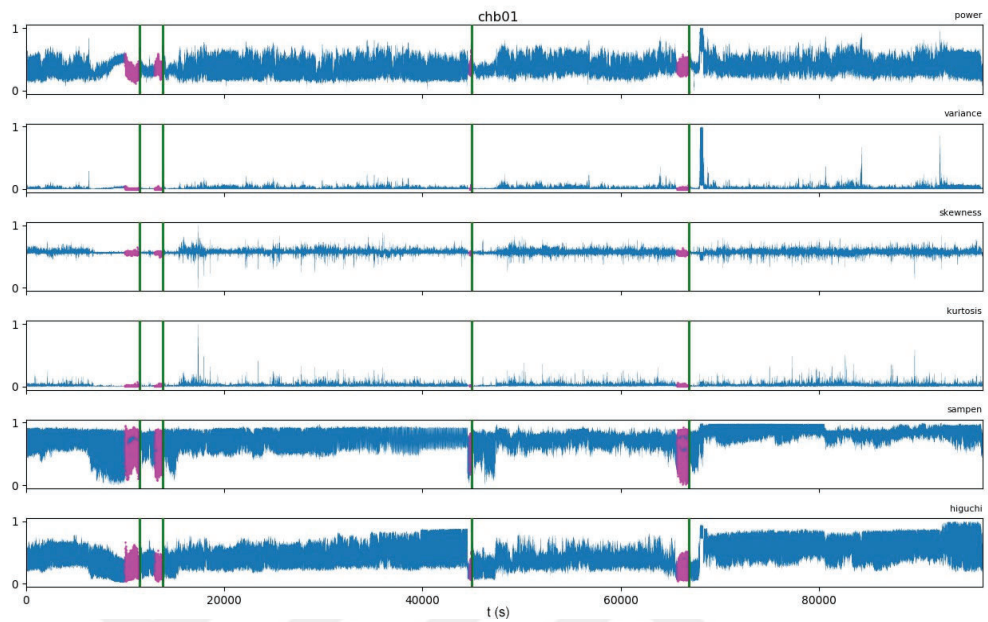
Table 3.1 Classification sensitivity and specificity of the patients. True positives are counted as the correct prediction as preictal.

Patient	% Sensitivity	% Specificity
<i>chb01</i>	70.3	85.3
<i>chb02</i>	6.3	56.3
<i>chb07</i>	17.3	53.4
<i>chb08</i>	27.3	50.6
<i>chb11</i>	0	71.3
<i>chb13</i>	5.2	52.8
<i>chb16</i>	20.8	46.7
<i>chb20</i>	16.2	80.3
<i>chb21</i>	11.2	40.1
<i>chb23</i>	24.3	54.6

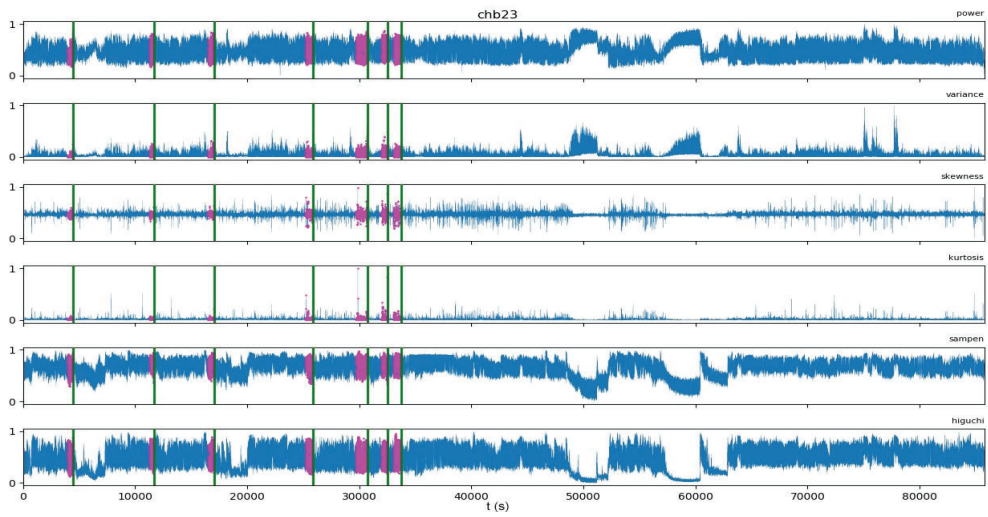
It is worth noticing that even with the class weight implementation, the classification process alone cannot successfully differentiate between preictal and non-preictal EEG signals. It is believed that this problem is originating from the fact that epileptic EEG recording from different patients is not similar in preictal stages, meaning that the change in a single feature of an epileptic EEG segment from one patient is not the same in a segment from another. For example, let us consider the features extracted from the

first monocomponents of the recordings of patients *chb01* and *chb23*, given in Figure 3.5

In Figure 3.5(a), it is seen that for the first monocomponent of *chb01*, there is a difference from the interictal segments that are observable, especially in spectral power and Higuchi's fractal dimension features. This change makes the interictal and preictal segments classifiable even to the naked eye. However, in Figure 3.5(b), it is seen that the same changes do not occur preictally for patient *chb01*. The preictal differences for this case are not manually observable but likely more subtle. This makes the classification process harder, especially in the case where the classifier is not provided information on the identity of the patient. In this work's non-patient specific method, as it was aimed to classify the preictal stages from all patients simultaneously, the classifier is exposed to both of these sections, and more, with the same label, where evidently, very little similarity is present. This difference in preictal trends are also seen in the classification scores in Table 1. The sensitivity of *chb01* is 70.3%, which is the highest achieved, while the sensitivity of *chb23* is 24.3%. The same case can be made with all other patients. The observable preictal trend of *chb01* is not seen in other patients. When the fact that many different types of epilepsy can trigger seizures in many different regions of or the entirety of the brain, this is not an unusual finding [1,67]. In addition to this, patients with an identified type of epilepsy can experience different types of seizures [68]. These observations previously made in the field of epilepsy are an indication that a non-patient specific algorithm may not be able to discern the preictal patterns from EEG signals of the patients. Furthermore, the localization of the EEG channels also affects the classification performance concerning epileptic seizures.



a)



b)

Figure 3.5 The features extracted from the first HVD subcomponent of (a) chb01 and (b) chb23. Green lines indicate the seizure onsets, and magenta sections indicate the signal points labeled as preictal.

In Figure 3.6, the classification sensitivity of epileptic seizures for different channels is illustrated. In Figure 3.6(a) it can be seen that most of the information related to preictal brain activity is contained in the parietal-occipital region of the patient's brain, with some preictal activity dominantly seen in the frontal and temporal lobes. Figure 3.6(b), it can be observed that the preictal activity is mostly contained to the right frontal cortex, but still minimally observed in the entire brain. Figure 3.6(c) shows that there is no discernible preictal activity in the entire cortex with the exceptions of the left temporal, right parietal and the right frontal-temporal regions, the latter conveying the most information. All three patients displayed in Figure 3.6 show preictal activity that is clear to the classifier in differing regions of the brain.

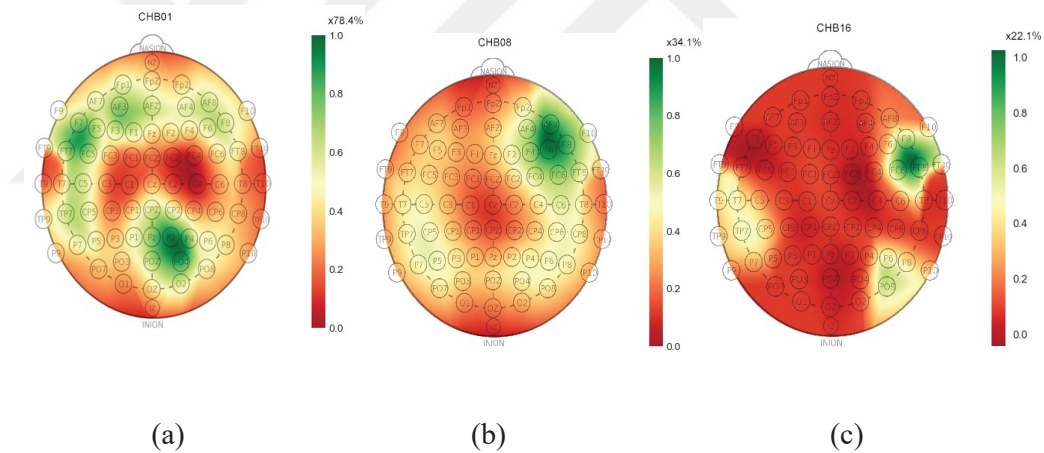


Figure 3.6 The classification sensitivity heatmaps from different channels of (a) *chb01*, (b) *chb08* and (c) *chb16*. All 7 HVD subcomponents have been used for this classification process.

