Analysis & Classification of Vibroarthrographic (VAG) Signals using Statistical Features

Author

UMER ABDUL REHMAN KHAN

MS-19 (EE) 00000319329

Supervisor

Dr. SHAHZAD AMIN SHEIKH

DEPARTMENT OF ELECTRICAL ENGINEERING

COLLEGE OF ELECTRICAL AND MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

July 2022

Analysis & Classification of Vibroarthrographic (VAG) Signals using Statistical Features

Author

UMER ABDUL REHMAN KHAN

MS-19 (EE) 00000319329

This thesis submitted in absolute accomplishment with the requirements regarding the degree of MS Electrical Engineering

Thesis Supervisor:

Dr. SHAHZAD AMIN SHEIKH

Supervisor’s Signature: ____

DEPARTMENT OF ELECTRICAL ENGINEERING

COLLEGE OF ELECTRICAL AND MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

July 2022

AUTHOR'S DECLARATION

I confirm that this dissertation titled "Analysis & Classification of Vibroarthrographic (VAG) Signals using Statistical Features" is my research-based work. I solemnly pronounce that the research work whether theoretical or practical presented in the MS Thesis is exclusively done by me. Moreover, I have not presented this thesis (or substantially similar research work) or any part of the thesis previously to any other degree awarding institution within Pakistan.

As I am totally aware of substantial value of no usage of unfair means whatsoever the reason along with NUST's well established policy of zero tolerance policy towards plagiarism. Therefore, I as author of the above mentioned thesis, solemnly declare that I have abstained from such means and no chapter of my thesis has been plagiarized. The content taken from other resources has been properly acknowledged/referenced accordingly.

Signature of Student

UMER ABDUL REHMAN KHAN

MS-19 (EE), 00000319329

LANGUAGE CORRECTNESS CERTIFICATE

This thesis has been read by an English expert and is free of typing, syntax, semantic, grammatical and spelling mistakes. Thesis is also according to the format given by the university.

____

Signature of Student

UMER ABDUL REHMAN KHAN

MS-19 (EE), 00000319329

____

Signature of Supervisor

Dr. SHAHZAD AMIN SHEIKH

COPYRIGHT STATEMENT

ACKNOWLEDGEMENTS

First of all, I thanks to Allah Almighty for His innumerable blessings and divine guidance throughout my life.

I particularly thanks to my supervisor Dr. Shahzad Amin Sheikh, and GEC members, Dr. Fahad Mumtaz Malik and Asst. Prof. Kamran Aziz Bhatti, for their excellent guidance and very consistent support during this research. Without their help, I may not be here today. Moreover, I would also like to pay my gratitude to all our faculty members who advised or helped us in any form during the thesis. Prof. Rangaraj M. Rangayyan of the University of Calgary in Canada generously granted the researchers permission to utilize their VAG signal data set for research purpose. I also want to express my thanks to him for his permission.

I also want to appreciate my family members who have a trust in me and support me in my continuous hard work in the most challenging times. I am very thankful to particularly my parents, for their constant support, prayers, inspiration during the educational career and for the continuous support in the difficulties I faced during this thesis.

COMPLETION CERTIFICATE

“It is certified that the data of this research document titled “Analysis & Classification of Vibroarthrographic (VAG) Signals using Statistical Features” submitted by Mr. Umer Abdul Rehman Khan, MS-19 (EE), Registration No. 00000319329 has been found acceptable for the requirements of degree.

Thesis Supervisor: ______________________________

(Dr. SHAHZAD AMIN SHEIKH)

ABSTRACT

Human Body joints are vital for normal body movements. Early diagnosis of any problem in joints helps the doctors for the timely treatment of the diseases. To identify these problems, many diagnosis methods are available but the easiest and most effective way is through Vibroarthrography. Vibroarthrography is the method of detecting the vibration signals from the knee joint to diagnose any disorders in it. Researchers are studying the usage of vibration signals from the human knee joints, known as Vibroarthrographic (VAG) Signals, for the diagnosis of the condition of the knee joint. There are various types of features and classifiers used for the classification of VAG signals into normal and abnormal signals. In this research, different types of features of the time domain and spectral domain are explored, and studied the combination of these features. These features include statistical features, Auto-Encoder Based features, and Continuous Wavelet Transform based features. The features are then selected by correlation coefficients and fed into classifiers models. Different classifiers are examined but the best results have been achieved by using the Decision Tree Classifier. The accuracy achieved using the Decision Tree Classifier is 93.26%. We have concluded that the proposed methodology performed very well in other performance evaluation parameters as well. We achieved the Sensitivity of 86.84%, Specificity of 98.04%, PPV of 97.06%, NPV of 90.91%, and a Matthews Correlation Coefficient (MCC) score of 0.8641. The proposed method has Area under the Curve of ROC approximately equal to 0.91. The proposed methodology gives us more accurate results, as compared to previous researches, without going into the Deep learning methods that are complex and time-consuming.

