SOME APPLICATIONS OF MACHINE LEARNING FOR BIOMEDICAL SIGNAL PROCESSING

M.Sc. Thesis

By

NOURHEVINUO VICTORIA ANGAMI

DISCIPLINE OF MATHEMATICS

INDIAN INSTITUTE OF TECHNOLOGY INDORE

MAY 2018

SOME APPLICATIONS OF MACHINE LEARNING FOR BIOMEDICAL SIGNAL PROCESSING

A THESIS

Submitted in partial fulfillment of the requirements

for the award of the degree

of

Master of Science

by

NOURHEVINUO VICTORIA ANGAMI

DISCIPLINE OF MATHEMATICS

INDIAN INSTITUTE OF TECHNOLOGY INDORE

MAY 2018

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled in the partial fulfillment of the requirements for the award of the degree of **MASTER OF SCIENCE** and submitted in the **DISCIPLINE OF MATHEMATICS, INDIAN INSTITUTE OF TECHNOLOGY INDORE**, is an authentic record of my own work carried out during the time period from July, 2016 to May, 2018 under the supervision of **Dr. M. TANVEER**, Ramanujan Fellow and Assistant Professor, Discipline of Mathematics, IIT Indore.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

Signature of the student with date

(Nourhevinuo Victoria Angami)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Signature of Thesis Supervisor with date

(Dr. M. Tanveer)

Nourhevinuo Victoria Angami has successfully given her M.Sc. Oral Examination held on 18 May, 2018.

Signature of Thesis Supervisor	Convener, DPGC
Date:	Date:
Sign. of PSPC Member #1	Sign. of PSPC Member #2
Date:	Date:

Acknowledgements

I would like to express my sincere gratitude to my thesis supervisor Dr. M. Tanveer for his patience, motivation and enthusiasm for helping me to complete this thesis.

I would also like to express my profound gratitude to Prof. Ram Bilas Pachori for his guidance and immense knowledge in the field of signal processing and other areas required in this thesis.

Their collaboration has given me the opportunity to work on this diverse exciting thesis and an insight to research atmosphere.

Sincere gratitude is also extended to the PSPC members, Dr. Md. Aquil Khan and Prof. Ram Bilas Pachori, for their valuable remarks and kind suggestions.

I would also like to thank Dr. Sk. Safique Ahmad, Head-Discipline of Mathematics, for ensuring that the research lab is facilitated with the state-of-the-art computational facilities to carry out research with ease.

Last but not the least, I would like to thank my lab-mates for their encouragement and support and those who have helped me directly or indirectly with this thesis.

Nourhevinuo Victoria Angami
1603141007
Discipline of Mathematics
Indian Institute of Technology Indore

Dedication

Dedicated to my mentor and my guide Dr. M. Tanveer.

Abstract

Flexible analytic wavelet transform (FAWT) is suitable for the study of oscillatory signals like electroencephalogram (EEG) signals with versatile features such as shift in-variance, tunable oscillatory properties and flexible time-frequency domain covering. In this thesis, we propose two automated methods for the classification of epileptic EEG signals using FAWT for decomposition of the EEG signals into sub-bands and suitable features were extracted. The obtained features are given as input to twin support vector machine (TSVM), least squares TSVM (LS-TSVM) and robust energy-based least squares twin support vector machines (RELS-TSVM) for classification. The proposed methods have been implemented on publicly available Bonn University EEG database [1] and the accuracy of RELS-TSVM was found to be better as compared to TSVM and LS-TSVM and is comparable to other existing methods with a maximum accuracy of 100% for the classification of seizure and non-seizure EEG signals and 98.33% for the classification of seizure and seizure-free EEG signals.

Contents

List of Figures **vi**

List of Tables **vii**

Abbreviations **viii**

1 Introduction **1**
1.1 Wavelets . . . 2
1.2 Flexible analytic wavelet transform (FAWT) . . . 2
1.3 Feature extraction . . . 4
1.3.1 Entropy-based features . . . 5
1.3.2 Hjorth parameters . . . 6
1.4 Classifiers . . . 7
1.4.1 Support vector machine (SVM) . . . 7
1.4.1.1 Linear SVM . . . 8
1.4.1.2 Nonlinear SVM . . . 9
1.4.2 Twin support vector machine (TSVM) . . . 10
1.4.2.1 Linear TSVM . . . 10
1.4.2.2 Nonlinear TSVM . . . 12
1.4.3 Least squares twin support vector machine (LS-TSVM) . . . 13
1.4.3.1 Linear LS-TSVM . . . 13
1.4.3.2 Nonlinear LS-TSVM . . . 14
1.4.4 Robust energy-based LS-TSVM (RELS-TSVM) . . . 15
1.4.4.1 Linear RELS-TSVM . . . 15
1.4.4.2 Nonlinear RELS-TSVM . . . 17

2 Literature review **19**

3 Proposed work and numerical experiments **22**
3.1 Discussion . . . 26

4 Conclusion and future work **30**

Bibliography **31**

List of Figures

1.1 Frequency response of FAWT filters [2]. 3

3.1 EEG signals for various subsets. 23

3.2 Proposed approach for the classification of EEG signals 23

List of Tables

3.1 p-values of sub-bands (Entropy-based features) 24

3.2 p-values of sub-bands (Hjorth parameters) 25

3.3 Accuracy % of the first proposed method (Entropy-based features) . . 26

3.4 Accuracy % of the second proposed method (Hjorth parameters) . . . 26

3.5 Comparison of existing methods for classification of seizure and non-seizure EEG signals 27

3.6 Comparison of existing methods for classification of seizure and seizure-free EEG signals 28

Abbreviations

EEG	**E**lectroencephalogram
FAWT	**F**lexible **A**nalytic **W**avelet **T**ransform
SURE	**S**tein's **U**nbiased **R**isk **E**stimator
SVM	**S**upport **V**ector **M**achines
TSVM	**T**win **S**upport **V**ector **M**achines
LS-TSVM	**L**east **S**quares **T**win **S**upport **V**ector **M**achines
RELS-TSVM	**R**obust **E**nergy-based **L**east **S**quares **T**win **S**upport **V**ector **M**achines
K-W	**K**ruskal-**W**allis

Chapter 1

Introduction

Epilepsy is a neurological disorder affecting almost 1% of the world population and can occur in any age group [3]. Epilepsy is noncontagious, treatable and preventable if the cause is known [3]. A person who has epilepsy may face social and economic discrimination and can fail to obtain certain benefits [3]. Epilepsy happens when the brain neurons undergo uncontrolled electrical impulses in the brain resulting in uncontrolled seizures in the person affected [4]. There are several methods of acquiring electrical signals from the brain. Detecting brain activity can be invasive, partially invasive, or non-invasive depending on the application and on the nature of the electrodes used to acquire the signals. Electroencephalography (EEG) is a standard non-invasive method of obtaining brain signals. The ease of use, portability, cost-efficient, and high temporal resolution makes EEG the most widely used technique for acquiring brain signals to monitor and diagnose epilepsy and other brain conditions [5]. Detection and interpretation of epileptic seizure in EEG signals perform manually is prone to error and requires experts for examination. Hence, automated methods are necessary for reliable and accurate detection of epileptic activity in the EEG signals.

In this chapter, we discuss flexible analytic wavelet transform (FAWT), entropy-based features, Hjorth parameters and various classification techniques.

1.1 Wavelets

A wave is defined as an oscillating function of time, such as a sinusoid. A wavelet is a small wave, which has finite energy concentrated around a point. A signal or function $f(t)$ can be often analyzed better if it is expressed as a linear combination by,

$$f(t) = \sum_j \sum_k c_{j,k} \psi_{j,k}(t), \quad j, k \in \mathbb{Z}$$

where $c_{j,k}$ are the real-valued expansion co-efficients, $\psi_{j,k}(t)$ are a set of real-valued functions of t.

The function space $L^2(\mathbb{R})$ is the space of square integrable functions $f(t)$. Mathematically defined as,

$$L^2(\mathbb{R}) = \Big\{ f \colon \mathbb{R} \to \mathbb{C} \colon \int_{\mathbb{R}} |f(t)|^2 dt < +\infty \Big\}.$$

A function $\psi \in L^2(\mathbb{R})$ is said to be wavelet in $L^2(\mathbb{R})$ if the system

$$\{\psi_{j,k}(\cdot) = 2^{j/2}\, \psi(2^j \cdot - k) \,|\, j, k \in \mathbb{Z}\}$$

forms an orthonormal basis for $L^2(\mathbb{R})$.

The wavelet system is a set of building blocks to represent a signal and the wavelet expansion gives a good time-frequency localization of the system [6].