All these factors, namely the type of epilepsy, the type of the seizures and the focal brain region, differ among patients. This makes a non-patient specific classification that, within the context of non-patient specific seizure process is non-trivially challenging. Therefore, it is concluded from the predictions with the HVD, it is not practical to rely on pure machine learning classifiers, and instead robust pre/postprocessing is needed. This can be seen in Table 3.2, with the sensitivity scores after alerting algorithm is applied. These results are critically indicative of the

performance improvement available with the use of additional methods to traditional classification.

Table 3.2 Alarm sensitivity and false alarm rate of the patients. True positives are counted as the alarms raised within 120 minutes before a seizure.

Patient	% Sensitivity	False Alarm/h
<i>chb01</i>	100	0.03
<i>chb02</i>	71.4	0.05
<i>chb07</i>	87.5	0.14
<i>chb08</i>	90	0.08
<i>chb11</i>	100	0.05
<i>chb13</i>	100	0.02
<i>chb16</i>	100	0.03
<i>chb20</i>	100	0.06
<i>chb21</i>	80	0.15
<i>chb23</i>	69.2	0.20
Mean	89.81	0.081

After the alerting process, the alarm sensitivity has shown at minimum 29.7% improvement for *chb01*, and at maximum 100% improvement for *chb11*. This drastic improvement of sensitivity, shows that even though the classification sensitivity can be considered low, or the number of correct predictions are not within a desirable range, the placement of the correct predictions are more frequently than not correct. Let us observe the predictions and alert placement of *chb01* on the features of the first extracted subcomponent.

In Figure 3.7, the preictal predictions can be seen to be spread throughout the entire signal but the preictal stages, which accounts for 0% classification sensitivity. However, it is seen that even though the classifier gives no correct preictal predictions, the alarm algorithm provides four alerts within the accepted 120-minute range preceding seizure onsets, while providing only two false alarms during interictal periods.

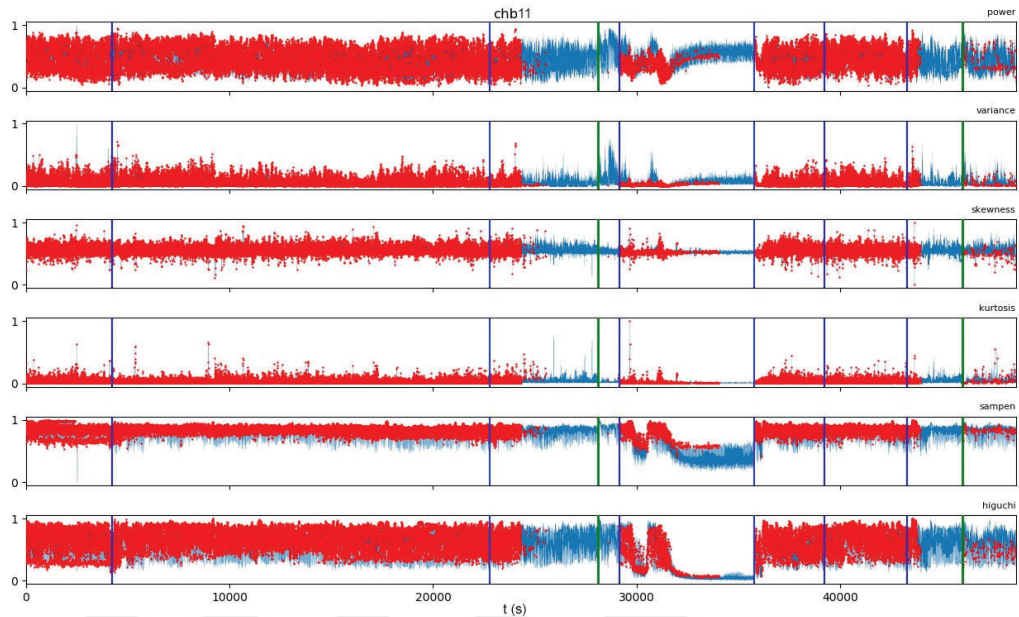


Figure 3.7 The alerts and classifier predictions for *chb11* illustrated on the features of the first subcomponent. Red points signify the preictal predictions by the classifier. The green vertical lines indicate seizure onsets and the blue vertical lines indicate alarms raised.

This indicates that even though the classifier performance alone cannot determine the correct preictal points, the frequency increase of the preictal predictions above the 0.9 thresholds still occurs towards, but not during, the preictal periods, therefore enabling the alarm algorithm to provide high sensitivity with low false alarm rate. This phenomenon likely stems from the fact that the non-patient specific classification approach imposes the decision criteria learned from the observations of other patients onto *chb11* in this example, therefore increasing the frequency of the preictal predictions outside of the zone labeled as preictal, therefore further indication to the notion that the classification alone may not provide high-performance seizure prediction, but together with a suitable postprocessing technique making correct seizure predictions is possible.

Furthermore, the epileptic seizure prediction performance of the framework proposed in this work is compared to other studies that employ various methods in order to achieve the same goal (Table 3.3). As can be seen in Table 3.3, the proposed framework was able to outperform other studies that utilize the same performance criteria. This framework can be seen to deliver higher sensitivity and lower false alarm

rates with the exception of [51], where the false alarm rates are equal. However, the proposed framework delivers an approximately 4% increase in the sensitivity metric. Therefore it is safe to say that the postprocessing step used in the presented framework does indeed increment upon the results that are previously reported in the literature concerning the prediction of epileptic seizures from EEG signals.

Table 3.3 Performance comparison between the methods of this work and other similar studies in the literature.

References	Sensitivity (%)	False Alarm Rate/h
[15]	83	0.17
[17]	70	0.15
[18]	-	0.41
[22]	85.11	0.08
[34]	75.8	0.10
This Study	89.81	0.08

4. CONCLUSION

This thesis addressed two different but complementary mechanisms for identifying and forecasting seizures. The first method consists of the extraction and analysis of EEG signals decomposed with Hilbert vibration decomposition in order to detect seizure activity with several classifiers. The results of the HVD process and other popular decomposition methods are also compared. The second approach expands on the technique of the first one by moving the classification goal to pre-seizure times in order to detect pre-seizure behavior and raise alarms using a convolutional neural network and a novel post-processing algorithm.

In the first framework, the classification of the epileptic states of the signals with the features extracted from the sub-signals of EEG signals obtained by HVD, EMD and MEMD methods. Signals from classes healthy, ictal and interictal are separated into sub-components and the attributes calculated from these sub-components are classified using MLP, SVM, RF, and kNN classifiers. In the results, higher metrics and classification accuracies were observed in the classification of HVD extracted features compared to MEMD and EMD extracted features. Therefore, the HVD method manifests itself as a reliable candidate in addition to the more common methods within the decomposition and analysis of epileptic EEG signals.