TABLE OF CONTENTS

AUTHOR'S DECLARATION iii

LANGUAGE CORRECTNESS CERTIFICATE iv

COPYRIGHT STATEMENT v

ACKNOWLEDGEMENTS vi

COMPLETION CERTIFICATE vii

ABSTRACT viii

TABLE OF CONTENTS ix

LIST OF FIGURES xii

LIST OF TABLES xiii

LIST OF ABBREVIATIONS xiv

CHAPTER 1: INTRODUCTION 1

1.1 Introduction 2

1.2 Block Diagram 3

1.3 Motivation 3

1.4 Problem statement 5

1.5 Resources used 5

1.6 Advantages 5

1.7 Applications 5

CHAPTER 02: LITERATURE REVIEW 6

2.1 Vibroarthography 7

2.2 Literature Review 7

CHAPTER 03: METHODOLOGY 11

3.1 Data Acquisition 12

3.2 Pre-processing on Data-set 14

3.3 Wavelet Decomposition (WD) and Signal Reconstruction 14

3.4 Feature Extraction 15

3.4.1 Statistical Features 16

3.4.1.1 Envelope Amplitude Signal Features 16

3.4.1.2 Symbolic Entropy 17

3.4.1.3 Approximate Entropy 18

3.4.1.4 Sample Entropy 19

3.4.1.5 Shannon Entropy 20

3.4.1.6 Renyi Entropy 21

3.4.1.7 Tsallis Entropy 21

3.4.1.8 Permutation Entropy 22

3.4.1.9 VAG Signal Standard Deviation 23

3.4.1.10 Auto-Correlation 23

3.4.2 Auto-Encoder based Features 24

3.4.3 Continuous Wavelet transform (CWT) based Features 25

3.4.3.1 Spectral Slope Feature of Magnitude of CWT 27

3.4.3.2 Spectral Flux Feature of Magnitude of CWT 28

3.4.3.3 Spectral Mean Feature of Magnitude of CWT 28

3.5 Feature Selection using Correlation Coefficient 29

3.6 Classification of VAG Signals 30

3.6.1 Decision Tree Classifier 30

3.6.2 Support vector Machine (SVM) 31

3.6.3 Ten-fold cross-validation 31

CHAPTER 04: RESULTS AND DISCUSSION 33

4.1 A Raw Abnormal and Normal VAG Signal 34

4.2 Original Signal Vs Smoothed Signal 34

4.3 Wavelet Decomposed Sub-bands and Reconstructed Signal 36

4.4 Performance Evaluation 40

4.5 Classification Results 43

4.6 Comparison of proposed methodology with previous literature 46

CONCLUSION 49

REFERENCES 50

LIST OF FIGURES

Figure 1: Block Diagram of Proposed Methodology 3

Figure 2: Sensor Placement in Data Acquisition 13

Figure 3 Flexion and Extension of Knee Joint 13

Figure 4: Daubechies (db4) 15

Figure 5: Auto-Encoder 25

Figure 6: Analytic Morlet (Gabor) Wavelet 27

Figure 7: Working of 10-Fold Cross Validation 32

Figure 8: Raw VAG signal of a Healthy subject 34

Figure 9: Smoothed VAG signal of a Healthy subject 35

Figure 10: Raw VAG Signal of an Unhealthy Subject 35

Figure 11: Smoothed VAG Signal of an Unhealthy subject 36

Figure 12: WD of a Healthy VAG subject 37

Figure 13: Reconstructed VAG Signal of a Healthy Subject 38

Figure 14: CWT of a Healthy Subject 38

Figure 15: WD of an Unhealthy VAG subject 39

Figure 16: Reconstructed VAG Signal of an Unhealthy Subject 39

Figure 17: CWT of an Unhealthy Subject 40

Figure 18: ROC Curve for the DT Classifier 45

Figure 19: ROC Curve for the SVM Classifier 45

Figure 20: Confusion Matrix of the DT Classifier 46

Figure 21: Confusion Matrix of the SVM Classifier 46

LIST OF TABLES

Table 1: Classification Results of DT & SVM 44

Table 2: Comparison of Proposed Methodology with Previous Researches 48

LIST OF ABBREVIATIONS

VAG …………………. Vibroarthrographic

WD ………………….. Wavelet Decomposition

TFR ……………….….. Time-frequency Representation

EAS ………………….. Envelope Amplitude Signal

STFT ………………… Short Time Fourier Transform

DL …………………… Deep Learning

SVM ………………… Support Vector Machine

DT ……………………Decision Tree

PPV………………….. Positive Predicted Value

NPV…………………. Negative Predicted Value

MCC………………… Mathews Correlation Coefficient

ROC ………………… Receiver Operating Characteristics

CHAPTER 1

INTRODUCTION

In this chapter, we will cover the introduction of our thesis. We will discuss the basic outline of thesis, what objectives we will achieve in this research and what are the benefits of our thesis in biomedical field. Following contents are covered in this chapter.

1.1 Introduction

1.2 Block Diagram

1.3 Motivation

1.4 Problem statement

1.5 Resources used

1.6 Advantages

1.7 Applications

1.1 Introduction

Healthy human body joints are essential for a healthy life. Due to the progression in a human society, the lifestyle of a person is also changed with it. Unhealthy lifestyle leads to many problems like in human knee joints. Human knee joints help the person in their daily movement. Any issue in human knee joint should be diagnosed early for the proper and timely treatment. There are many methods available for this purpose. These methods are of two types: One in Invasive and second is Non-invasive method. In Invasive method, the physician physically enter in the body part and examined it by cutting the skin or by place in the instruments into the body. While, in Non-invasive method, there is not involved any cutting or puncturing of skin. In addition, it does not come into contact with the mucosa or internal organs other than through a natural or artificial body orifice. Many non-invasive methods are used commercially to check the condition of knee joint and diagnose any problem in it. For example, X-Rays, CT scan, MRI and ultrasound are used in this procedure. The most efficient non-invasive method is Vibroarthrography. It is cost effective, radiation free and results are obtained quickly as compared to other methods [1].

In Vibroarthrography, the signals are generated by using the sensors placed on the mid patella of human body knee joint. The movement of knee joint generate the sound based vibrational signals, which are captured by the sensor. These signals are known as Vibroarthrographic (VAG) signals. These signals give the information about the condition of knee joint. This approach is used to identify anomalies in the human body knee joint.

VAG signals are nonlinear and nonstationary in nature. So, that is why, we cannot apply typical signal processing techniques. In order to accomplish this, VAG signal analysis employs high level processing methods.

1.2 Block Diagram

The block diagram of proposed methodology is shown in the Figure 1. This block diagram clearly shows three parts of work area. First part is for Signal Preprocessing on VAG signals, Second part is Feature Extraction where we extract different types of features and third part is Classification Models to predict the labels into normal VAG (healthy subject) and abnormal VAG (unhealthy subject) signals.