1.2 Flexible analytic wavelet transform (FAWT)

The iterated filter banks of FAWT [7] is constructed by one lowpass channel and two highpass channels, where one of the highpass channel analyzes ‘positive frequencies’ and the other ‘negative frequencies’. This separation of negative and positive frequencies obtains Hilbert transform pairs of wavelet bases. Let $f(\theta)$ be a polynomial [8],

$$f(\theta) = \frac{1}{2}(1 + \cos(\theta))\sqrt{2 - \cos(\theta)}, \tag{1.1}$$

which satisfies the perfect construction condition [7]:

$$|f(\theta)|^2 + |f(\theta - \pi)|^2 = 1,\ \theta \in [0, \pi]. \tag{1.2}$$

Daubechies' orthonormal wavelet filters with 2 vanishing moments is used to construct $f(\theta)$ [2]. Let $H(w)$ and $G(w)$ be the frequency response of the lowpass fil-

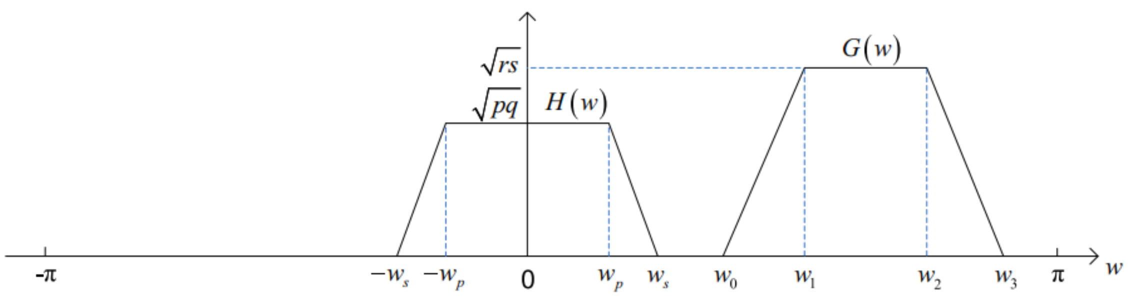

FIGURE 1.1: Frequency response of FAWT filters [2].

ter and highpass filters respectively for the underlying perfect reconstruction filter banks. The attributes of the filters can be controlled by the parameters p, q, r and s. The positive constants β and ε and the parameters p, q, r and s should satisfy the below constraints [2]:

$$1 - \frac{p}{q} \leq \beta \leq \frac{r}{s},\ \varepsilon \leq (\frac{p - q + \beta q}{p + q})\pi. \tag{1.3}$$

Figure 1.1 can be mathematically written as [7]:

$$H(w) = \begin{cases} \sqrt{pq}, & |w| < w_p, \\ \sqrt{pq} f\left[\frac{(w - w_p)}{(w_s - w_p)}\right], & w_p \leq w \leq w_s, \\ \sqrt{pq} f\left[\frac{(\pi - w + w_p)}{(w_s - w_p)}\right], & -w_s \leq w \leq -w_p, \\ 0, & |w| > w_s, \end{cases} \tag{1.4}$$

where

$$w_p = \frac{1-\beta}{p}\pi + \frac{\varepsilon}{p},\ w_s = \frac{\pi}{q}.$$

$$G(w) = \begin{cases} \sqrt{rs}f\left[\frac{(\pi-w+w_0)}{(w_1-w_0)}\right], & w_0 \le w \le w_1, \\ \sqrt{rs}, & w_1 < w < w_2, \\ \sqrt{rs}f\left[\frac{(w+w_2)}{(w_3-w_2)},\right. & w_2 \le w \le w_3, \\ 0, & w \in [-\pi, w_0] \cup [w_3, \pi], \end{cases} \tag{1.5}$$

where

$$w_0 = \frac{1-\beta}{r}\pi + \frac{\varepsilon}{r}, w_1 = \frac{p}{qr}\pi, w_2 = \frac{\pi}{r} - \frac{\varepsilon}{r}, w_3 = \frac{\pi}{r} + \frac{\varepsilon}{r}.$$

The parameters p, q, r, s and β with desired Q-factor Q, dilation factor d and redundancy factor R give flexibility to design wavelets where the accurate extraction of the impulse intervals can be extracted. The factors Q, d and R are related as given [2],

$$Q \approx \frac{2-\beta}{\beta},\ d \approx p/q,\ \beta \le 1,\ R \approx \frac{r/s}{1-d},\ R > \frac{\beta}{1-d}. \tag{1.6}$$

1.3 Feature extraction

Extraction of suitable features is the most important part for better classification purposes. Features are attributes or characteristics that help us to study the complex EEG signals in a better way. In this section, the features selected for this thesis are briefly described.

1.3.1 Entropy-based features

It is observed that the analysis of EEG signals using a combination of joint time-frequency tools and nonlinear features yields an excellent performance [9]. In the seizure detection, the performance of wavelets used with nonlinear features has been found promising. Since EEG signals are highly complex and nonlinear [21], we have chosen nonlinear features viz., entropy-based features. The entropies we have chosen are log energy entropy [10], SURE entropy [11] and Shannon entropy [12] and mathematically defined as follows:

- The computation of log energy entropy [13] is performed to evaluate the degree of complexities in EEG signals. It is given by [10],

$$E_{LogEn} = \sum_{i=1}^{N} \log(y_i^2),$$

  where y_i is the i^{th} sample of the signal and N represents the length of the signal.

- SURE entropy is based on Stein's unbiased risk estimate (SURE) [11,13] which is a common measuring tool for quantifying properties related to information for an accurate representation of the signal defined as [11],

$$E_{SURE} = N - \#\{i : |y_i| \leq \varepsilon\} + \sum_{i=1}^{N} \min(y_i^2, \varepsilon^2),$$

  where y_i is the i^{th} sample of the signal, $\#$ is the number of times $i : |y_i| \leq \varepsilon$, $\varepsilon > 0$ and N represents the length of the signal.

- Shannon entropy is an efficient tool to measure uncertain information [14]. The formula of the Shannon entropy can be given as [12],

$$H(X) = -\sum_{i=1}^{N} y_i^2 \log_2 y_i^2,$$

where y_i is the i^{th} sample of the signal and N represents the length of the signal.

1.3.2 Hjorth parameters

Hjorth parameters are commonly used features for classification of EEG signals introduced by Hjorth in 1970 [15]. It has three parameters to study an EEG signal viz., activity, mobility and complexity.

- Activity is the measure of the average power of the signal (variance of the signal). Mathematically, it can be expressed as [15]:

$$Activity = \sum_{i=1}^{N} \frac{(y(i) - \mu)}{N}.$$

- Mobility is an estimate of the mean frequency. It can be defined as follows [15]:

$$Mobility = \sqrt{\frac{var(y')}{var(y)}}.$$

- Complexity is an estimate of the bandwidth of the signal. It can be defined as follows [15]:

$$Complexity = \frac{Mobility(y')}{Mobility(y)},$$

where $y(t)$ is the signal, $y' = \frac{dy(t)}{dt}$ is the first derivative of the signal, μ is the mean of the signal and N is the number of samples.

As the Hjorth parameters can be calculated easily, it is capable for real-time application.

1.4 Classifiers

In machine learning, one of the primary tasks is classification. Support vector machines introduced by Vapnik [16] is a machine learning technique which is based on statistical learning theory. Due to the good generalization performance, it has been widely used for classification problems. The objective of SVM is to find the optimal hyperplane with maximum margin between the data points of different classes. The data points which lie on the supporting hyperplane are called support vectors. The optimization problem of SVM is convex in nature which gives a global and unique solution. In this section, we briefly discuss support vector machines (SVMs) [16] and its variants [17–19] used for classification purposes. In this thesis we have worked on two class classification of EEG signals, hence only the binary classification algorithms of SVM and its variants (TSVM [17], LS-TSVM [18] and RELS-TSVM [19]) are briefly discussed.

1.4.1 Support vector machine (SVM)

Consider the training set $T = \{(x_1, y_1), (x_2, y_2), \ldots, (x_\ell, y_\ell)\}$, where $x_i \in \mathbb{R}^n$ are inputs and $y_i \in Y = \{-1, 1\}$ are corresponding outputs for $i = 1, 2, \ldots, \ell$. If there exist $w \in \mathbb{R}^n$, $b \in \mathbb{R}$ and a positive number ϵ such that for any subscript i with $y_i = 1$, we have $w^T x_i + b \geq \epsilon$, and for any subscript i with $y_i = -1$, we have $w^T x_i + b \leq -\epsilon$, we say the training set and corresponding classification problem is linearly separable. If the training data are not linearly separable then to enhance the linear separability, we introduce $\phi(\cdot)$ a nonlinear function that maps the original input space into higher-dimension Hilbert space $\mathscr{H}$ called the feature space.