In the second framework, we have designed, applied, and demonstrated an epileptic seizure prediction system that extracts features from the subcomponents of EEG recordings from the CHB-MIT database presented on PhysioNet by employing the HVD, and uses these features for the binary classification as preictal or non-preictal with an CNN classifier. After the classification is done, a frequency-based alarm algorithm is used for alerting an oncoming epileptic seizure. This approach is shown to surpass the performances of other methods reported in the literature. The performance of the HVD method in epileptic seizure prediction is also shown in this study. The classification performance, in the conventional sense, was undesirable due

to the fact that many different factors affect the preictal trends across different patients. However, with a supporting alarm algorithm, correct seizure predictions were made. This shows that a single classifier may not be able to handle the intrinsic complexity of the problem of seizure detection, but a suitable postprocessing algorithm may build upon the classification performance and provide satisfactory sensitivity and false alarm rates. With the approach followed in this framework, a superior success is captured in epileptic seizure prediction; thus, a potential foundation to a robust and reliable epileptic seizure warning system which can be utilized in future studies is demonstrated.



REFERENCES

1. Fisher RS, Van Emde Boas W, Blume W, Elger C, Genton P, Lee P, et al. Response: Definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE) [4]. *Epilepsia*. 2005;46(10):1701–2.
2. WHO. Epilepsy - Fact sheet N°999 [Internet]. World Organization Health. 2015 [cited 2019 Feb 26]. Available from: <http://www.who.int/mediacentre/factsheets/fs999/en/>
3. Fisher RS, Cross JH, French JA, Higurashi N, Hirsch E, Jansen FE, et al. Operational classification of seizure types by the International League Against Epilepsy: position paper of the ILAE Commission for Classification and Terminology. *Zeitschrift fur Epileptol*. 2018;31(4):272–81.
4. Beghi E, Cornaggia C, Specchio LM, Specchio N, Boati E, Defanti CA, et al. Morbidity and accidents in patients with epilepsy: Results of a European cohort study. *Epilepsia*. 2002;43(9):1076–83.
5. Strzelczyk A, Griebel C, Lux W, Rosenow F, Reese JP. The burden of severely drug-refractory epilepsy: A comparative longitudinal evaluation of mortality, morbidity, resource use, and cost using German health insurance data. *Front Neurol*. 2017;8(DEC):712.
6. Ramchoun H, Amine M, Idrissi J. Multilayer Perceptron: Architecture Optimization and Training multi-criteria learning and nonlinear optimization View project. *Artic Int J Interact Multimed Artif Intell* [Internet]. 2016 [cited 2020 Jan 26];4:1–26. Available from: <https://www.researchgate.net/publication/292996667>
7. Suykens JAK, Vandewalle J. Least squares support vector machine classifiers. *Neural Process Lett*. 1999;9(3):293–300.
8. Pal M. Random forest classifier for remote sensing classification. *Int J Remote Sens*. 2005;26(1):217–22.
9. Andrzejak RG, Lehnertz K, Mormann F, Rieke C, David P, Elger CE. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Phys Rev E - Stat Physics, Plasmas, Fluids, Relat Interdiscip Top*. 2001;64(6):8.
10. Hammond JK, White PR. The analysis of non-stationary signals using time-frequency methods. *J Sound Vib*. 1996;190(3):419–47.
11. Huang NE, Shen Z, Long SR, Wu MC, Snin HH, Zheng Q, et al. The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis. *Proc R Soc A Math Phys Eng Sci*. 1998;454(1971):903–95.
12. Joshi V, Pachori RB, Vijesh A. Classification of ictal and seizure-free EEG signals using fractional linear prediction. *Biomed Signal Process Control*. 2014;9(1):1–5.

13. Feldman M. Time-varying vibration decomposition and analysis based on the Hilbert transform. *J Sound Vib.* 2006;295(3–5):518–30.
14. Liu B, Riemenschneider S, Xu Y. Gearbox fault diagnosis using empirical mode decomposition and Hilbert spectrum. *Mech Syst Signal Process.* 2006;20(3):718–34.
15. Sharma H, Sharma KK. Baseline wander removal of ECG signals using Hilbert vibration decomposition. *Electron Lett.* 2015;51(6):447–9.
16. Bertha M, Golinval JC. Identification of a time-varying beam using hilbert vibration decomposition. In: *Conference Proceedings of the Society for Experimental Mechanics Series.* Springer; 2014. p. 71–81.
17. Cicone A. Nonstationary Signal Decomposition for Dummies. In 2019. p. 69–82.
18. Hine GE, Maiorana E, Campisi P. Resting-State EEG: A Study on Its Non-Stationarity for Biometric Applications. In: *Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft fur Informatik (GI). Gesellschaft fur Informatik (GI);* 2017.
19. Rehman N, Mandic DP. Multivariate empirical mode decomposition. *Proc R Soc A Math Phys Eng Sci.* 2010;466(2117):1291–302.
20. Cho D, Min B, Kim J, Lee B. EEG-Based Prediction of Epileptic Seizures Using Phase Synchronization Elicited from Noise-Assisted Multivariate Empirical Mode Decomposition. *IEEE Trans Neural Syst Rehabil Eng.* 2017 Aug 1;25(8):1309–18.
21. Farahmand S, Sobayo T, Mogul DJ. Noise-Assisted Multivariate EMD-Based Mean-Phase Coherence Analysis to Evaluate Phase-Synchrony Dynamics in Epilepsy Patients. *IEEE Trans Neural Syst Rehabil Eng.* 2018 Dec 1;26(12):2270–9.
22. Braun S, Feldman M. Decomposition of non-stationary signals into varying time scales: Some aspects of the EMD and HVD methods. *Mech Syst Signal Process.* 2011;25(7):2608–30.
23. Feldman M. Theoretical analysis and comparison of the Hilbert transform decomposition methods. *Mech Syst Signal Process.* 2008;22(3):509–19.
24. Daoud HG, Abdelhameed AM, Bayoumi M. Automatic epileptic seizure detection based on empirical mode decomposition and deep neural network. In: *Proceedings - 2018 IEEE 14th International Colloquium on Signal Processing and its Application, CSPA 2018.* 2018. p. 182–6.
25. Mardia K V. Applications of Some Measures of Multivariate Skewness and Kurtosis in Testing Normality and Robustness Studies Published by : Springer on behalf of the Indian Statistical Institute Your use of the JSTOR archive indicates your. *Sankhyā Indian J Stat.* 1974;36(2):115–38.
26. DeCarlo LT. On the Meaning and Use of Kurtosis. *Psychol Methods.* 1997;2(3):292–307.
27. Dietterich TG, Kong EB. *Machine Learning Bias, Statistical Bias, and Statistical Variance of Decision Tree Algorithms.* 1995.
28. Jost L. Entropy and diversity. *Oikos.* 2006;113(2):363–75.
29. Maszczyk T, Duch W. Comparison of shannon, renyi and tsallis entropy used in decision trees. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* 2008. p. 643–51.