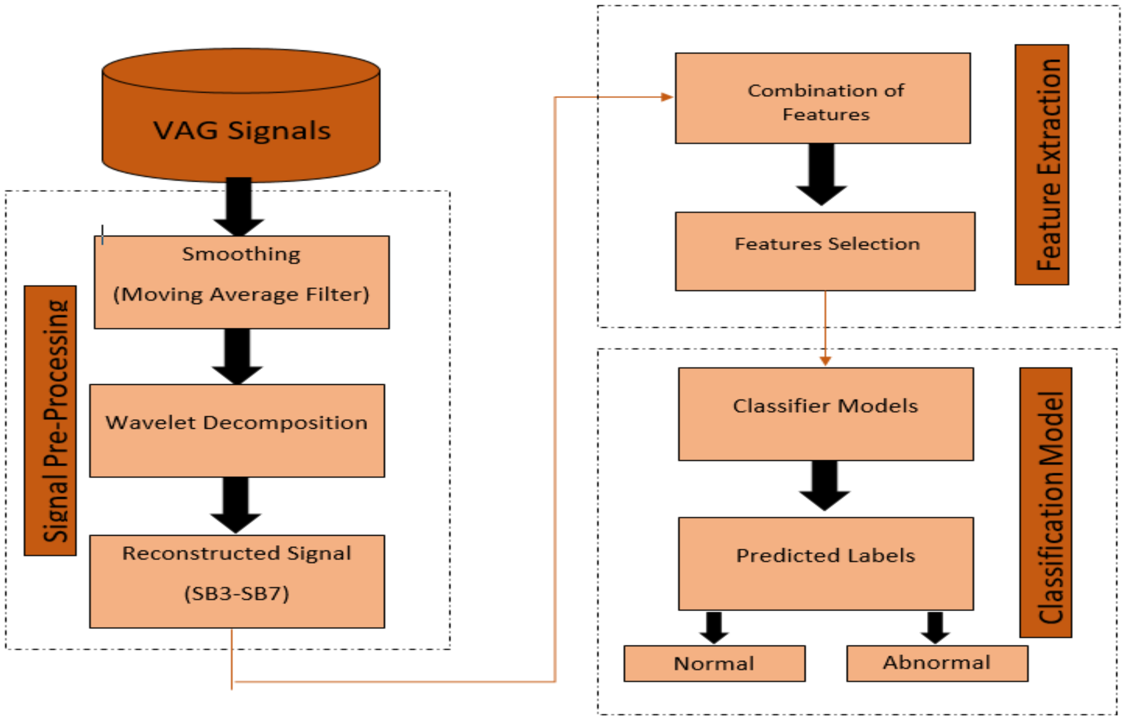

Figure 1: Block Diagram of Proposed Methodology

1.3 Motivation

The human knee joint is amongst the most important joints of body. This joint aids in flexible movements and body weight maintenance. The human knee joint carries a lot of weight that's why it is prone to injury more than any other joint. Among elderly people, these knee-joint disorders are most common. The rapid progress of any type of diseases in human body knee joint is greatly affecting the lifestyle of patients. Therefore, the earlier detection of knee joint pathology makes it easy for physician to provide the proper clinical treatment and hence to control the degeneration activity of that pathology. There are several non-invasive and invasive techniques for diagnosing any injury related to knee-joint. Non-invasive methods of Diagnosis of diseases are gaining importance in Health Care Sector. Magnetic Resonance Imaging (MRI), Computer Tomography (CT) scan, and X-Rays are a few non-invasive diagnostic methods. These methods are not suitable for routine diagnostic since these are image-based techniques and these kind of diagnosis techniques provides only static information and do not give the varying knee-joint characteristics [13]. There is also an involvement of radiation that are harmful for the person. Moreover, MRI and CT scans are expensive diagnosis methods, and the scan and reports take time to be ready for discussion [2].

So, there should be a diagnostic system that not only provides early diagnosis but also should be cost effective, less time taking and gives as much accurate result results as possible. This will help the medical practitioners to perform further clinical treatment to stop degenerative process of patients having knee-joint pathology. Many researchers have been working in this area of research and exploring other ways to diagnose problems in human body. Among different Non-invasive methods, Researchers preferred to use Vibroarthography in which Vibroarthrographic signals are acquired from the Human Body Joints, especially knee joints in our research. Vibroarthrography is the best way as compared to traditional methods because it is cost effective, radiation free and the results are obtained quickly as compared to other methods. Moreover, the use of Digital Signal Processing techniques with Machine Learning Algorithms help to create more room for researchers to make the whole process efficient and also enhance the accuracy of the whole diagnosis system.

1.4 Problem statement

- To find the best suitable features and classification algorithms learning models for developing a Configurable diagnostic system for early diagnosis of any pathology related to knee-joint.
- To develop a non-invasive diagnostic method capable of classifying healthy and unhealthy subjects as accurate as possible.

1.5 Resources used

The proposed research work including preprocessing of VAG signals, different types of features extraction and classification learners has been carried out in MATLAB R2020a. The whole research has been done on the MATLAB software.

1.6 Advantages

The proposed methodology of Vibroarthography will help in early diagnostic of any knee-joint pathology, and it will also help in early treatment of that pathology.

1.7 Applications

The proposed system can be used in health care system. A non-invasive method for early diagnosing of any knee-joint pathology may be used in any diagnostic center.

CHAPTER 02

LITERATURE REVIEW

The literature review of our thesis will be covered in this chapter. We discuss the previous researches methodology and find the level of prior study has been done in relation to our research.

2.1 Vibroarthography

2.2 Literature Review

2.1 Vibroarthography

If the pathology of the knee joint is detected early, the therapist can offer the appropriate clinical approaches to manage the worsening process of arthritis. Computational approaches have been used to diagnose knee joint disorders in addition to traditional medical studies. The Vibroarthrography technique is one of the methods used to diagnose knee joint disorders [2].

Vibroarthrography, also known as vibration arthrography, is a technique for detecting vibration signals from the knee joint in order to diagnose abnormalities. It is a non-invasive method, affordable and radiation free way of determining the extent of cartilage injury in the knee during extension and flexion movements [14].

2.2 Literature Review

Many researchers are used VAG Signals for the classification of abnormal and normal knee joints. In [2], Chen J-C used the Empirical Decomposition Method (EMD) for signal pre-processing. The VAG signals are decomposed in different Intrinsic Mode Functions (IMFs) and signal is reconstructed by using dominant IMFs. The SVM classifier has achieved accuracy of 85.30%. They employed the Hilbert Huang transform (HHT) that has been around for a long time. In [3], Yunfeng Wu et al. used a kernel-based estimation model of probability density for the illustration of signal distributions in a feature space which is bi-variate. Authors in [15] described a telehealth system of VAG analysis. The proposed system can condense VAG data into 3-bits. From the encoded signals, various time-domain features like spiky index, band power, rapid change factor etc. are calculated. Experimental findings revealed that the classification accuracy of extracted features through 3-bit encoded signal is more than raw VAG signal dataset.