1.4.1.1 Linear SVM

The optimization problem for the linear case of SVM can be formulated as [16]:

$$\begin{aligned} \min_{w,\,b,\,\xi} \quad & \frac{1}{2}\|w\|^2 + a\sum_{i=1}^{\ell} \xi_i \\ s.t. \quad & y_i(w^T x_i + b) \geq 1 - \xi_i,\ \xi_i \geq 0,\ i = 1, 2, \ldots, \ell, \end{aligned} \tag{1.7}$$

where $\xi = (\xi_1, \xi_2, \ldots, \xi_\ell)$ is slack variable and a is the penalty parameter. To solve the above optimization problem we consider its dual formulation by introducing the Lagrange function,

$$L(w,\, b,\, \xi,\, \alpha,\, \beta) = \frac{1}{2}\|w\|^2 + a\sum_{i=1}^{\ell}\xi_i - \sum_{i=1}^{\ell}\beta_i\xi_i - \sum_{i=1}^{\ell}\alpha_i(y_i(w^T x_i + b) + \xi_i - 1), \tag{1.8}$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_\ell)^T \geq 0$ and $\beta = (\beta_1, \beta_2, \ldots, \beta_\ell)^T \geq 0$ are the Lagrangian multipliers.

Using the Karush-Kuhn-Tucker (KKT) [20] conditions, we get the dual formulation of (1.7) as,

$$\begin{aligned} \max_{\alpha} \quad & -\frac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}\alpha_i\alpha_j y_i y_j x_i^T x_j + \sum_{i=1}^{\ell}\alpha_i \\ s.t. \quad & \sum_{i=1}^{\ell}\alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq a. \end{aligned} \tag{1.9}$$

The following decision function $\mathscr{D}(x)$ assigns an unknown data point $x \in \mathbb{R}^n$ to the class $\{+1, -1\}$ accordingly as,

$$\mathscr{D}(x) = \operatorname{sgn}\Big(\sum_{i=1}^{\ell} y_i\alpha_i(x_i \cdot x) + b\Big). \tag{1.10}$$

1.4.1.2 Nonlinear SVM

Suppose the training points are not linearly separable then we introduce the kernel function and the optimization problem is formulated as [16],

$$\begin{aligned} \min_{w,\,b,\,\xi} \quad & \frac{1}{2}\|w\|^2 + a\sum_{i=1}^{\ell}\xi_i \\ s.t. \quad & y_i(w^T\phi(x_i)+b) \geq 1-\xi_i,\ \xi_i \geq 0,\ i=1,2,\ldots,\ell, \end{aligned} \tag{1.11}$$

where $\xi = (\xi_1, \xi_2, \ldots, \xi_\ell)$ is the slack variable, a is the penalty parameter and $\phi(\cdot)$ is a non-linear function that maps input space into feature space. To solve the above optimization problem we consider its dual formulation by introducing the Lagrange function,

$$L(w,\, b,\, \xi,\, \alpha,\, \beta) = \frac{1}{2}\|w\|^2 + a\sum_{i=1}^{\ell}\xi_i - \sum_{i=1}^{\ell}\beta_i\xi_i - \sum_{i=1}^{\ell}\alpha_i(y_i(w^T\phi(x_i)+b)+\xi_i-1), \tag{1.12}$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_\ell)^T \geq 0$ and $\beta = (\beta_1, \beta_2, \ldots, \beta_\ell)^T \geq 0$ are the Lagrangian multipliers. After solving (1.12) as similar to the linear case, the dual formulation of (1.11) is given by,

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^{\ell}\alpha_i - \frac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}\alpha_i\alpha_j y_i y_j \phi(x_i)^T\phi(x_j) \\ s.t. \quad & \sum_{i=1}^{\ell}\alpha_i y_i = 0,\ 0 \leq \alpha_i \leq a. \end{aligned} \tag{1.13}$$

The following decision function $\mathscr{D}(x)$ assigns an unknown data point $x \in \mathbb{R}^n$ to the class $\{+1, -1\}$ accordingly as,

$$\mathscr{D}(x) = \operatorname{sgn}\Big(\sum_{i=1}^{\ell} y_i\alpha_i\mathscr{K}(x_i, x) + b\Big) \tag{1.14}$$

where $\mathscr{K}(x_i, x)$ is the kernel function.

1.4.2 Twin support vector machine (TSVM)

Twin support vector machine (TSVM) [17] seeks two non-parallel hyperplanes instead of a single hyperplane as in the case of conventional SVM by solving two quadratic programming problems (QPPs) where the whole data points are divided to formulate the objective function and the constraints. In this manner, the run time of TSVM is much faster as compared to the computational time of conventional SVM [17].

Consider the binary classification problem where data points belonging to class 1 is represented by the matrix $\mathcal{A}$ and the data points belonging to class -1 is represented by the matrix $\mathcal{B}$ with ℓ_1 and ℓ_2 number of data points respectively in the n-dimensional real space $\mathbb{R}^n$.

1.4.2.1 Linear TSVM

Linear TSVM seeks a pair of non parallel hyperplanes,

$$w_+^T x + b_+ = 0 \quad \text{and} \quad w_-^T x + b_- = 0, \tag{1.15}$$

where $x \in \mathbb{R}^n$, $w_+ \in \mathbb{R}^n$, $w_- \in \mathbb{R}^n$, $b_+ \in \mathbb{R}$ and $b_- \in \mathbb{R}$. Each hyperplane is close to the data points of one class and far from the data points of the other class. The two optimization problems for the linear case can be expressed as follows [17]:

$$\begin{aligned} \min_{w_+, b_+, \xi_1} \quad & \frac{1}{2}\|\mathcal{A}w_+ + e_1 b_+\|^2 + a_1 e_2^T \xi_1 \\ s.t. \quad & -(\mathcal{B}w_+ + e_2 b_+) + \xi_1 \geq e_2, \quad \xi_1 \geq 0 \end{aligned} \tag{1.16}$$

and

$$\begin{aligned} \min_{w_-, b_-, \xi_2} \quad & \frac{1}{2}\|\mathcal{B}w_- + e_2 b_-\|^2 + a_2 e_1^T \xi_2 \\ s.t. \quad & (\mathcal{A}w_- + e_1 b_-) + \xi_2 \geq e_1, \quad \xi_2 \geq 0, \end{aligned} \tag{1.17}$$

a_1, a_2 are positive parameters, ξ_1 and ξ_2 are the slack variables and e_1, e_2 are vectors of ones of appropriate dimensions. Similarly, we introduce the Lagrangian multipliers α and β and get the dual formulation of TSVM as,

$$\begin{aligned} \max_{\alpha} \quad & e_2^T \alpha - \frac{1}{2} \alpha^T G (H^T H)^{-1} G^T \alpha \\ s.t. \quad & 0 \leq \alpha \leq a_1 \end{aligned} \tag{1.18}$$

and

$$\begin{aligned} \max_{\beta} \quad & e_1^T \beta - \frac{1}{2} \beta^T H (G^T G)^{-1} H^T \beta \\ s.t. \quad & 0 \leq \beta \leq a_2, \end{aligned} \tag{1.19}$$

where $G = [\mathcal{B} \quad e_2]$; $H = [\mathcal{A} \quad e_1]$; $\alpha \in \mathbb{R}^{\ell_2}$ and $\beta \in \mathbb{R}^{\ell_1}$ are Lagrange multipliers. On solving (1.18) and (1.19) we get,

$$\begin{bmatrix} w_+ \\ b_+ \end{bmatrix} = -(H^T H + \delta I)^{-1} G^T \alpha \tag{1.20}$$

and

$$\begin{bmatrix} w_- \\ b_- \end{bmatrix} = (G^T G + \delta I)^{-1} H^T \beta, \tag{1.21}$$

where I is an identity matrix of appropriate dimension and δ is a very small positive scalar introduced in order to deal with the case when $(H^T H)$ or $(G^T G)$ is singular and avoid the possible ill-conditioning of the matrices in finding the inverse.