30. Jain AK, Mao J, Mohiuddin KM. Artificial neural networks: A tutorial. Vol. 29, Computer. 1996. p. 31–44.
31. Fesghandis GS. Comparison of Multilayer Perceptron and Radial Basis Function Neural Networks in Predicting the Success of New Product Development. Eng Technol Appl Sci Res. 2017;
32. Noriega L. Multilayer perceptron tutorial. Sch Comput Staff Univ. 2005;
33. Hecht-Nielsen R. Theory of the backpropagation neural network. In 1989.
34. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995 Sep;20(3):273–97.
35. Amari S, Wu S. Improving support vector machine classifiers by modifying kernel functions. Neural Networks. 1999 Jul;12(6):783–9.
36. Rutkowski L, Jaworski M, Pietruczuk L, Duda P. The CART Decision Tree for Mining Data Streams.
37. Raileanu LE, Stoffel K. Theoretical comparison between the Gini Index and Information Gain criteria *. Vol. 41, Annals of Mathematics and Artificial Intelligence. Kluwer Academic Publishers; 2004.
38. Liaw A, Wiener M. Classification and Regression by RandomForest [Internet]. 2002 [cited 2020 Jan 27]. Available from: <https://www.researchgate.net/publication/228451484>
39. Rodríguez JD, Pérez A, Lozano JA. Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. IEEE Trans Pattern Anal Mach Intell. 2010;32(3):569–75.
40. Tsochantaridis I, Hofmann T, Joachims T, Altun Y. Support vector machine learning for interdependent and structured output spaces. In: Proceedings, Twenty-First International Conference on Machine Learning, ICML 2004. 2004.
41. Tatum IV WO. Handbook of EEG Interpretation. Handbook of EEG Interpretation. Demos Medical Publishing; 2014.
42. Iasemidis LD, Sackellares JC. The evolution with time of the spatial distribution of the largest Lyapunov exponent on the human epileptic cortex. Meas chaos Hum brain. 1991;49–82.
43. Mormann F, Lehnertz K, David P, Elger C. Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. Phys D Nonlinear Phenom. 2000;144(3):358–69.
44. Iasemidis LD, Shiao DS, Chaovalitwongse W, Sackellares JC, Pardalos PM, Principe JC, et al. Adaptive Epileptic Seizure Prediction System. IEEE Trans Biomed Eng. 2003;50(5):616–27.
45. Gigola S, Ortiz F, D’Attellis CE, Silva W, Kochen S. Prediction of epileptic seizures using accumulated energy in a multiresolution framework. J Neurosci Methods. 2004;138(1–2):107–11.
46. Schelter B, Winterhalder M, Maiwald T, Brandt A, Schad A, Schulze-Bonhage A, et al. Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction. Chaos. 2006;16(1):13108.
47. Chisci L, Mavino A, Perferi G, Sciandrone M, Anile C, Colicchio G, et al. Real-time epileptic seizure prediction using AR models and support vector machines. IEEE Trans Biomed Eng. 2010;57(5):1124–32.
48. Song Y, Crowcroft J, Zhang J. Automatic epileptic seizure detection in EEGs based on optimized sample entropy and extreme learning machine. J Neurosci

- Methods. 2012;210(2):132–46.
49. Parvez MZ, Paul M. Epileptic seizure prediction by exploiting spatiotemporal relationship of EEG signals using phase correlation. *IEEE Trans Neural Syst Rehabil Eng.* 2016;24(1):158–68.
 50. Yang Y, Zhou M, Niu Y, Li C, Cao R, Wang B, et al. Epileptic seizure prediction based on permutation entropy. *Front Comput Neurosci.* 2018;12.
 51. Yuan S, Zhou W, Chen L. Epileptic Seizure Prediction Using Diffusion Distance and Bayesian Linear Discriminate Analysis on Intracranial EEG. *Int J Neural Syst.* 2018;28(1):1750043.
 52. Shoeb AH. Application of machine learning to epileptic seizure onset detection and treatment. Massachusetts Institute of Technology; 2009.
 53. PhysioBank P. PhysioNet: components of a new research resource for complex physiologic signals. *Circ v101 i23 e215-e220.* 2000;
 54. Shoeb A, Edwards H, Connolly J, Bourgeois B, Ted Treves S, Gutttag J. Patient-specific seizure onset detection. *Epilepsy Behav.* 2004;5(4):483–98.
 55. Richman JS, Lake DE, Moorman JR. Sample Entropy. *Methods Enzymol.* 2004;384:172–84.
 56. Yentes JM, Hunt N, Schmid KK, Kaipust JP, McGrath D, Stergiou N. The appropriate use of approximate entropy and sample entropy with short data sets. *Ann Biomed Eng.* 2013;41(2):349–65.
 57. Park Y, Luo L, Parhi KK, Netoff T. Seizure prediction with spectral power of EEG using cost-sensitive support vector machines. *Epilepsia.* 2011;52(10):1761–70.
 58. Bandarabadi M, Teixeira CA, Rasekhi J, Dourado A. Epileptic seizure prediction using relative spectral power features. *Clin Neurophysiol.* 2015;126(2):237–48.
 59. Welch PD. The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms. *IEEE Trans Audio Electroacoust.* 1967;15(2):70–3.
 60. Mandelbrot BB. Self-affine fractals and fractal dimension. *Phys Scr.* 1985;
 61. Higuchi T. Approach to an irregular time series on the basis of the fractal theory. *Phys D Nonlinear Phenom.* 1988;31(2):277–83.
 62. Esteller R, Vachtsevanos G, Echauz J, Litt B. A Comparison of waveform fractal dimension algorithms. *IEEE Trans Circuits Syst I Fundam Theory Appl.* 2001;48(2):177–83.
 63. Kesić S, Spasić SZ. Application of Higuchi’s fractal dimension from basic to clinical neurophysiology: A review. *Comput Methods Programs Biomed.* 2016;133:55–70.
 64. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems.* 2012. p. 1097–105.
 65. Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Li FF. Large-scale video classification with convolutional neural networks. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* 2014. p. 1725–32.
 66. Simard PY, Steinkraus D, Platt JC. Best practices for convolutional neural networks applied to visual document analysis. In: *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR.*

2003. p. 958–63.
67. Gastaut H, Gastaut JL, Silva GEG e., Sanchez GRF. Relative Frequency of Different Types of Epilepsy: A Study Employing the Classification of the International League Against Epilepsy. *Epilepsia*. 1975;16(3):457–61.
 68. Williamson PD, Thadani VM, Darcey TM, Spencer DD, Spencer SS, Mattson RH. Occipital lobe epilepsy: Clinical characteristics, seizure spread patterns, and results of surgery. *Ann Neurol*. 1992;31(1):3–13.

APPENDIX

Results of 10-fold cross validation with all classifiers are provided in this section.

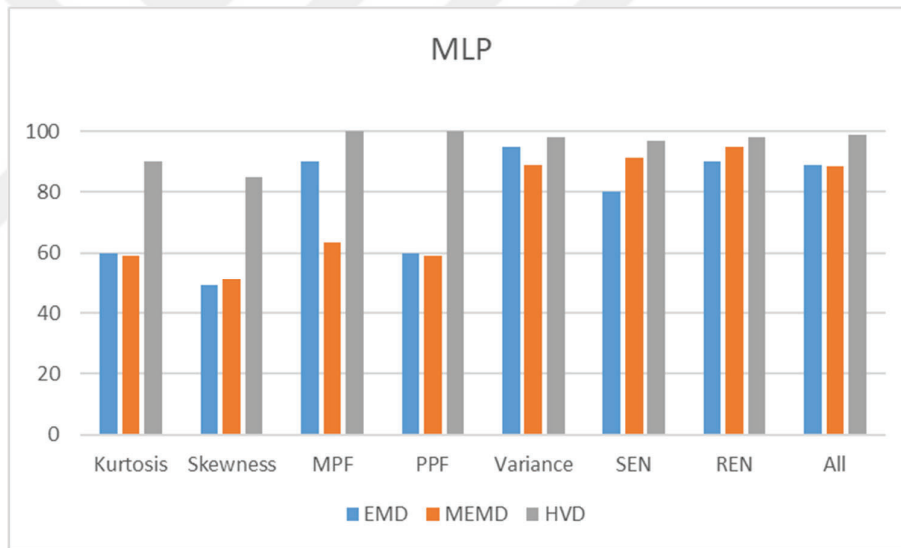


Figure A.1 10-fold cross-validation accuracies of the classification process with MLP classifier