There is an issue of mode mixing in EMD and it can be resolved by the new updated method that is Ensemble Empirical Mode Decomposition (EEMD). In [4], EEMD technique is employed to do a signal pre-processing for the removal of noise and baseline wanders. The VAG signal is reconstructed by using dominant IMFs through Detrended Fluctuation Analyses (DFA). Then the 3 features are calculated from each reconstructed VAG signals that are Envelope Amplitude signal mean, Standard deviation and Root Mean Square and 3 entropy based features that are Symbolic Entropy, Approximate Entropy and Fuzzy Entropy. The best results are obtained using the Generalized Logistic Regression Analyzer (GLRA) that has accuracy of 83%. The EEMD technique is also employed in [5] [19] for obtaining the IMFs. The DFA is used for reconstruction of signals from dominant IMFs. Three entropy based features are used for classification, which are spectral entropy, permutation entropy, Tsallis entropy. The Random Forest classifier has given the accuracy of 86.52%. The EEMD technique for signal pre-processing has an issue of allowing residual noise in the reconstructed signal. To overcome this issue, an advanced technique for signal decomposition, known as Complete Ensemble Empirical Mode Decomposition with Additive Noise (CEEMDAN) is implemented. In [6], S. Nalband et al. used CEEMDAN technique to reconstruct the VAG Signals in which there is no residual noise in it. Dominant IMFs (IMF5+IMF6+IMF7) are used to reconstruct the VAG signal. Total 6 entropy based entropies are comuted. These entropies are are Sample Entropy, Approximate Entropy, Shannon Entropy, Renyi Entropy, Tsallis Entropy, and Permutation Entropy. The classification accuracy achieved in this research is 86.61%.

In [7], researchers used a wavelet decomposition method for the division of signals into sub-bands. The signal is reconstructed using suitable sub-bands and some entropy-based features are extracted from it. The features are selected using a genetic and apriori algorithm to classify VAG signals using LS-SVM and Random Forest. Rangayyan et al. [8] used an Artificial Neural Network for the classification of healthy and pathological VAG subjects. The proposed method can accurately screen out the signals around 78% subjects based on AUC-ROC curve. In [9], Rui Gong et al. used Empirical Mode Decomposition method for the VAG signal decomposition into IMFs and reconstructed the signals from dominant IMFs. Then 4 features are extracted from Continuous Wavelet Transform (CWT) in different frequency ranges and 2 features are extracted from the Auto-Correlation Function. The LS-SVM classifier gave the accuracy of 86.67%. In [10], Wu et al. used a system of multiple classifiers that based on (LS-SVM) and Recurrent Neural Network (RNN) for the classification of 89 VAG signals [11].

Another method for inferring the healthy and diseased status of the knee joint is to use the Hilbert-Huang Transform (HHT) [16]. The Hilbert Transform is applied to IMFs in the z-plane that are represented in circular geometry. Each circle's geometrical area is considered a set of features to perform classification. S. Nalband et al. in [17] used two separate techniques for representing of VAG signals in time-frequency terms. (SPWVD) and Hilbert Huang Transform were implemented by the authors with (CEEMDAN). From CEEMDAN-based HHT and SPWVD, statistical features extraction is done from dominant IMFs and sub-band, respectively. Pattern classification is done via LS-SVM. Wu and Krishnan [10] employed a system of multiple classifier which is based on (LS-SVM) and recurrent neural network (RNN) for classification of a VAG signal dataset comprising of 89 signals [11]. They achieved an accuracy figure of 80.9 percent by using a Parzen window [20] for estimation of signal's probability density function.

In [12], M.M. Shidore et al. used standard deviation thresholding to reduce the datasets from 89 signals to 65 signals. They extracted the time domain and spectral domain features from the reduced datasets. The features are selected by different feature selection techniques that are ANOVA Test, Mutual Information and Chi Square Test. The classification of VAG signals into normal and abnormal signals have been done by SVM, RF and GNB. Better results are obtained using the Random Forest classifier, trained with features selected by the Mutual Information (MI) test and achieved a classification accuracy of 89.23%.

In all of above mentioned researches, the number of features were very less. That is why, there is not any significant improvement in the accuracy of classification of VAG signals. Exploring different types of features and their combination help to get more accuracy. Moreover, adding more features will enable us to obtain extremely accurate findings.. In this research, we have used all the VAG signals (89) of dataset for training and testing the systems, without reducing the size of dataset by removal of VAG signals that are difficult to classify. In this way, we have improved our results without going to a Deep Learning Techniques, which are complex in nature and have high computation time.

CHAPTER 03

METHODOLOGY

In this chapter, we will discuss different methodologies we have used in our thesis. What are different methods we used and how we implement them are described in this chapter.

3.1 Data Acquisition

3.2 Pre-processing on Data-Set

3.3 Wavelet Decomposition (WD) and Signal Reconstruction

3.4 Feature Extraction

3.4.1 Statistical Features

3.4.2 Auto-Encoder Based Features

3.4.3. CWT Based Features

3.5 Features Selection

3.6 Classification of Abnormal and Normal Subjects

3.1 Data Acquisition

The dataset containing VAG signals was developed by Prof. Rangaraj M. Rangayyan at the Research Centre in University of Calgary, Canada [4, 6, 9, 12]. The authorized dataset was created by a setup of data acquisition that included people with normal and abnormal conditions of knee joint, a miniature accelerometer (Model 3115A, Dytran Instruments, Inc., Chatsworth, CA, USA) for capturing vibration signals, a data acquisition board from National Instruments, and LABVIEW software [18]. Among these 89 subjects:

- 51 subjects have healthy knee condition (22 males, 29 females, age 28 ± 9.5 years)
- 38 subjects were having some type of knee joint disorder (20 males, 18 females, age 35 ± 13.8 years).

Each of the 89 subjects went through the same experimental protocol. Vibration signals are captured from mid patella by placing an accelerometer as the participant swung the leg from full flexion (135 degrees) to full extension (0 degrees) and then back to 135 degrees in 4 seconds. The volunteers were given verbal instructions for completing the flexion to extension and back to flexion cycle in as little time as feasible, with each flexion and extension lasting two seconds. The first half of a VAG signal corresponds to extension of leg whereas the second half corresponds to flexion of leg [4, 6]. The VAG signal was pre-filtered (10Hz–1 kHz) followed by amplification. The VAG signal had a sample frequency of 2kHz and was digitized with a 12-bit resolution. The 89 VAG samples contains either 7500 or 8000 timestamps, each of which corresponds to a single axis accelerometer acceleration value. We did zero padding to the signals with 7500 timestamps, by adding zero to the last 500 timestamps, to ensure the same length across the samples, which is 8000.