A new point $x \in \mathbb{R}^n$ is assigned to class $\{+1, -1\}$ depending on which of the two hyperplanes (1.15) is closer to the decision function $\mathscr{D}(x)$,

$$\mathscr{D}(x) = \text{ sgn } \Big(\frac{w_+^T x + b_+}{\|w_+\|} + \frac{w_-^T x + b_-}{\|w_-\|} \Big). \tag{1.22}$$

1.4.2.2 Nonlinear TSVM

The nonlinear problem can be formulated by considering the following kernel generated surfaces:

$$\mathscr{K}(x^T, D^T)w_+ + b_+ = 0 \;\text{ and }\; \mathscr{K}(x^T, D^T)w_- + b_- = 0, \tag{1.23}$$

where $D = [\mathcal{A}; \mathcal{B}]$ and $\mathscr{K}(\cdot, \cdot)$ is an arbitrary kernel function. The primal problem of nonlinear TSVM can be expressed as follows [17]:

$$\begin{aligned} \min_{w_+, b_+, \xi_1} \quad & \frac{1}{2}\|\mathscr{K}(\mathcal{A}, D^T)w_+ + e_1 b_+\|^2 + a_1 e_2^T \xi_1 \\ s.t. \quad & -(\mathscr{K}(\mathcal{B}, D^T)w_+ + e_2 b_+) + \xi_1 \geq e_2, \quad \xi_1 \geq 0 \end{aligned} \tag{1.24}$$

and

$$\begin{aligned} \min_{w_-, b_-, \xi_2} \quad & \frac{1}{2}\|\mathscr{K}(\mathcal{B}, D^T)w_- + e_2 b_-\|^2 + a_2 e_1^T \xi_2 \\ s.t. \quad & (\mathscr{K}(\mathcal{A}, D^T)w_- + e_1 b_-) + \xi_2 \geq e_1, \quad \xi_2 \geq 0. \end{aligned} \tag{1.25}$$

By introducing the Lagrangian multipliers α and β, we can derive their dual problems as follows:

$$\begin{aligned} \max_{\alpha} \quad & e_2^T \alpha - \frac{1}{2}\alpha^T N(M^T M)^{-1} N^T \alpha \\ s.t \quad & 0 \leq \alpha \leq a_1 \end{aligned} \tag{1.26}$$

and

$$\begin{aligned} \max_{\beta} \quad & e_1^T \beta - \frac{1}{2}\beta^T M(N^T N)^{-1} M^T \beta \\ s.t. \quad & 0 \leq \beta \leq a_2, \end{aligned} \tag{1.27}$$

where $M = [\mathscr{K}(\mathcal{A}, D^T) \quad e_1]; N = [\mathscr{K}(\mathcal{B}, D^T) \quad e_2]$; ξ_1, ξ_2 are slack variables and $e_i (i = 1, 2)$ is the vectors of ones of appropriate dimensions. After solving (1.26) and (1.27) we get,

$$\begin{bmatrix} w_+ \\ b_+ \end{bmatrix} = -[M^T M]^{-1} N^T \alpha \tag{1.28}$$

and

$$\begin{bmatrix} w_- \\ b_- \end{bmatrix} = [N^T N]^{-1} M^T \beta. \tag{1.29}$$

A new point $x \in \mathbb{R}^n$ is assigned to class $\{+1, -1\}$ depending on which of the two hyperplanes (1.23) is closer to the decision function $\mathscr{D}(x)$,

$$\mathscr{D}(x) = \operatorname{sgn}\Big(\frac{\mathscr{K}(x^T, D^T)w_+ + e_1 b_+}{\|w_+\|} + \frac{\mathscr{K}(x^T, D^T)w_- + e_2 b_-}{\|w_-\|}\Big). \tag{1.30}$$

1.4.3 Least squares twin support vector machine (LS-TSVM)

The idea of TSVM is extended to formulate least squares TSVM (LS-TSVM) [18] algorithm where the inequality constraints in TSVM is replaced by equality constraints by the addition of squares of 2-norm of the slack variables. This makes the algorithm simple and fast since the objective function reduces to a linear system of equations.

1.4.3.1 Linear LS-TSVM

The optimization problem for the linear LS-TSVM can be expressed as [18]:

$$\begin{aligned} \min_{w_+, b_+, \xi_1} \quad & \frac{1}{2}\|\mathcal{A}w_+ + e_1 b_+\|^2 + \frac{a_1}{2}\|\xi_1\|^2 \\ s.t. \quad & -(\mathcal{B}w_+ + e_2 b_+) + \xi_1 = e_2 \end{aligned} \tag{1.31}$$

and

$$\begin{aligned} \min_{w_-, b_-, \xi_2} \quad & \frac{1}{2}\|\mathcal{B}w_- + e_2 b_-\|^2 + \frac{a_2}{2}\|\xi_2\|^2 \\ s.t. \quad & (\mathcal{A}w_- + e_1 b_-) + \xi_2 = e_1. \end{aligned} \tag{1.32}$$

The two hyperplanes are obtained by solving the two system of linear equations (1.31) and (1.32) by replacing the values of ξ_1 and ξ_2 in their respective objective functions and we get,

$$\begin{bmatrix} w_+ \\ b_+ \end{bmatrix} = -\Big(\frac{1}{a_1}H^T H + G^T G\Big)^{-1} G^T e_1, \tag{1.33}$$

$$\begin{bmatrix} w_- \\ b_- \end{bmatrix} = \Big(\frac{1}{a_2}G^T G + H^T H\Big)^{-1} H^T e_2, \tag{1.34}$$

where a_1 and a_2 are positive penalty parameters, $G = [\mathcal{B} \quad e_2]$ and $H = [\mathcal{A} \quad e_1]$.

A new point $x \in \mathbb{R}^n$ is assigned to class $\{+1, -1\}$ depending on which of the two hyperplanes is closer to $\mathscr{D}(x_i)$,

$$\mathscr{D}(x) = \ \mathrm{sgn}\ \Big(\frac{w_+^T x + b_+}{\|w_+\|} + \frac{w_-^T x + b_-}{\|w_-\|}\Big). \tag{1.35}$$

1.4.3.2 Nonlinear LS-TSVM

Following the same idea as nonlinear TSVM, nonlinear LSTSVM can be formulated by considering the following kernel generated surfaces:

$$\mathscr{K}(x^T, D^T)w_+ + b_+ = 0 \quad \text{and} \quad \mathscr{K}(x^T, D^T)w_- + b_- = 0. \tag{1.36}$$

The primal QPPs of nonlinear TSVM can be modified in the same way with 2-norm of slack variables and inequality constraints replaced by equality constraints [18].

$$\begin{aligned} \min_{w_+, b_+, \xi_1} \quad & \frac{1}{2}\|\mathscr{K}(\mathcal{A}, D^T)w_+ + e_1 b_+\|^2 + \frac{a_1}{2}\|\xi_1\|^2 \\ s.t. \quad & -(\mathscr{K}(\mathcal{B}, D^T)w_+ + e_1 b_+) + \xi_1 = e_1, \end{aligned} \tag{1.37}$$

and

$$\begin{aligned} \min_{w_-, b_-, \xi_2} \quad & \frac{1}{2}\|\mathscr{K}(\mathcal{B}, D^T)w_- + e_2 b_-\|^2 + \frac{a_2}{2}\|\xi_2\|^2 \\ s.t. \quad & (\mathscr{K}(\mathcal{A}, D^T)w_- + e_2 b_-) + \xi_2 = e_2. \end{aligned} \tag{1.38}$$

Like the linear case, the solution of QPPs (1.37) and (1.38) can be derived to be,

$$\begin{bmatrix} w_+ \\ b_+ \end{bmatrix} = -(N^T N + \frac{1}{a_1} M^T M)^{-1} N^T e_1 \tag{1.39}$$

and

$$\begin{bmatrix} w_- \\ b_- \end{bmatrix} = (M^T M + \frac{1}{a_2} N^T N)^{-1} M^T e_2, \tag{1.40}$$

where $M = [\mathscr{K}(\mathcal{A}, D^T) \quad e_2]$ and $N = [\mathscr{K}(\mathcal{B}, D^T) \quad e_1]$.

A new point $x \in \mathbb{R}^n$ is assigned to class $\{+1, -1\}$ depending on which of the two hyperplanes distance $|\mathscr{K}(x^T, D^T)w_+ + e_1 b_+|$ or $|\mathscr{K}(x^T, D^T)w_- + e_2 b_-|$ is minimum.

1.4.4 Robust energy-based LS-TSVM (RELS-TSVM)

The algorithm of RELS-TSVM [19] uses positive definite matrix in the formulation to maximize the margin by addition of regularization term to each of the objective constraints and by considering different energy parameters for each hyperplane unlike the constraints of LS-TSVM which requires the hyperplane to be exactly 1 distance away from the data points of the other class.

1.4.4.1 Linear RELS-TSVM

In linear case, the objective functions of RELS-TSVM are expressed as follows [19]:

$$\begin{aligned} \min_{w_+, b_+, \xi_1} \quad & \frac{1}{2}\|\mathcal{A}w_+ + eb_+\|^2 + \frac{a_1}{2}\xi_1^T \xi_1 + \frac{a_3}{2}\left\| \begin{bmatrix} w_+ \\ b_+ \end{bmatrix} \right\|^2 \\ s.t. \quad & -(\mathcal{B}w_+ + eb_+) + \xi_1 = E_1 \end{aligned} \tag{1.41}$$

and

$$\begin{aligned} \min_{w_-,b_-,\xi_2} \quad & \frac{1}{2}\left\|\mathcal{B}w_- + eb_-\right\|^2 + \frac{a_2}{2}\xi_2^T\xi_2 + \frac{a_4}{2}\left\|\begin{bmatrix} w_- \\ b_- \end{bmatrix}\right\|^2 \\ s.t. \quad & (\mathcal{A}w_- + eb_-) + \xi_2 = E_2, \end{aligned} \tag{1.42}$$

where e is a vector of ones of appropriate dimensions; a_1, a_2, a_3, and a_4 are positive parameters; E_1 and E_2 represents energy parameters of the hyperplanes; w_+, $w_- \in \mathbb{R}^n$ and b_+, $b_- \in \mathbb{R}$.