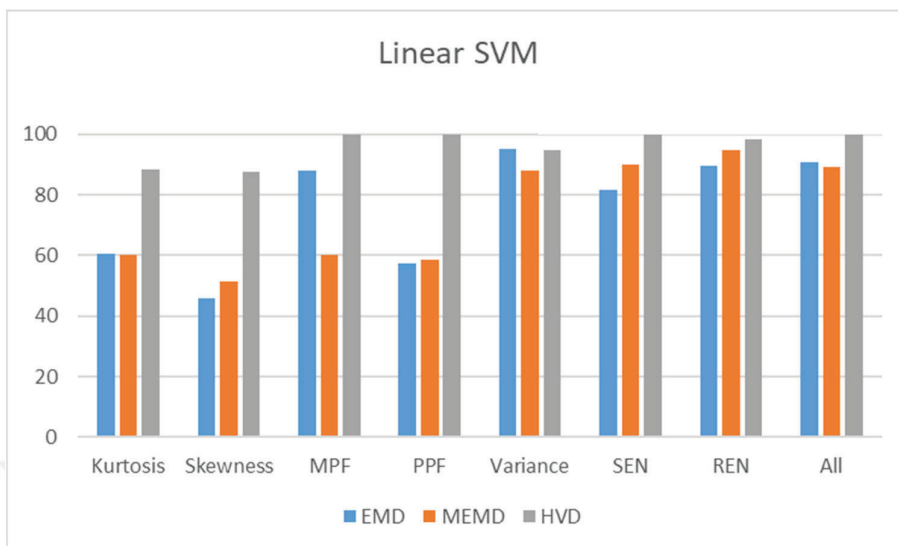


Figure A.2 10-fold cross validation accuracies of classification process with linear kernel SVM classifier

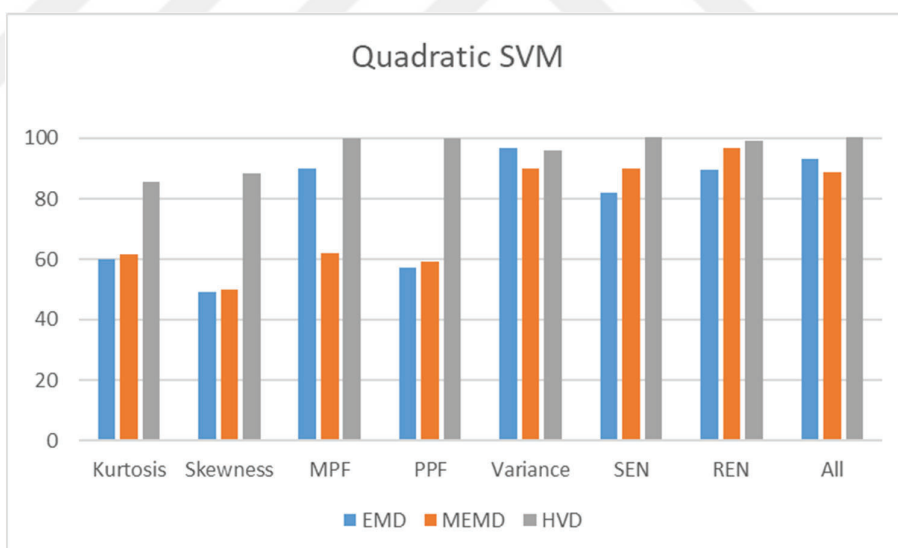


Figure A.3 10-fold cross validation accuracies of classification process with quadratic kernel SVM classifier

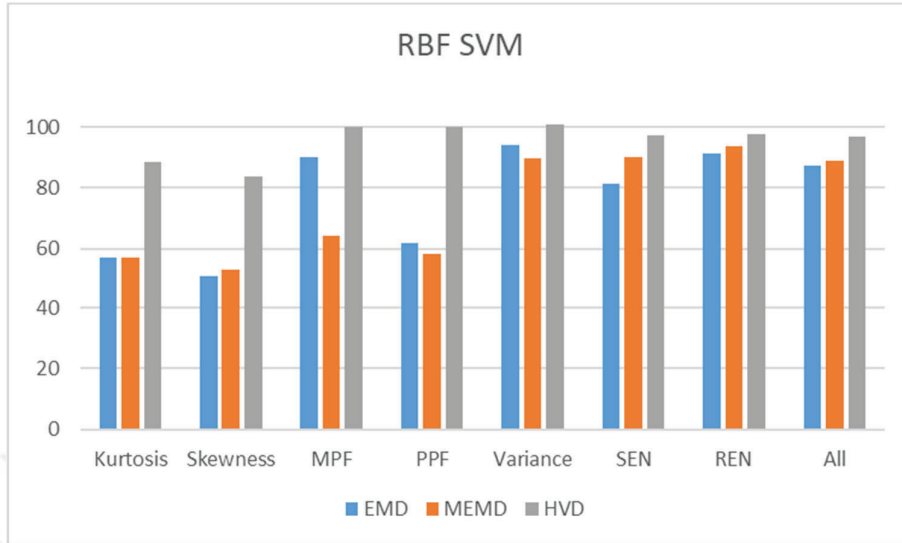


Figure A.4 10-fold cross validation accuracies of classification process with RBF kernel SVM classifier

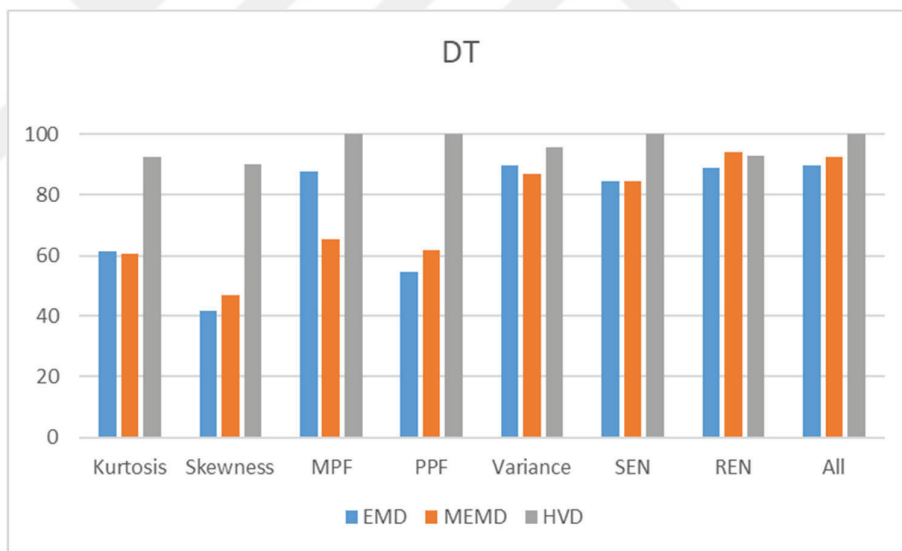


Figure A.5 10-fold cross validation accuracies of the classification process with DT classifier

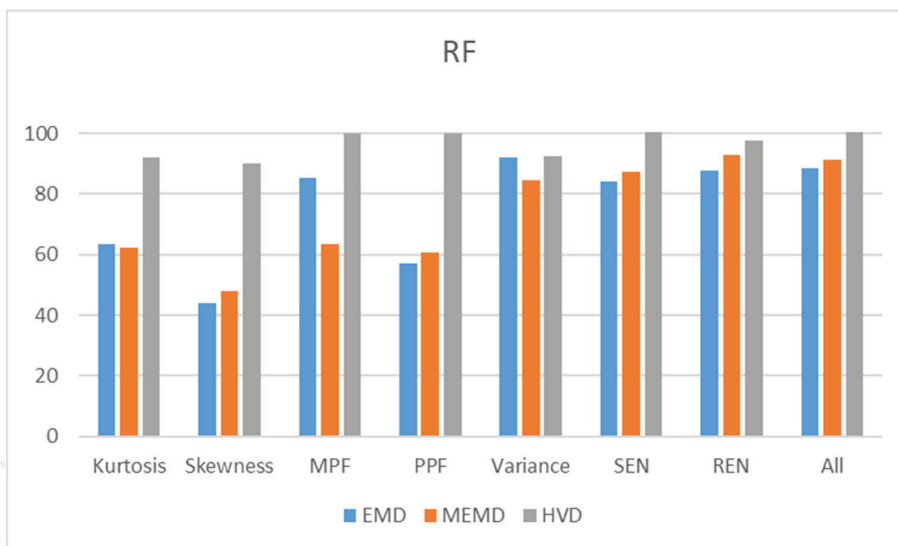


Figure A.6 10-fold cross validation accuracies of classification process with RF classifier