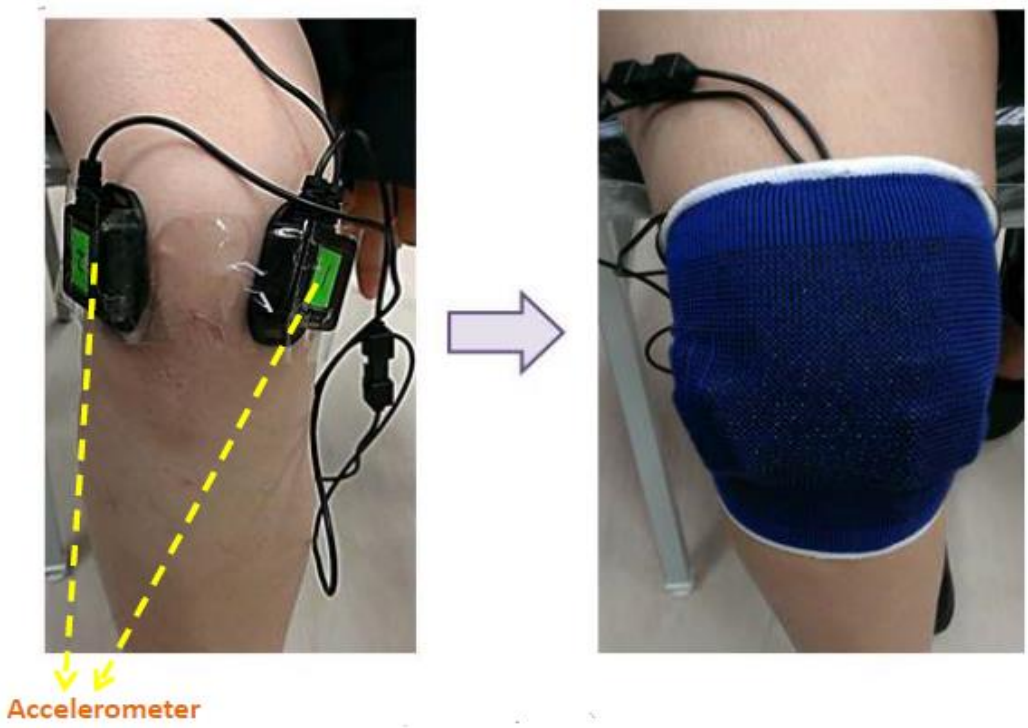

Figure 2: Sensor Placement in Data Acquisition

The movement of flexion and extension of human body knee-joint is shown in figure 3.

Figure 3: The movement of Flexion and Extension of Knee Joint

3.2 Pre-processing on Data-set

It's difficult to keep the real sensor signal from being tampered with by numerous sorts of noise. Noise can come from a variety of sources, including supply lines, electrical circuitry, measurement devices, the environment, and thermal impacts. Such systems create noise that accumulates in the raw VAG signal and make the signal erroneous that means it varies from the baseline threshold resulting in baseline wandering. Before further processing and analysis, it is critical to remove the VAG signal's drift [12].

In this thesis, we applied smoothing on all 89 VAG signals to get resulting signals free of noise.

Smoothing is done via the Signal Analyzer Application in MATLAB R2020a.

3.3 Wavelet Decomposition (WD) and Signal Reconstruction

Due to the multicomponent composition, nonstationary and nonlinear nature of vibroarthographic signals, a nonstationary signal processing technique has been used. Wavelet decomposition (WD) is a technique that we used. For the examination of Vibroarthographic signals, it is utilized as a preprocessing approach. At the first stage of decomposition, this algorithm decomposes the VAG signal into the detail coefficients and approximations. These coefficients are then further split into approximation and coefficients. As a result, the signal disintegrates like a tree. These coefficients belong to a specific frequency at each stage. The signal was decomposed using a 4th order (db4) wavelet from the Debaucheries family.

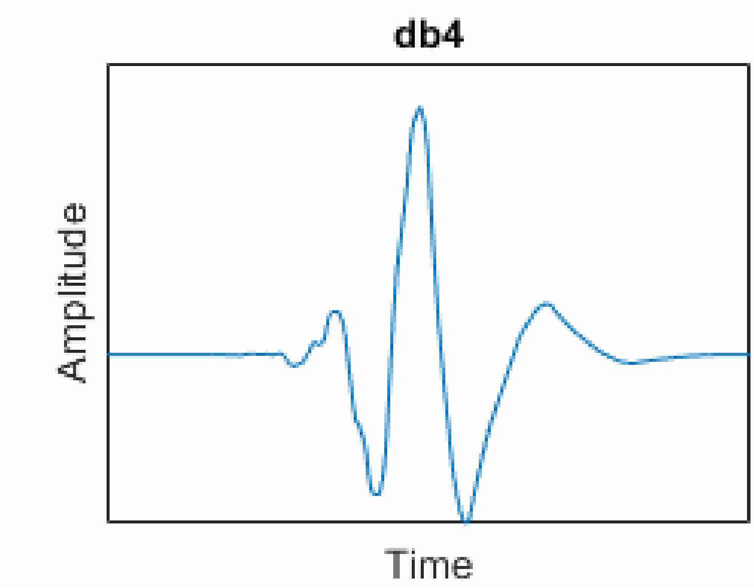

Figure 4: Daubechies (db4)

We have used ten decomposition levels to produce 11 coefficients (10 detail and an approximation coefficient). The study of biomedical signals nowadays is commonly done by implying Wavelet decomposition including EEG, EMG, and VAG etc. [10, 27].

From the 10 sub-bands (S1-S10) that we get from Wavelet Decomposition, 5 of them (S3-S7) were used to reconstruct the original VAG Signal. Then that reconstructed VAG Signal is used for further purposes.

3.4 Feature Extraction

The features extraction is an important part in the classification process. A good feature discriminate the two different type of signal with a high accuracy. Different types of features are studied and used to get a set of features that will input to classifiers. The features used in this research are Statistical Features, Auto-Encoder Based Features and CWT based features. These features help to quantify the hidden parameters of signals and analyze the signals.

3.4.1 Statistical Features

Statistical features help to explore the data statistically. It gives information about the nature of data and help to differentiate different types of data after analyzing their characteristics like mean, median, variance, and many others. Different types of statistical features are used to study the abnormal and normal behavior of VAG signals. These features quantify the irregular behavior of the signals and give us information about our dataset.