The solution of the QPP (1.41) is obtained as follows:

$$\begin{bmatrix} w_+ \\ b_+ \end{bmatrix} = -\left(a_1G^TG + H^TH + a_3I\right)^{-1} a_1\mathcal{B}^TE_1 \tag{1.43}$$

and the solution of QPP (1.42) is obtained as follows:

$$\begin{bmatrix} w_- \\ b_- \end{bmatrix} = \left(a_2H^TH + G^TG + a_4I\right)^{-1} a_2\mathcal{A}^TE_2, \tag{1.44}$$

where I is an identity matrix of appropriate dimension, $H = [\mathcal{A} \quad e]$ and $G = [\mathcal{B} \quad e]$.
The label of an unknown data point $x \in \mathbb{R}^n$ is assigned to the class $\{+1, -1\}$, depending on the following decision function $\mathscr{D}(x)$,

$$\mathscr{D}(x) = \begin{cases} +1, & \text{if } \left|\frac{x^Tw_+ + eb_+}{x^Tw_- + eb_-}\right| \leq 1 \\ -1, & \text{if } \left|\frac{x^Tw_+ + eb_+}{x^Tw_- + eb_-}\right| > 1 \end{cases}$$

where $|.|$ is the absolute value.

1.4.4.2 Nonlinear RELS-TSVM

For the nonlinear RELS-TSVM, the following kernel-generated surfaces are considered [19]:

$$\mathscr{K}(x^T, D^T)w_+ + b_+ = 0 \quad \text{and} \quad \mathscr{K}(x^T, D^T)w_- + b_- = 0, \tag{1.45}$$

where $D = [\mathcal{A}; \mathcal{B}]$ and $\mathscr{K}$ is a suitably chosen kernel.

The two optimization problems for the nonlinear RELS-TSVM can be expressed as [19]:

$$\begin{aligned} \min_{w_+, b_+, \xi_1} \quad & \frac{1}{2}\|\mathscr{K}(\mathcal{A}, D^T)w_+ + eb_+\|^2 + \frac{a_1}{2}\xi_1^T\xi_1 + \frac{a_3}{2}\left\| \begin{bmatrix} w_+ \\ b_+ \end{bmatrix} \right\|^2 \\ s.t. \quad & -(\mathscr{K}(\mathcal{B}, D^T)w_+ + eb_-) + \xi_1 = E_1 \end{aligned} \tag{1.46}$$

and

$$\begin{aligned} \min_{w_-, b_-, \xi_2} \quad & \frac{1}{2}\|\mathscr{K}(\mathcal{B}, D^T)w_- + eb_-\|^2 + \frac{a_2}{2}\xi_2^T\xi_2 + \frac{a_4}{2}\left\| \begin{bmatrix} w_- \\ b_- \end{bmatrix} \right\|^2 \\ s.t. \quad & (\mathscr{K}(\mathcal{A}, D^T)w_- + eb_-) + \xi_2 = E_2. \end{aligned} \tag{1.47}$$

The solution of the QPP (1.46) is obtained as follows:

$$\begin{bmatrix} w_+ \\ b_+ \end{bmatrix} = -(a_1 N^T N + M^T M + a_3 I)^{-1} a_1 N^T E_1 \tag{1.48}$$

and the solution of QPP (1.47) is obtained as follows:

$$\begin{bmatrix} w_- \\ b_- \end{bmatrix} = (a_2 M^T M + N^T N + a_4 I)^{-1} a_2 M^T E_2, \tag{1.49}$$

where I is identity matrix of appropriate dimension, $N = [\mathscr{K}(\mathcal{B}; D^T) \quad e]$ and $M = [\mathscr{K}(\mathcal{A}, D^T) \quad e]$.

In comparison to the algorithm of LS-TSVM, due to the addition of extra regularization term the matrices $(a_1 N^T N + M^T M + a_3 I)$ and $(a_2 M^T M + N^T N + a_4 I)$ are positive definite, which makes the solution of nonlinear RELS-TSVM more stable.

The following decision function $\mathscr{D}(x)$ labels an unknown data point $x \in \mathbb{R}^n$ to the class $\{+1, -1\}$ as,

$$\mathscr{D}(x) = \begin{cases} +1 & \text{if } \left|\frac{\mathscr{K}(x^T, D^T)w_+ + eb_+}{\mathscr{K}(x^T, D^T)w_- + eb_-}\right| \leq 1 \\ \\ -1 & \text{if } \left|\frac{\mathscr{K}(x^T, D^T)w_+ + eb_+}{\mathscr{K}(x^T, D^T)w_- + eb_-}\right| > 1 \end{cases}$$

where $|.|$ is the absolute value.

Chapter 2

Literature review

Initially, EEG signals were studied by Fourier transformation assuming that they were stationary. However, studies show that EEG signals are non-stationary [21] and highly nonlinear [22]. Hence, various time-frequency domain methods have been proposed for the detection of epileptic seizure in EEG signals. Many researchers have proposed methods for the study of EEG signals and the automated detection. Several methods are discussed briefly. Empirical mode decomposition (EMD) is applied for the analysis of EEG signals based on Hilbert transformation [23]. Classification of the epileptic seizure in EEG signals using second-order difference plot of intrinsic mode function (IMF) by artificial neural network (ANN) has been performed [24]. The method of classifying EEG signals based on fractional-order calculus and support vector machine (SVM) with different kernel function has been proposed in [25]. Features from phase space representations (PSRs) of IMFs of EEG signals and employed LS-SVM for classification [26]. A method based on one-dimensional local binary pattern (1D-LBP) features for the classification of seizure and seizure-free EEG signals by nearest neighbour classifier has been performed in [27]. Empirical wavelet transform (EWT) has been explored for the analysis of multivariate signals with six classifiers namely: random forest (RF), functional tree (FT), C4.5, Bayes-net, Naive-Bayes, and K-nearest neighbours (K-NN) in [28]. EEG signal classification using universum support vector machine has been performed in [29].

Recent works are focusing on wavelet transforms (WT) for signal processing. Since signals are ephemeral and non-stationary, the properties of wavelets make it easier to analyze the signals. WT has good time-frequency concentration and multi-resolution analysis capacity is a powerful mathematical tool for signal processing. Flexible analytic wavelet transform (FAWT) was also used to study various bio-medical problems other than EEG signals [30–32] which yielded good results. The methods in [8] and [13] proposes study of EEG signals by FAWT with fractal dimension [8] and entropy-based features [13] with LS-SVM classifier. Many intrinsic limitations in other WTs are enhanced by FAWT [2]. Employing arbitrary and fractional scaling and translation factors, FAWT enjoys flexible TF covering manner of the bases. Hence we are motivated to explore the properties of FAWT for decomposing the EEG signals.

For classification purpose we used variants of support vector machines ((TSVM) [17], (LS-TSVM) [18] and (RELS-TSVM [19])). In the recent years, some variants have emerged based on the idea of constructing non-parallel hyperplanes other than searching for an optimal hyperplane in the convention SVM [17–19, 33–37]. TSVM is one of the powerful classification methods. TSVM seeks two non-parallel proximal hyperplanes such that each hyperplane is closer to one of the two classes and as far as possible from the other one [17]. Experimental results have shown the promising performance over SVM and LS-SVM with lesser training time due to its strong generalization ability. Recently, few studies [18, 19, 33–37] have proposed variants of TSVM. LS-TSVM has been proposed by replacing the convex quadratic programming problem (QPP) with convex linear system leading to very fast computational speed [18]. Different energy for each class has been considered in energy-based LS-TSVM (ELS-TSVM) [34] proposed by Nasiri et al., and can also be applied to unbalanced datasets. To overcome the problems in TSVM, LSTSVM and ELS-TSVM like satisfying only empirical risk-minimization principle and the matrices used in the formulation to be always positive semi-definite, RELS-TSVM algorithm has been proposed by Tanveer et al. [19] in which the margin is maximized with a positive definite matrix formulation. Furthermore, the algorithm for RELS-TSVM does not need any specific optimizer and applies energy parameters to minimize the effect of the noise and outliers which makes the classification not only robust to noise and outliers but also more stable [19]. With all these advantages, in this thesis, we explore the performances of TSVM, LS-TSVM and RELS-TSVM.