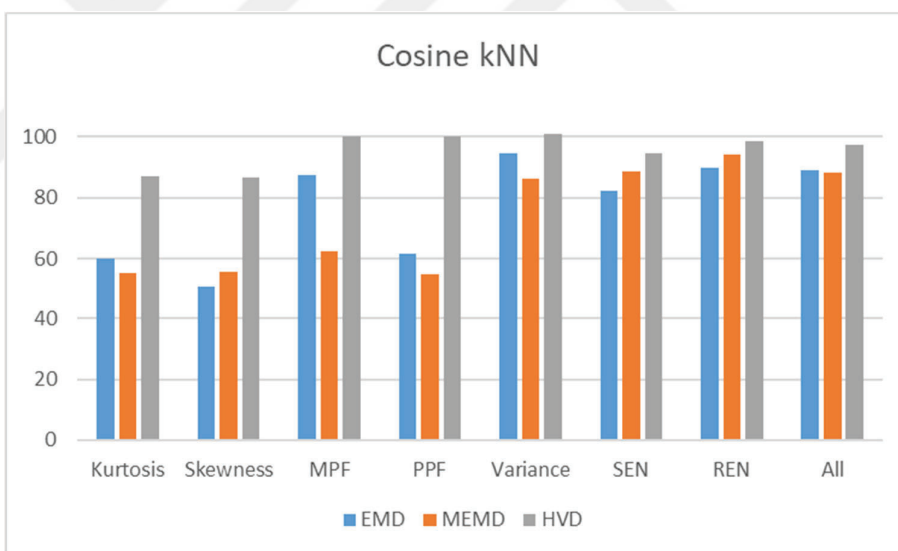


Figure A.7 10-fold cross validation accuracies of classification process Cosine distance kNN classifier

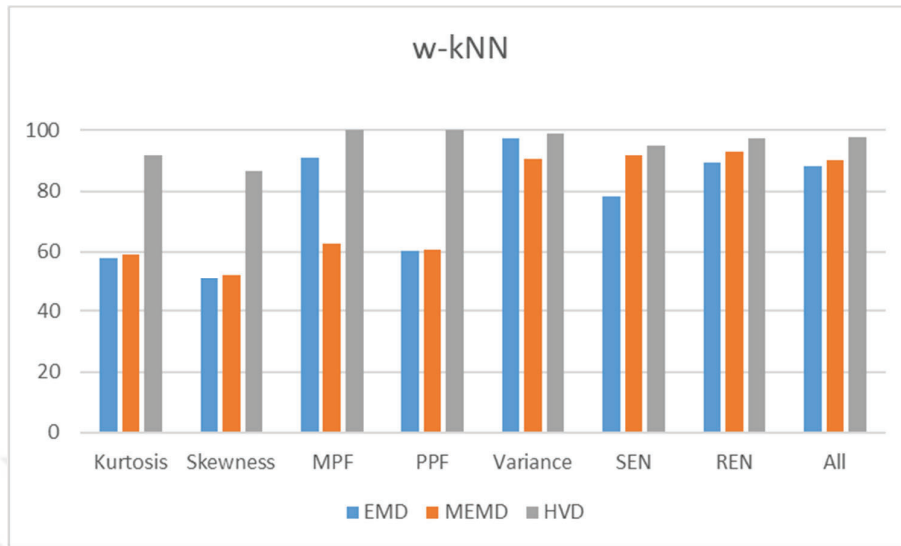


Figure A.8 10-fold cross validation accuracies of classification process with weighted Euclidean distance kNN classifier

Table A.1 MLP classification metrics of all classes

MLP		EMD			HVD		
		Z	S	F	Z	S	F
Kurtosis	<i>P</i>	0.53	0.45	0.82	0.82	0.85	0.99
	<i>R</i>	0.48	0.50	0.81	0.85	0.83	0.97
	<i>F1</i>	0.51	0.47	0.81	0.83	0.84	0.98
Skewness	<i>P</i>	0.40	0.41	0.66	0.80	0.96	0.80
	<i>R</i>	0.39	0.43	0.63	0.80	0.95	0.81
	<i>F1</i>	0.39	0.42	0.64	0.80	0.95	0.81
MPF	<i>P</i>	0.65	0.67	0.82	1.00	1.00	1.00
	<i>R</i>	0.66	0.69	0.79	1.00	1.00	1.00
	<i>F1</i>	0.66	0.68	0.81	1.00	1.00	1.00
PPF	<i>P</i>	0.55	0.54	0.77	1.00	1.00	1.00
	<i>R</i>	0.64	0.52	0.68	1.00	1.00	1.00
	<i>F1</i>	0.59	0.53	0.72	1.00	1.00	1.00
Variance	<i>P</i>	0.83	0.82	0.90	0.96	0.96	0.97
	<i>R</i>	0.75	0.89	0.92	0.95	0.96	0.98
	<i>F1</i>	0.79	0.86	0.91	0.95	0.96	0.98
SEN	<i>P</i>	0.83	0.82	0.90	0.96	0.96	0.97
	<i>R</i>	0.75	0.89	0.92	0.95	0.96	0.98
	<i>F1</i>	0.79	0.86	0.91	0.95	0.96	0.98
REN	<i>P</i>	0.93	0.88	0.97	0.97	0.98	1.00
	<i>R</i>	0.91	0.91	0.96	1.00	0.98	0.97
	<i>F1</i>	0.92	0.90	0.96	0.99	0.98	0.98
All	<i>P</i>	0.92	0.88	0.94	0.99	1.00	0.97
	<i>R</i>	0.85	0.94	0.94	0.98	0.98	1.00
	<i>F1</i>	0.88	0.91	0.94	0.98	0.99	0.99

Table A.2 Linear SVM classification metrics of all classes

Linear SVM		EMD			HVD		
		Z	S	F	Z	S	F
Kurtosis	<i>P</i>	0.54	0.47	0.80	0.87	0.90	0.99
	<i>R</i>	0.45	0.57	0.79	0.91	0.84	0.92
	<i>F1</i>	0.51	0.51	0.81	0.81	0.91	0.93
Skewness	<i>P</i>	0.47	0.37	0.72	0.77	1.00	0.78
	<i>R</i>	0.39	0.42	0.67	0.85	0.97	0.74
	<i>F1</i>	0.32	0.42	0.61	0.75	0.95	0.87
MPF	<i>P</i>	0.63	0.61	0.81	1.00	1.00	1.00
	<i>R</i>	0.66	0.76	0.79	1.00	1.00	1.00
	<i>F1</i>	0.66	0.62	0.88	1.00	1.00	1.00
PPF	<i>P</i>	0.48	0.58	0.71	1.00	1.00	1.00
	<i>R</i>	0.65	0.54	0.66	1.00	1.00	1.00
	<i>F1</i>	0.62	0.47	0.68	1.00	1.00	1.00
Variance	<i>P</i>	0.86	0.75	0.87	0.91	0.91	1.00
	<i>R</i>	0.80	0.96	0.89	0.95	0.98	1.00
	<i>F1</i>	0.84	0.79	0.92	0.95	0.98	0.97
SEN	<i>P</i>	0.87	0.87	0.86	0.97	0.91	0.93
	<i>R</i>	0.76	0.93	0.99	0.99	0.93	0.97
	<i>F1</i>	0.86	0.79	0.89	0.98	0.90	0.99
REN	<i>P</i>	0.95	0.92	0.95	1.00	1.00	1.00
	<i>R</i>	0.98	0.84	0.99	1.00	0.91	0.93
	<i>F1</i>	0.93	0.95	1.00	0.98	0.97	0.99
All	<i>P</i>	0.97	0.95	0.95	1.00	1.00	0.93
	<i>R</i>	0.84	0.93	0.95	0.96	0.93	1.00
	<i>F1</i>	0.90	0.95	1.00	1.00	0.95	0.95