3.4.1.1 Envelope Amplitude Signal Features

The VAG signal are transformed into a new type of signal, which is known as Envelope Amplitude Signal [4]. In this process, the VAG signal is divided into sequential and non-overlapping segments. Each segment of VAG contains 20 samples. The local upper extreme or peak is extracted for each segment. In the same way, local lower extreme or peak is extracted for each segment of a VAG signal. Then, we use the piece-wise cubic Hermite Interpolation method to get the samples between local upper and lower peaks. In this way, we get upper envelope and lower envelope of a VAG signal. The difference between upper envelope and lower envelope at each time samples gives us a new signal that is "Envelope Amplitude Signal" (EAS). This new signal describes the temporal variations in an overall signal and helps to differentiate the abnormal and normal signals.

$$EA\ (t) = Upper\ Envelope\ Sig(t) - Lower\ Envelope\ Sig(t) \qquad (1)$$

There are three features that are calculated from Envelope Amplitude Signal. These features are Mean of EAS, Standard Deviation of EAS and Root Mean Square (RMS) of EAS.

3.4.1.2 Symbolic Entropy

The Symbolic Entropy is a parameter that is used to quantify the abnormality of a VAG signal. It depicts the coarse-grained signal dynamics of a physiological process [4]. In this feature, each signal "vag (t)" sample values are converted into two Quantization Levels that are 0 and 1 using a predefined threshold δ.

$$s^q(n) = \begin{cases} 0, & \left| vags\,(t) - \overline{vags\,(t)} \right| < \delta \\ 1, & \left| vags\,(t) - \overline{vags\,(t)} \right| > \delta \end{cases} \tag{2}$$

In this study, value of "q" is 2 and the threshold "δ" value is 0.123. Then the whole signal is converted into a series of 0's and 1's sequence. Then the signal is divided into the small segments, which are known as words. The word's length "M" in this research is 4. A word is a sequential set that contains "M" number of signal symbols. Then find the probability of the occurrence of each symbolic word as $P_b\left(s_M^q\right)$. The entropy of these words are calculated by using the probability density function of words s_m^q (n).

$$H_s\left(q,\ M\right) = -\sum_{b=1}^{W} P_b\left(s_M^q\right) log_2\left[P_b\left(s_M^q\right)\right] \tag{3}$$

Due to the presence of systematic errors, there is a biasness in the empirical probability estimates of $P_b\left(s_M^q\right)$. Therefore, Eguia et al. [27], introduced an additional bias term that helps to correct the value of entropy. The formula of corrected Shannon entropy is given below:

$$H_C\,(q, M) = \ H_s\,(q, M) + \frac{W_R - 1}{2W log_e 2} \tag{4}$$

Here W_R represents the overall number of actual words that appear in the symbol sequence s_M^q. The maximum value of H_C is calculated by following formula.

$$H_C^{max}(q, M) = log_2 W + \frac{W-1}{2Wlog_e 2} \tag{5}$$

Then, the Symbolic Entropy is calculated using a following formula.

$$\mathrm{SyEn}(\mathrm{q},\ \mathrm{M}) = \frac{\mathrm{H}_c(q,\ M)}{H_c^{max}(q, M)} \tag{6}$$

3.4.1.3 Approximate Entropy

Approximate entropy gives us a statistical measure about the irregular dynamics in a VAG signals [4, 6]. Pincus [30] first put up the idea with the intention of parameterizing the abnormal dynamics in biological data. In our case, we calculate the approximate entropy of a signal length N=8000. The window size "i" is 4, which is used for the similarity comparison between different segments of a signal. For similarity comparison, there is a similarity criteria "t" to use it as a tolerance threshold. The value of threshold is 0.2 in this research.

After transforming a VAG signal, we have a vector sequence contains overlapping vectors where each vector consists of "i" consecutive time samples of VAG signal.

$$v_i\ (n) = [vags(n), vags(n+1), \dots \dots, vags(n+i-1)] \tag{7}$$

$$vector\ seq = [vec_1(n), vec_2(n), vec_3(n), \dots, vec_{N-i+1}(n)\ ] \tag{8}$$

Then, we calculate the maximum distance between the vectors in a vector sequence.

$$d[u^i(j), u^i(k)] = \max_{l=1,2,.,i} |x(j+l-1) - x(k+l-1)| \tag{9}$$

Each distance value is compared with the tolerance threshold "t" value and its probability is calculated by following formula [4].

$$C_j^i(t) = \frac{number\ of\ k\ satisfying\ d\left[u^i(j), u^i(k)\right] < t}{N - i + 1} \qquad (10)$$

It displays how often two vectors within a tolerance level are similar to one another. Then, we define the function as "ϕi(t)" that gives us the average of the natural logarithm of probability values.

$$\phi^i(\mathrm{t}) = \frac{\sum_{j=1}^{N-i+1} log_e C_j^i(t)}{N - i + 1} \qquad (11)$$

At last, we calculate the approximate value using following formula.

$$\mathrm{ApEn}(\mathrm{i}, \mathrm{t}, \mathrm{N}) = \phi^i(\mathrm{t}) - \phi^{i+1}(\mathrm{t})$$

$$= \frac{\sum_{j=1}^{N-i+1} log_e C_j^i(t)}{N - i + 1} - \frac{\sum_{j=1}^{N-i} log_e C_j^{i+1}(r)}{N - i} \qquad (12)$$

3.4.1.4 Sample Entropy

Sample Entropy is a modified version of Approximate entropy. It addresses issues such as relative dissimilarity and biasing in approximate entropy [6]. It is a novel sort of entropy that provides information about VAG signals by comparing them to their delayed counterparts.

In the sample entropy, the time signal is divided into a templates of size "v". The total samples of the VAG signal are represented by "B". In the same way, the total samples are calculated, where each sample has a size of "v + 1". Then the sample entropy is calculated by using following formula.

$$\text{SampleEn}\left(v,\ u,\ N\right) = -\ln\frac{A^{v}\ (u)}{B^{v}\ (u+1)} \tag{13}$$

Researchers like Pincus [6], suggested that value of v should be around 2 and tolerance window u should be equal to 0.1 to 0.2 multiply with the standard deviation of a signal. Every signal has its unique properties and not these values can be used in every signal. In this research, we selected the “v” value equal to 4 and “u” equal to 0.2.