Chapter 3

Proposed work and numerical experiments

In this chapter, two automated methods are proposed for the classification of epileptic EEG signals using FAWT for decomposition of the EEG signals into sub-bands and suitable features were extracted. The EEG database has been taken from Bonn University which is publicly available compiled by Andrzejak et al. [1, 38]. The sets Z, O, N, F, S contain 100 EEG segments with each segment have a duration of 23.6 s with 4097 samples, sampled at a sampling rate of 173.61 Hz. The sets Z and O contain EEG recordings of five normal volunteers in a relaxed and awake state with eyes open and eyes closed respectively. The sets N and F are recorded from five patients during seizure free intervals. The EEG signals in dataset N were recorded from the epileptogenic zone while the signals in F were recorded from the opposite hemisphere of the brain. The set S comprises EEG recordings during seizure activity. In total we have 500 EEG segments.

In the first proposed method we have taken (Z, O, N, F) as the non-seizure class and S as seizure class with entropy-based features to classify the seizure and non-seizure EEG signals. The class (Z, O, N, F) versus S is a more generalized way of classification of EEG signals hence in the second proposed method we have restricted

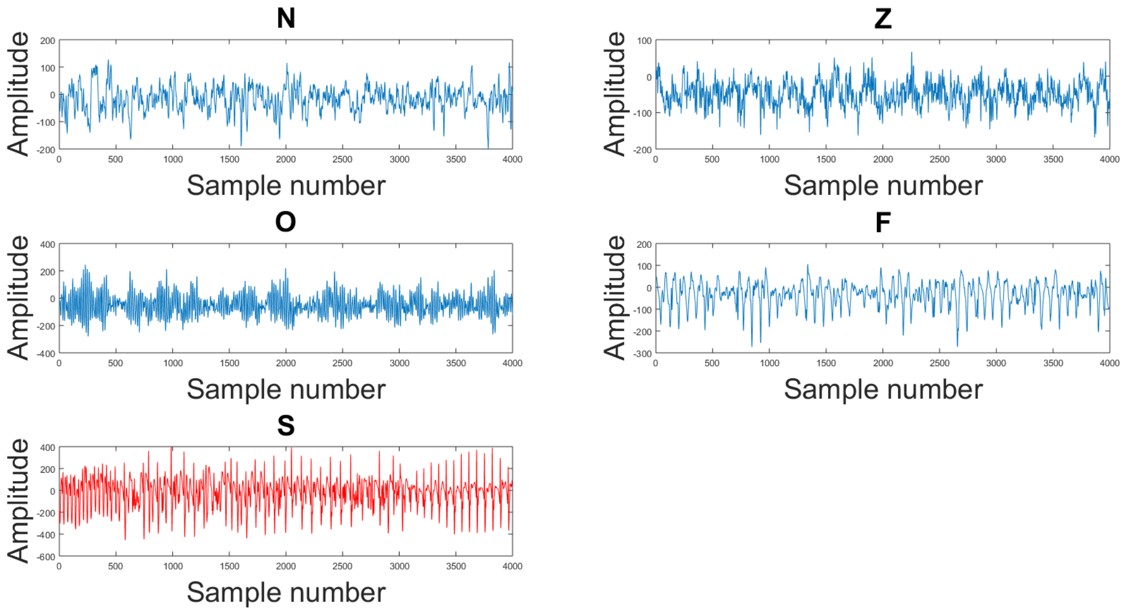

FIGURE 3.1: EEG signals for various subsets.

the sets to (N, F) termed as the seizure-free class versus seizure class (S) and used Hjorth parameters feature for the classification of EEG signals. Although Hjorth parameters are applied in the time domain, they can be interpreted in the frequency domain as well and better accuracy was achieved after decomposition of EEG signals by empirical mode decomposition (EMD) in [5]. The figure below shows the steps involved in the automated detection applied in this thesis.

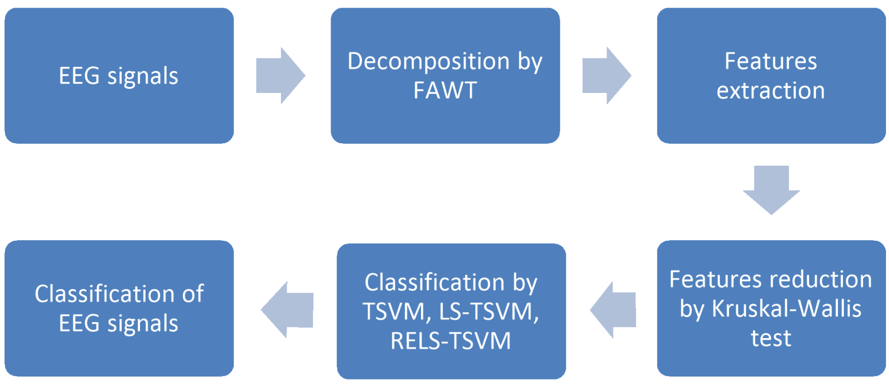

FIGURE 3.2: Proposed approach for the classification of EEG signals

Each of these EEG segments has been decomposed up to $J = 15^{\text{th}}$ level [13] by FAWT which gives $(J + 1)$ sub-bands and one additional sub-band is of the reconstructed signal, i.e., 17 sub-bands were produced for each of the EEG segment. The parameter of FAWT has been set to $\varepsilon = \frac{1}{32}\left(\frac{p-q+\beta q}{p+q}\right)\pi$, $p = 3$, $q = 4$, $r = 1$ and $s = 2$ [2, 13].

After the decomposition of EEG signals into sub-bands, features has been extracted from each of these sub-bands. To reduce the computational time of the classifiers, we have applied Kruskal-Wallis (K-W) [39] test to discriminate the significant sub-bands with p-value ≤ 0.05 to be taken as input for classification of EEG signals. Table 3.1 shows the p-values of the sub-bands with entropy based features tested against seizure class versus non-seizure class. As seen in table the p values are less than $p \leq 0.5$, hence all the sub-bands are considered as features for classification purpose.

TABLE 3.1: p-values of sub-bands (Entropy-based features)

Sub-band	Log energy entropy	Shannon entropy	SURE entropy
1	04.98×10^{-39}	1.07×10^{-80}	8.91×10^{-27}
2	07.78×10^{-06}	7.43×10^{-45}	0.034845084
3	01.43×10^{-98}	6.07×10^{-110}	2.46×10^{-88}
4	01.82×10^{-110}	2.51×10^{-120}	5.36×10^{-99}
5	03.91×10^{-122}	2.67×10^{-126}	1.17×10^{-114}
6	04.51×10^{-121}	4.03×10^{-122}	4.56×10^{-116}
7	43.35×10^{-116}	1.43×10^{-117}	2.12×10^{-113}
8	04.20×10^{-120}	3.16×10^{-120}	5.15×10^{-118}
9	08.81×10^{-129}	3.86×10^{-128}	3.22×10^{-129}
10	05.49×10^{-129}	1.41×10^{-128}	3.04×10^{-127}
11	02.44×10^{-113}	3.25×10^{-115}	2.65×10^{-109}
12	04.94×10^{-107}	4.82×10^{-109}	1.57×10^{-104}
13	01.53×10^{-95}	1.71×10^{-99}	4.88×10^{-89}
14	05.68×10^{-70}	1.16×10^{-74}	1.51×10^{-62}
15	01.48×10^{-52}	1.33×10^{-57}	2.01×10^{-47}
16	01.39×10^{-17}	2.04×10^{-32}	5.11×10^{-10}
17	03.38×10^{-126}	2.63×10^{-128}	4.03×10^{-110}

Similarly, we applied K-W test to discriminate the significant sub-bands tested against seizure class versus seizure-free class. Table 3.2 shows the p-values of the sub-bands with Hjorth parameters features. As seen below, p-values of some of the sub-bands are greater than $p = 0.05$, hence the highlighted sub-bands were not considered for feeding into the classifiers.