Table A.3 Quadratic SVM classification metrics of all classes

Quadratic SVM		EMD			HVD		
		Z	S	F	Z	S	F
Kurtosis	<i>P</i>	0.55	0.45	0.77	0.82	0.96	1.00
	<i>R</i>	0.44	0.60	0.74	0.85	0.88	0.86
	<i>F1</i>	0.53	0.56	0.79	0.74	0.92	0.95
Skewness	<i>P</i>	0.47	0.41	0.74	0.77	1.00	0.83
	<i>R</i>	0.40	0.36	0.64	0.84	0.94	0.81
	<i>F1</i>	0.30	0.34	0.58	0.76	0.97	0.92
MPF	<i>P</i>	0.66	0.55	0.86	1.00	1.00	1.00
	<i>R</i>	0.72	0.78	0.75	1.00	1.00	1.00
	<i>F1</i>	0.70	0.69	0.84	1.00	1.00	1.00
PPF	<i>P</i>	0.42	0.63	0.71	1.00	1.00	1.00
	<i>R</i>	0.59	0.47	0.68	1.00	1.00	1.00
	<i>F1</i>	0.55	0.47	0.64	1.00	1.00	1.00
Variance	<i>P</i>	0.93	0.69	0.91	0.88	0.89	1.00
	<i>R</i>	0.77	0.94	0.94	0.94	1.00	1.00
	<i>F1</i>	0.88	0.86	0.91	1.00	1.00	1.00
SEN	<i>P</i>	0.94	0.91	0.92	0.96	0.90	0.92
	<i>R</i>	0.77	0.85	0.96	0.99	1.00	0.91
	<i>F1</i>	0.82	0.77	0.91	1.00	0.95	0.96
REN	<i>P</i>	1.00	0.87	0.95	0.99	0.96	1.00
	<i>R</i>	1.00	0.81	1.00	1.00	0.90	0.87
	<i>F1</i>	0.97	0.96	1.00	0.95	0.95	0.93
All	<i>P</i>	0.98	0.96	1.00	1.00	1.00	0.99
	<i>R</i>	0.87	0.95	0.98	0.91	0.95	1.00
	<i>F1</i>	0.90	0.98	0.96	0.94	1.00	0.94

Table A.4 RBF SVM classification metrics of all classes

RBF SVM		EMD			HVD		
		Z	S	F	Z	S	F
Kurtosis	<i>P</i>	0.49	0.39	0.73	0.78	0.94	1.00
	<i>R</i>	0.49	0.58	0.75	0.89	0.86	0.87
	<i>F1</i>	0.59	0.57	0.77	0.76	0.96	0.90
Skewness	<i>P</i>	0.51	0.45	0.73	0.79	0.94	0.76
	<i>R</i>	0.35	0.40	0.58	0.80	0.86	0.84
	<i>F1</i>	0.24	0.27	0.65	0.69	0.94	0.96
MPF	<i>P</i>	0.60	0.62	0.82	1.00	1.00	1.00
	<i>R</i>	0.68	0.84	0.82	1.00	1.00	1.00
	<i>F1</i>	0.69	0.65	0.85	1.00	1.00	1.00
PPF	<i>P</i>	0.50	0.70	0.75	1.00	1.00	1.00
	<i>R</i>	0.60	0.43	0.62	1.00	1.00	1.00
	<i>F1</i>	0.62	0.43	0.69	1.00	1.00	1.00
Variance	<i>P</i>	0.92	0.72	0.96	0.84	0.92	1.00
	<i>R</i>	0.78	0.96	0.93	0.94	1.00	1.00
	<i>F1</i>	0.95	0.82	0.95	0.98	1.00	1.00
SEN	<i>P</i>	0.88	0.90	0.85	0.92	0.87	0.90
	<i>R</i>	0.79	0.93	0.91	0.98	0.92	0.91
	<i>F1</i>	0.84	0.77	0.86	0.97	0.92	1.00
REN	<i>P</i>	1.00	0.94	0.92	0.92	0.94	1.00
	<i>R</i>	1.00	0.88	1.00	1.00	0.95	0.87
	<i>F1</i>	0.95	1.00	1.00	0.95	0.91	1.00
All	<i>P</i>	1.00	1.00	0.99	1.00	1.00	0.97
	<i>R</i>	0.89	1.00	0.92	0.91	0.97	1.00
	<i>F1</i>	0.94	1.00	0.96	0.90	1.00	0.89

Table A.5 DT classification metrics of all classes

DT		EMD			HVD		
		Z	S	F	Z	S	F
Kurtosis	<i>P</i>	0.51	0.37	0.80	0.80	0.99	1.00
	<i>R</i>	0.49	0.59	0.70	0.89	0.93	0.93
	<i>F1</i>	0.58	0.55	0.82	0.77	0.95	0.85
Skewness	<i>P</i>	0.50	0.42	0.75	0.74	0.98	0.74
	<i>R</i>	0.30	0.38	0.51	0.81	0.81	0.87
	<i>F1</i>	0.28	0.31	0.59	0.66	1.00	1.00
MPF	<i>P</i>	0.59	0.62	0.89	1.00	1.00	1.00
	<i>R</i>	0.64	0.86	0.75	1.00	1.00	1.00
	<i>F1</i>	0.63	0.66	0.85	1.00	1.00	1.00
PPF	<i>P</i>	0.43	0.75	0.79	1.00	1.00	1.00
	<i>R</i>	0.54	0.37	0.59	1.00	1.00	1.00
	<i>F1</i>	0.60	0.48	0.68	1.00	1.00	1.00
Variance	<i>P</i>	0.90	0.74	0.89	0.88	0.86	1.00
	<i>R</i>	0.84	0.99	0.96	0.87	1.00	0.99
	<i>F1</i>	0.97	0.83	0.95	0.92	1.00	1.00
SEN	<i>P</i>	0.84	0.96	0.83	0.90	0.94	0.98
	<i>R</i>	0.77	0.92	0.85	0.98	0.94	0.84
	<i>F1</i>	0.81	0.69	0.89	0.93	0.89	0.98
REN	<i>P</i>	1.00	0.98	0.85	0.87	0.97	1.00
	<i>R</i>	1.00	0.83	1.00	1.00	0.91	0.86
	<i>F1</i>	1.00	1.00	1.00	0.93	0.86	1.00
All	<i>P</i>	1.00	1.00	0.97	1.00	1.00	0.91
	<i>R</i>	0.85	0.98	0.97	0.84	1.00	1.00
	<i>F1</i>	0.87	1.00	0.99	0.87	1.00	0.91

Table A.6 RF classification metrics of all classes

RF		EMD			HVD		
		Z	S	F	Z	S	F
Kurtosis	<i>P</i>	0.71	0.80	0.72	0.92	0.97	0.94
	<i>R</i>	0.73	0.86	0.65	0.94	0.98	0.91
	<i>F1</i>	0.72	0.83	0.68	0.93	0.98	0.92
Skewness	<i>P</i>	0.49	0.57	0.38	0.78	0.76	0.76
	<i>R</i>	0.52	0.51	0.40	0.85	0.77	0.68
	<i>F1</i>	0.50	0.54	0.39	0.81	0.77	0.72
MPF	<i>P</i>	0.87	0.80	0.88	1.00	1.00	1.00
	<i>R</i>	0.89	0.85	0.81	1.00	1.00	1.00
	<i>F1</i>	0.88	0.83	0.84	1.00	1.00	1.00
PPF	<i>P</i>	0.80	0.68	0.68	1.00	1.00	1.00
	<i>R</i>	0.78	0.69	0.68	1.00	1.00	1.00
	<i>F1</i>	0.79	0.68	0.68	1.00	1.00	1.00
Variance	<i>P</i>	0.89	0.93	0.95	0.94	0.96	1.00
	<i>R</i>	0.92	0.92	0.93	0.98	0.95	0.97
	<i>F1</i>	0.91	0.92	0.94	0.96	0.95	0.98
SEN	<i>P</i>	0.75	0.69	0.93	0.96	0.96	0.96
	<i>R</i>	0.80	0.66	0.91	0.99	0.94	0.95
	<i>F1</i>	0.77	0.68	0.92	0.98	0.95	0.95
REN	<i>P</i>	0.92	0.91	0.96	0.96	0.94	0.98
	<i>R</i>	0.95	0.88	0.96	0.99	0.95	0.94
	<i>F1</i>	0.94	0.89	0.96	0.98	0.95	0.96
All	<i>P</i>	0.86	0.91	0.92	0.97	0.98	0.99
	<i>R</i>	0.90	0.88	0.90	0.99	0.97	0.98
	<i>F1</i>	0.88	0.89	0.91	0.98	0.97	0.98