3.4.1.5 Shannon Entropy

Shannon proposed the Shannon entropy that is used to access the temporal complexity of a signal and its measure of information [6]. The more entropy value indicate more randomness in a given signal. First, a fast Fourier transform is used to compute the VAG signal's spectrum. After that, the signal's fourier transform is multiplied with a signal to obtain the power spectrum. The Power level of the frequency of the component is denoted by Er. The computed Power Spectral Density is normalized to obtain Probability density function. Power spectrum density e_f is used to calculate the Shannon entropy.

$$e_f = \frac{E_f}{\sum E_r}, \tag{14}$$

The Probability density function is used to calculate the Shannon entropy as given in the following formula:

$$ShannonEn = \sum_{f} e_f \log\left(\frac{1}{e_f}\right) \tag{15}$$

3.4.1.6 Renyi Entropy

The Renyi Entropy is used to define the diversity of a signal in the form of indices. It is a generalized form of Shannon entropy, which is introduced by Renyi. In abnormal VAG signals, there are high spikes in the Renyi entropy that help the classifier for differentiation of normal and abnormal signals. The formula of the Renyi entropy [6] is mentioned below:

$$\text{ReEn}\,(\propto) = \frac{1}{1 - \propto}\,(\sum_{f} e_f^{\propto}), \qquad \propto > 0,\ \propto\ \neq 1 \qquad (16)$$

Here also fourier transform is involved in the above expression. “α” is a factor that represents the enabling measurement of uncertainty in the signal distribution. In our research, the value of “α” is selected and have a value equal to 2.

3.4.1.7 Tsallis Entropy

Tsallis entropy identifies the signal complexities of VAG. The VAG signals are multi-component signals and having non-Gaussian trends, which makes the usage of Tsallis entropy best to describe its behavior [6]. The Tsallis entropy is calculated using the following formula:

$$\text{TsEn} = \frac{1 - \sum_{i=1}^{W} a_j^b}{b - 1} \qquad (17)$$

The parameter ‘b’, known as non-extensively index, is used to model the uncertainties in the VAG signals. Its value is greater than 1 for a sub-extensive system. The probability of the jth state of the signal is represented by “a_j”. As VAG signals are non-linear and non-stationary in nature and act as a sub-extensive system, we have chosen the value of “b” equal to 2 in this research.

3.4.1.8 Permutation Entropy

The permutation entropy is computationally simple and robust in nature, which is used to tell about the signal irregular behavior. It is proposed by the Bandt and Pompe [6] for the entropy calculation of a non-linear signal and a non-stationary signal. Some parameters values should be selected for the computation of permutation entropy. The first parameter is the embedding dimension “m” and second parameter is a time delay “td”. Embedding dimension is represented by “m”, which specifies how much information is stored in each vector. The value of “m” give the possible number of symbols in a given signal by calculating the permutation of “m”. As a result, an embedded vector is produced and sorted in ascending order. The probability of each symbol in a vector is now determined using the formula:

$$P_k = \frac{p_k}{L - m + 1} \tag{18}$$

In the above formula, pk is the number of instances of the kth symbol of the given data length L. At last, the Permutation entropy of VAG signal is calculated by using the pk values as given below:

$$\mathrm{PeEn} = -\sum_{1}^{k} P_k \log(P_k) \tag{19}$$

In this research, we selected embedding dimension “m” value equal to 3 as it gives more accurate results.

3.4.1.9 VAG Signal Standard Deviation

The standard deviation is a statistic value that expresses how much variance or dispersion are present in a group of numbers. It tells the data's dispersion around the mean. A low standard deviation implies that the data are grouped around the mean, whereas a large standard deviation shows that the data are more dispersed. The Standard Deviation of VAG signal is also used as a feature. The abnormal VAG signal has more variance around its mean due to its irregular behavior. If we take a mean of Standard Deviation of all abnormal VAG signals and all normal signals, then we get high mean value in case of abnormal signals as compared to normal signals. This make it a suitable feature to enhance the classification accuracy.

3.4.1.10 Auto-Correlation

The degree of similarity between a time series and a lagged version of itself across consecutive time intervals is represented mathematically as autocorrelation. There is a similarity in between autocorrelation and correlation. In Autocorrelation, the same time series is used two times to calculate autocorrelation. One time series is used in its original form and second time series is same but there is a time lag in it. The ACF (τ) of VAG signal is calculated using the following formula:

$$A(\tau) = \frac{1}{N-\tau} \sum_{i=0}^{N-\tau-1} (vag(i)vag(i+\tau)) \tag{20}$$

It is an important parameter that is used in signal processing to look at sequences of values, like time domain signals to analyze the repeating patterns present in the given signal. Sometimes, these repeating patterns represent the presence of a periodic signal hidden by noise or the missing of fundamental frequency in a signal that is deduced from its harmonic frequencies.

In this research, we calculated the Autocorrelation of VAG signals and obtained two features from it. These features are Skewness and Kurtosis. In [9] Rui Gong et al., also employed Auto correlation function based features that are skewness and kurtosis with CWT based features. The formulas of these two features are given below:

$$Skewness = \frac{\sum_{i}^{N}(X_i - \bar{X})^3}{(N-1) * \sigma^3} \qquad (21)$$

Where "N" represents the number of samples, X_i represents the random variable, $\bar{X}$ represents the mean of the samples and "σ" represents the standard deviation.

$$Kurtosis = \frac{\mu^4}{\sigma^4} \qquad (22)$$

Where μ^4 represents the fourth moment about the mean of A(τ) and "σ" represents the standard deviation of A(τ).

In our methodology, these 2 features compare the Autocorrelation function with normal probability density function. They are used to identify the changes in the Autocorrelation function of VAG signals.

3.4.2 Auto-Encoder based Features

Autoencoder is a type of unsupervised machine learning algorithm. It converts the input data signal values effectively into a latent space representation. This representation is reconstructed back using the decoder at output layer to get a data as close as input values. In this way, we get the best salient features of input data that represent our input data. These features are enough for the reconstruction of input signals. A typical Autoencoder is shown below with a single hidden layer.