TABLE 3.2: p-values of sub-bands (Hjorth parameters)

Sub-band	Activity	Mobility	Complexity
1	1.25×10^{-44}	1.25×10^{-19}	8.57×10^{-05}
2	1.35×10^{-44}	0.909100447	1.29×10^{-08}
3	9.03×10^{-59}	7.17×10^{-44}	0.00614096
4	2.31×10^{-64}	3.86×10^{-27}	2.42×10^{-07}
5	3.02×10^{-65}	9.08×10^{-40}	2.10×10^{-10}
6	2.68×10^{-64}	0.417180712	0.513175327
7	7.49×10^{-66}	2.87×10^{-13}	7.93×10^{-07}
8	2.27×10^{-63}	1.46×10^{-25}	0.43020936
9	1.03×10^{-63}	8.65×10^{-11}	0.524383843
10	1.80×10^{-63}	0.018210469	0.036643259
11	1.03×10^{-51}	1.11×10^{-10}	0.451747071
12	1.52×10^{-46}	0.675482692	0.00047207
13	3.01×10^{-38}	0.021999018	0.474950302
14	4.69×10^{-23}	1.18×10^{-05}	0.315696464
15	1.91×10^{-16}	0.020824924	0.010202237
16	3.40×10^{-13}	6.88×10^{-07}	0.018381077
17	2.08×10^{-63}	1.39×10^{-52}	6.71×10^{-65}

We now have the required significant feature matrix for classification. Certain parameters have to be set in order to get a better classification performance. In both the methods, for the selection of optimal parameters we have used the grid search method [40]. The Gaussian kernel parameter σ is selected from the set $\{2^j | j = -10, -9, ..., 10\}$. The values of the parameters (a_1, a_2, a_3, a_4) and (E_1, E_2) for the kernel were selected from the sets $\{2^j | = -5, -3, -1, 0, 1, 3, 5\}$ and $\{0.6, 0.7, 0.8, 0.9, 1\}$ respectively. For the training phase, we set $a_1 = a_2$ and $a_3 = a_4$ to reduce the training time [19]. The classifiers used in both the methods are TSVM, LS-TSVM and RELS-TSVM. The ten-fold cross-validation methodology has been applied to evaluate the efficiency of the classifiers [41]. We choose 70% of the data for training and rest 30% for testing. The results achieved by these methods are shown in Table 3.3 and Table 3.4.

TABLE 3.3: Accuracy % of the first proposed method (Entropy-based features)

Entropy-based features	TSVM	LS-TSVM	RELS-TSVM
Log energy entropy	**100**	**100**	**100**
Shannon entropy	**98.75**	**98.75**	97.92
SURE entropy	96.25	98.33	**98.75**
Combined	99.58	96.25	**100**
Shannon-log energy entropy	**99.58**	95.83	99.17
Log energy-SURE entropy	97.5	**99.58**	**99.58**
SURE-Shannon entropy	98.33	**99.58**	99.17

TABLE 3.4: Accuracy % of the second proposed method (Hjorth parameters)

Hjorth parameters	TSVM	LS-TSVM	RELS-TSVM
Activity	97.5	**98.33**	97.5
Mobility	**98.33**	97.5	**98.33**
Complexity	95	93.33	**96.67**

3.1 Discussion

In this section we briefly discuss some existing methods and the comparison with the proposed method. Table 3.5 contains some of the works done on seizure versus non-seizure classification of EEG signals. In [21], the ability of time-frequency domain has been explored on EEG signals containing epileptic seizures classified by artificial neural network (ANN) provided an accuracy of 97.73%. In [42], line length features were extracted by discrete wavelet transform (DWT) and classified by multi-layer perceptron neural network (MLPNN) given an accuracy of 97.77%. In [43], discrimination of the EEG signals were done both by SVM and probabilistic neural network (PNN), signals decomposed by DWT with energy standard and entropy features which gave an accuracy of 95.44%. In [45], decomposition of EEG signals were done by DWT with fuzzy approximate entropy and classified by SVM with accuracy 97.38%. In [14], DTCWT with features energy, standard deviation (std), root-mean-square (rms), Shannon entropy are applied through general regression neural network (GRNN) yielding an accuracy of 100%. In [44], signal decomposition based on empirical mode decomposition for classifying seizure and non-seizure EEG signals by LS-SVM gave accuracy 99.50%-100% for different intrinsic mode functions obtained after decomposition. Recent work [8], analytic time-frequency flexible wavelet transform (ATFFWT) with fractal dimension (FD) features with least squares support vector machine (LSSVM) gave an accuracy of 99.20%. In [46], local binary pattern (LBP) with histogram features fed to SVM gave an accuracy of 99.31%. In this work, EEG signals decomposed by FAWT with log energy entropy and the combination of all these entropies viz., log energy entropy, Shannon entropy

Table 3.5: Comparison of existing methods for classification of seizure and non-seizure EEG signals

Works	Year	Method	Classifier	Accuracy (%)
Tzallas et al. [21]	2007	Time-frequency analysis	ANN	97.73
Guo et al. [42]	2010	Line length	MLPNN	97.77
Gandhi et al. [43]	2011	DWT and energy- and entropy features	SVM and PNN	95.44
Bajaj and Pachori [44]	2012	EMD	LS-SVM	99.50-100
Kumar et al. [45]	2014	DWT-fuzzy approximate entropy	SVM	97.38
Swami et al. [14]	2016	DTCWT and energy, std, rms Shannon entropy features	GRNN	100
Sharma et al. [8]	2017	ATFFWT and FD features	LS-SVM	99.20
Tiwari et. al [46]	2017	LBP and histogram feature	SVM	99.31
Our proposed work	**2018**	**FAWT with log energy entropy**	**TSVM, LSTSVM, RELS-TSVM**	**100, 100, 100 respectively**

and SURE entropy fed into TSVM, LSTSVM, RELS-TSVM gave an maximum accuracy of 100%.

Table 3.6 consists of some works done on seizure and seizure-free classification of EEG signals. In [47], various features viz., power spectral features, Petrosian fractal dimension and Higuchi fractal dimension, Hjorth parameters were used as features of EEG signals and further classified using PNN with accuracy of 97%. Empirical mode decomposition (EMD) is applied for the analysis of EEG signals based on Hilbert transformation with energy of linear prediction and classification by SVM yielded an accuracy of 95.33% in [23]. Classification of the epileptic seizure in EEG signals using second-order difference plot of intrinsic mode function (IMF) by MLPNN with an accuracy of 97.75% in [24]. Features from phase space representations (PSRs)

Table 3.6: Comparison of existing methods for classification of seizure and seizure-free EEG signals

Works	Year	Method	Classifier	Accuracy (%)
Bao et al. [47]	2008	Power spectral features, Petrosian fractal dimension and Higuchi, Hjorth parameters	PNN	97
Joshi et al. [25]	2014	Energy of linear prediction	SVM	95.33
Pachori and Patidar [24]	2014	95% Second-order difference plot intrinsic mode functions	MLPNN	97.75
Sharma and Pachori [26]	2014	95% PSR of IMFs	SVM	98.67
Gupta et al. [13]	2017	FAWT with cross correntropy, log energy entropy, SURE entropy	LS-SVM	94.41
Our proposed work	**2018**	**FAWT with Hjorth parameters**	**TSVM,** **LSTSVM,** **RELS-TSVM**	**98.33** **98.33** **98.33**

of IMFs of EEG signals and employed LS-SVM for classification with accuracy of 98.67% in [26]. Similar work with cross correntropy, log energy entropy and SURE entropy implemented by Gupta et al. in [13] gave an accuracy of 94.41% with LSSVM as classifier. In this work, EEG signals decomposed by FAWT with Hjorth parameters features fed into TSVM, LSTSVM, RELS-TSVM gave a maximum accuracy of 98.33%.

Chapter 4

Conclusion and future work

The manual analysis of epilepsy in EEG signals is not efficient and time-consuming. Hence the methodologies proposed in this project can overcome these shortcomings and minimize the errors too. Early detection and timely diagnosis may help a person to prevent serious brain disorders. The experimental results presented in this project demonstrates that FAWT is a useful tool to decompose EEG signals into sub-bands for better analysis. The features used in both the methods are found to be effective and RELS-TSVM as classification model with Gaussian kernel gives a maximum accuracy of 100% for classifying seizure and non-seizure EEG signals and 98.33% for the classification of seizure and seizure-free EEG signals.

In this thesis, the approach of the two methods are novel. FAWT is useful as a decomposition tool for biomedical signals. Variants of support vector machines is widely used as classifiers for various classification problems. As discussed and by the experimental results, the performance of RELS-TSVM is better and comparable with other classifier. It is also worth mentioning that RELS-TSVM is used as a classifier for EEG signals for the first time.
I would be very interested to explore more on these areas of machine learning and its applications in various areas of research.

Bibliography

[1] Andrzejak, R.G., Lehnertz, K., Rieke, C., Mormann, F., David, P., Elger, C.E. (2001), *Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state*, Phys. Rev. E., vol. 64, no. 6 (Part 1) 061907 (8 pages).

[2] Zhang, C., Li, B., Chen, B., Cao, H., Zi, Y., He, Z. (2015),*Weak fault signature extraction of rotating machinery using flexible analytic wavelet transform*, Mech. Syst. Signal Process., vol. 64, pp. 162-187.