Table A.7 Cosine kNN classification metrics of all classes

Cos kNN		EMD			HVD		
		Z	S	F	Z	S	F
Kurtosis	<i>P</i>	0.50	0.33	0.92	0.87	0.99	1.00
	<i>R</i>	0.56	0.70	0.73	0.95	0.88	1.00
	<i>F1</i>	0.54	0.45	0.79	0.87	0.98	0.94
Skewness	<i>P</i>	0.51	0.34	0.76	0.84	0.88	0.78
	<i>R</i>	0.26	0.40	0.45	0.82	0.71	0.96
	<i>F1</i>	0.27	0.23	0.57	0.72	1.00	1.00
MPF	<i>P</i>	0.72	0.48	0.85	1.00	1.00	1.00
	<i>R</i>	0.52	0.86	0.77	1.00	1.00	1.00
	<i>F1</i>	0.64	0.75	0.76	1.00	1.00	1.00
PPF	<i>P</i>	0.32	0.74	0.80	1.00	1.00	1.00
	<i>R</i>	0.53	0.38	0.58	1.00	1.00	1.00
	<i>F1</i>	0.66	0.39	0.67	1.00	1.00	1.00
Variance	<i>P</i>	1.00	0.86	0.88	0.81	0.91	1.00
	<i>R</i>	0.82	0.94	0.92	0.92	1.00	1.00
	<i>F1</i>	1.00	0.77	1.00	0.85	1.00	0.96
SEN	<i>P</i>	0.71	0.95	0.94	0.92	0.88	1.00
	<i>R</i>	0.81	0.96	0.81	1.00	0.95	0.89
	<i>F1</i>	0.78	0.68	0.94	0.91	0.92	0.90
REN	<i>P</i>	1.00	1.00	0.79	0.92	1.00	1.00
	<i>R</i>	1.00	0.84	1.00	1.00	0.81	0.84
	<i>F1</i>	1.00	1.00	0.95	1.00	0.86	0.93
All	<i>P</i>	1.00	1.00	0.90	1.00	1.00	0.91
	<i>R</i>	0.90	0.96	0.91	0.83	1.00	1.00
	<i>F1</i>	1.00	0.99	0.90	0.87	1.00	0.87

Table A.8 Weighted kNN classification metrics of all classes

w-kNN		EMD			HVD		
		Z	S	F	Z	S	F
Kurtosis	<i>P</i>	0.45	0.32	0.97	0.82	1.00	1.00
	<i>R</i>	0.49	0.67	0.77	1.00	0.86	1.00
	<i>F1</i>	0.47	0.51	0.82	0.91	0.97	0.89
Skewness	<i>P</i>	0.56	0.39	0.80	0.86	0.94	0.79
	<i>R</i>	0.20	0.36	0.43	0.83	0.66	0.98
	<i>F1</i>	0.30	0.18	0.64	0.66	1.00	1.00
MPF	<i>P</i>	0.71	0.48	0.86	1.00	1.00	1.00
	<i>R</i>	0.60	0.91	0.80	1.00	1.00	1.00
	<i>F1</i>	0.69	0.82	0.74	1.00	1.00	1.00
PPF	<i>P</i>	0.27	0.69	0.82	1.00	1.00	1.00
	<i>R</i>	0.56	0.32	0.64	1.00	1.00	1.00
	<i>F1</i>	0.73	0.38	0.61	1.00	1.00	1.00
Variance	<i>P</i>	1.00	0.91	0.92	0.75	0.84	1.00
	<i>R</i>	0.75	0.99	0.99	0.98	1.00	1.00
	<i>F1</i>	1.00	0.73	1.00	0.90	1.00	0.94
SEN	<i>P</i>	0.65	0.90	0.96	0.88	0.86	1.00
	<i>R</i>	0.87	0.93	0.83	0.96	0.99	0.83
	<i>F1</i>	0.82	0.65	0.93	0.85	0.89	0.89
REN	<i>P</i>	1.00	1.00	0.77	0.92	1.00	1.00
	<i>R</i>	0.94	0.87	1.00	1.00	0.87	0.85
	<i>F1</i>	1.00	1.00	0.91	1.00	0.84	0.99
All	<i>P</i>	1.00	1.00	0.84	0.98	1.00	0.99
	<i>R</i>	0.97	1.00	0.90	0.84	0.98	1.00
	<i>F1</i>	1.00	1.00	0.89	0.89	1.00	0.89

CURRICULUM VITAE

Barkın BÜYÜKÇAKIR was born in Ankara on 10 February 1995. He graduated from İzmir Tevfik Fikret Elemenraty School in 2009 and from İzmir Cihat Kora Highschool in 2014. He started attending İzmir Katip Çelebi University, Electrical and Electronics Engineering Department in 2014. He focused on control, programming and machine learning fields in his undergraduate studies until 2018, when he graduated 5th of his class with outstanding achievement award. During this time, he took place in two projects supported by The Scientific and Technological Research Council of Turkey (TÜBİTAK), as project leader and as a researcher. He also participated in Turkey University Tennis League for 3 years as captain of tennis team, where he and the team brought home two second place trophies in 2018 and 2019. After he graduated, he was accepted to the Electrical and Electronics Engineering Master's Degree Program in İzmir Katip Çelebi University in the 1st place. His passion for artificial intelligence lead him to study in the field of machine learning, where he joined another project supported by TÜBİTAK. In the project, he had to opportunity to heavily focus on machine learning and deep learning, where he was able to put his enthusiasm for the field to work. He also participated in other studies where he was able to apply his knowledge in machine learning. His M.Sc. studies currently continue with 4.00/4.00 GPA.

List of Research Projects:

1. Design and Implementation of Intel Atom Processor-Based Embedded System for Furuta Reverse Pendulum Control Experiment Set. TÜBİTAK, Project No: 1919B11700425.
2. Control of Multi-Input Multi-Output Twin Rotor Set with LattePanda-Based Embedded System. TÜBİTAK, Project No: 1919B11700425.
3. Epileptic Seizure Classification in EEG Recordings Using Signal Decomposition Techniques. TÜBİTAK, Project No: 117E137.

List of Publications:

1. Elmaz F, Büyükçakır B, Yücel Ö, Mutlu AY. Classification of solid fuels with machine learning. *Fuel*. 2020;266(July 2019):117066.
2. Büyükçakır B, Elmaz F, Mutlu AY. Hilbert Vibration Decomposition-based Epileptic Seizure Prediction with Neural Network. *Computers in Biology and Medicine*. 2020. (Under Revision)
3. Büyükçakır B, Mutlu AY. Comparison of Hilbert Vibration Decomposition with Empirical Mode Decomposition for Classifying Epileptic Seizures. In: *Conference Record - Asilomar Conference on Signals, Systems and Computers*. 2019. p. 357–62.
4. Gökçen A, Büyükçakır B, Şahin S. Implementation and application of logistic map on field programmable analog array board. In: *26th IEEE Signal*

Processing and Communications Applications Conference, SIU 2018. 2018. p. 1–4.

5. Buyukcakil B, Mutlu AY. Epileptic seizure classification with multivariate empirical mode decomposition and hilbert vibration decomposition. In: 27th Signal Processing and Communications Applications Conference, SIU 2019. Institute of Electrical and Electronics Engineers Inc.; 2019.
6. Buyukcakil B, Elmaz F, Sahin S, Aydin L. Stochastic optimization of PID parameters for twin rotor system with multiple nonlinear regression. In: 2018 6th International Conference on Control Engineering and Information Technology, CEIT 2018. Institute of Electrical and Electronics Engineers Inc.; 2018