[3] Epilepsy, *http://www.who.int/mental_health/neurology/epilepsy/en/index.html*

[4] Epilepsy 2017, *www.who.int/mediacentre/factsheets/fs999/en/.*

[5] Gaur, P., Pachori, R.B., Wang, H., Prasad, G. (2015), *An Empirical Mode Decomposition Based Filtering Method for Classification of Motor-Imagery EEG Signals for Enhancing Brain-Computer Interface*, IJCNN Neural Networks.

[6] Burrus, C.S., Gopinath, R.A., Guo, H. (1998), *Introduction to Wavelets and Wavelet Transforms-A Primer*, Prentice Hall.

[7] Bayram, I. (2013), *An analytic wavelet transform with flexible analytical time-frequency covering*, IEEE Trans. Signal Process., vol. 61, no. 5, pp. 1131-1142.

[8] Sharma, M., Pachori, R.B., Acharya, U.R. (2017), *A new approach to characterize epileptic seizures using analytic time-frequency flexible wavelet transform and fractal dimension*, Pattern Recogn. Lett., vol. 94, pp. 172-179.

[9] Lemons, D.S. (2013), *A student's guide to entropy*, Cambridge University Press.

[10] Sharma, A., Amarnath, M., Kankar, P.K. (2016), *Feature extraction and fault severity classification in ball bearings*, J. Vib. Control, vol. 22, no. 1, pp. 176-192.

[11] Anci, D. (2009), *An expert system for speaker identification using adaptive wavelet SURE entropy*, Expert Syst. Appl., vol. 36, no. 3 (Part 2), pp. 6295-6300.

[12] Shannon, C. (1948), *A mathematical theory of communication*, Bell Syst. Tech. J., vol. 27, pp. 379-423.

[13] Gupta, V., Priya, T., Yadav, A.K., Pachori, R.B., Acharya, U.R. (2017), *Automated detection of focal EEG signals using features extracted from flexible analytic wavelet transform*, Pattern Recogn. Lett., vol. 9, pp. 180-188.

[14] Swami, P., Gandhi, T. K., Panigrahi, B.K., Tripathi, M., Anand, S. (2016), *A novel robust diagnostic model to detect seizures in electroencephalography*, Expert Syst. Appl., vol. 56, pp. 116-130.

[15] Hjorth, B. (1970), *EEG analysis based on time domain properties*, Electroencephalography and Clinical Neurophysiology, vol. 29, no. 3, pp. 306-310.

[16] Cortes, C., Vapnik, V.N. (1995), *Support Vector Networks*, Machine Learning, vol. 20, pp. 273-297.

[17] Jayadeva, Khemchandani, R., Chandra, S. (2007), *Twin support vector machines for pattern classification*, IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 5, pp. 905-910.

[18] Kumar, M.A., Gopal, M. (2009), *Least squares twin support vector machines for pattern classification*, Expert Syst. Appl., vol. 36, no. 4, pp. 7535-7543.

[19] Tanveer, M., Khan, M.A., Ho, S-S. (2016), *Robust energy-based least squares twin support vector machines*, Appl. Intell., vol. 45, no. 1, pp. 174-186.

[20] Kuhn, H.W., Tucker, A.W. (1951), *Nonlinear Programming*, Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, pp. 481-492.

[21] Tzallas, A.T., Tsipouras, M. G., Fotiadis, D. I. (2007), *Automatic seizure detection based on time-frequency analysis and artificial neural networks*, Computat. Intell. Neurosci., vol. 2007.

[22] Ocak, H. (2009), *Automatic detection of epileptic seizures in EEG signals using discrete wavelet transform and approximate entropy*, Expert Syst. Appl., vol. 36, pp. 2027-2036.

[23] Pachori, R.B., Bajaj, V. (2011), *Analysis of normal and epileptic seizure EEG signals using empirical mode decomposition*, Computer Methods and Programs in Biomed., vol. 104, no. 3, pp. 373-381.

[24] Pachori, R.B., Patidar, S. (2014), *Epileptic seizure classification in EEG signals using second-order difference plot of intrinsic mode functions*, Computer Methods and Programs in Biomed., vol. 113, no. 2, pp. 494-502.

[25] Joshi, V., Pachori, R.B., Vijesh, A. (2014), *Classification of ictal and seizure-free for EEG signals using fractional linear prediction*, Biomed. Signal Process. and Control, vol. 9, pp. 1-5.

[26] Sharma, R., Pachori, R.B. (2015), *Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode function*, Expert Syst. Appl., vol. 42, no. 3, pp. 1106-1117.

[27] Kumar, T. S., Kanhangad, V., Pachori, R.B. (2015), *Classification of seizure and seizure-free EEG signals using local binary patterns*, Biomed. Signal Process. and Control, vol. 15, pp. 33-40.

[28] Bhattacharyya, A., Pachori, R.B. (2017), *A multivariate approach for patient-specific EEG seizure detection using empirical wavelet transform*, IEEE Trans. on Biomed. Eng., vol. 64, no. 9, pp. 2003-2015.

[29] Richhariya, B., Tanveer, M. (2018), *EEG signal classification using universum support vector machine*, Expert Syst. Appl., vol. 106, pp. 169-182.

[30] Kumar, M., Pachori, R.B., Acharya, U.R. (2016), *An efficient automated technique for CAD diagnosis using flexible analytic wavelet transform and entropy features extracted from HRV signals*, Expert Syst. Appl., vol. 63, pp. 165-172.

[31] Kumar, M., Pachori, R.B., Acharya, U.R. (2017), *Characterization of coronary artery disease using flexible analytic wavelet transform applied on ECG signals*, Biomed. Signal Process. and Control, vol. 31, pp. 301-308.

[32] Kumar, M., Pachori, R.B., Acharya, U.R. (2017), *Use of accumulated entropies for automated detection of congestive heart failure in flexible analytic wavelet transform framework based on short-term HRV signals*, Entropy, vol. 19, no. 3 (92).

[33] Tanveer, M. (2015), *Robust and sparse linear programming twin support vector machines*, Cogn. Comput., vol. 7, pp. 137-149.

[34] Nasiri, J.A., Charkari, N.M., Mozafari, K. (2014), *Energy-based model for least squares twin support machines for human action recognition*, Signal Process., vol. 104, pp. 248-257.

[35] Shao, Y.H., Zhang, C.H., Wang, X.B., Deng, N.Y. (2011), *Improvements on twin support vector machine for classification*, IEEE Trans. Neural Netw., vol. 22, no. 6, pp. 962-968.

[36] Tian, Y., Qi, Z., Ju, X.,Shi, Y.,Liu, X.(2014), *Nonparallel support vector machines for pattern classification*, IEEE Trans. Cybern., vol. 44, no. 7, pp. 1067-1079.

[37] Tanveer, M. (2015), *Application of smoothing techniques for linear programming twin support vector machines*, Knowledge Info. Syst., vol. 45, no. 1, pp. 191-214.

[38] University of Bonn Germany, EEG Time Series Data *http://epileptologie-bonn.de/cmsfront_content.php?idcat 193*.

[39] P.E. McKight, P.E., Najab, J. (2010), *Kruskal-Wallis test*, Corsini Encycl. Psychol.

[40] Hsu, C.W., Lin, C.J. (2002), *A comparison of methods for multi-class support vector machines*, IEEE Trans. Neural Netw., vol. 13, no. 2, pp. 415-425.

[41] Duda, R.O., Hart, P.R., Stork, D.G. (2012), *Pattern classification*, Wiley Second edition, New York.

[42] Guo, L., Rivero, D., Dorado, J., Rabuna, J.R.,Pazos, A. (2010), *Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks*, J. Neurosci. Methods, vol. 191, no. 1, pp. 101-109.

[43] Gandhi, T., Panigrahi, B.K., Anand, S. (2011), *A comparative study of wavelet families for EEG signal classification*, Neurocomputing , vol. 74, no. 17, pp. 3051-3057.

[44] Bajaj, V., Pachori, R.B. (2012), *Classification of seizure and non-seizure EEG signals using empirical mode decomposition*, IEEE Trans. Inf. Technol. Biomed., vol. 16, no. 6, pp. 1135–1142.

[45] Kumar, Y., Dewal, M.L., Anand, R.S. (2014), *Epileptic seizure detection using DWT based fuzzy approximate entropy and support vector machine*, Neurocomputing, vol. 133, pp. 271-279.

[46] Tiwari, A.K., Pachori, R.B., Kanhangad, V., Panigrahi, B.K. (2017), *Automated diagnosis of epilepsy using key-point-based local binary pattern of EEG signals*, IEEE J. Biomed. and Health Inform., vol. 21, no. 4, pp. 888-896.

[47] Boa, F.S., Lie, D.Y.-C., Zhang, Y. (2008), *A new approach to automated epileptic diagnosis using EEG and probabilistic neural networks*, 20th IEEE International Conference on Tools with Artificial Intelligence, vol. 2, pp. 482-486.