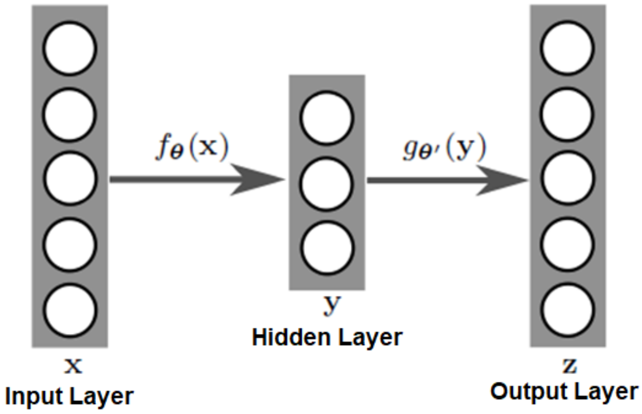

Figure 5: Auto-Encoder

The typical Auto encoder consists of three layers. These layers are known as Input layer, Hidden layer and Output layer. The features are extracted from the Input layer in a Hidden layer. The encoded features are decoded by the decoder. We have used the Hidden layers features which are extracted from the VAG input signals. These features describe the inner behavior of a signal. The size of hidden layer is selected after applying a hit and trial method. The size of hidden layer that gives the best result is 15. So we extract 15 features from the Autoencoder hidden layer.

3.4.3 Continuous Wavelet transform (CWT) based Features

Transforming a time domain signal into Time Frequency Representation, to extract important information about the signal dynamics, has been done by calculating the Continuous Wavelet transform of a VAG signal. The CWT has a similarity with the other type of transform, which is known as Short Time Fourier Transform (STFT). STFT also gives a time frequency distribution but there, it used a fixed duration window function. While in CWT, a wavelet [21] of varying scale and size is used, that slide over the entire time domain signal at different scales and time axis to get the 2 Dimensional time frequency distribution of a signal. The CWT of signal x(t) is defined as [24]:

$$CWT_x(b,a) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} x(t)\; g \, *\left(\frac{t - b}{a}\right) dt \qquad (23)$$

The above formula clearly show the convolution of a given time domain signal "x(t)" with a predefined wavelet window "g(t)". There are two factors in wavelet part; one is a shift in time axis by parameter "b" and second is a dilation by a scale parameter "a". To compare the distribution of STFT and CWT, the parameter "a" should be inversely proportional to frequency. The CWT has a tradeoff between temporal and spectral resolution. At different values of frequencies, it gives different results. The CWT has great spectral resolution but poor temporal resolution at low frequencies. The CWT, on the other hand, gives good temporal resolution but low spectral resolution at high frequencies. Due to this reason, CWT helps to clearly observe the singularities. To observe this phenomenon, Scalogram is used to get the energy density function of a CWT, which is defined as |CWTx(b, a)|^2.

The wavelet used in this research to calculate the CWT of VAG signal is Analytic Morlet (Gabor) Wavelet, named as "amor" in MATLAB. It has equal variance in time and frequency domain. The wavelet is shown in time and frequency domain in Figure 6.

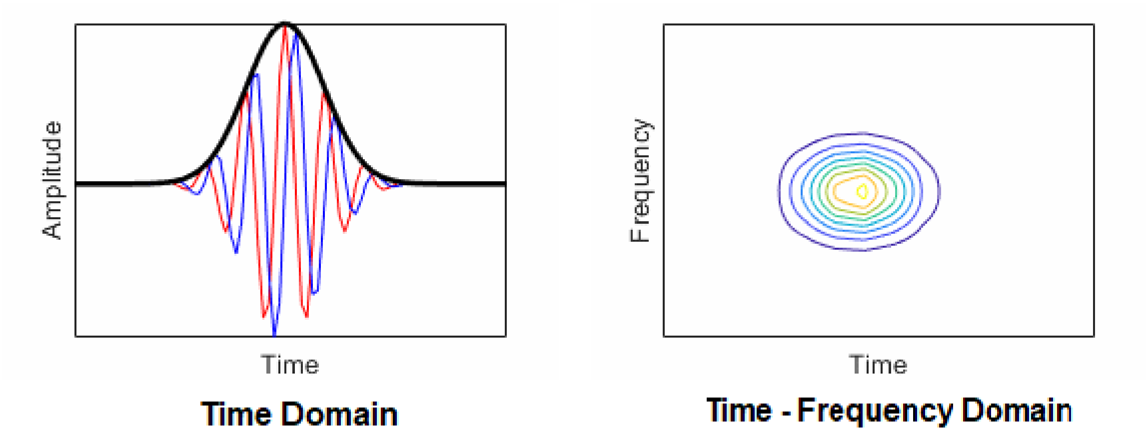

Figure 6: Analytic Morlet (Gabor) Wavelet

There are 6 features that are calculated by CWT of a VAG signals. These six features are the mean of Spectral Slope of CWT, variance of Spectral Slope of CWT, mean of Spectral Flux of CWT, variance of Spectral Flux of CWT, mean of Spectral mean of CWT and variance of Spectral mean of CWT [12].

3.4.3.1 Spectral Slope Feature of Magnitude of CWT

The impulsive components of signals are where the spectral slope reveals important information [12]. The spectral envelope's property known as spectral slope denotes the change in relative intensities of linear component across the signal's frequency range. The linear regression methodology underlies this linear calculation of the spectral slope. Because the power of signals like vibration signals diminishes with increasing harmonic number, they exhibit negative spectral slopes. The signal's low-frequency components have a relatively higher power than its high-frequency ones. The following equation is used to determine the slope of the linear function derived from the magnitude spectrum:

$$v_{SSL}(n) = \frac{\kappa \sum_{k=0}^{\frac{\kappa}{2}-1} k|X(k, \;\; n)| \; - \; \sum_{k=0}^{\frac{\kappa}{2}-1} k. \; \sum_{k=0}^{\frac{\kappa}{2}-1} |X(k, \, n)|}{\kappa \sum_{k=0}^{\frac{\kappa}{2}-1} k^2 \; - (\sum_{k=0}^{\frac{\kappa}{2}-1} k)^2} \qquad (24)$$

The amplitude range of the spectral magnitude determines the final value of the spectral slope. It improves with disappearing higher harmonics and is ideal for noisy troughs. There is zero spectral slope when there is a zero magnitude or equal magnitude at all bins all bins [12].

3.4.3.2 Spectral Flux Feature of Magnitude of CWT

The spectral flux measures the quantitative change in the spectral shape and tell us about the variation between successive CWT frames as given is given by following equation:

$$v_{SF}(n) = \frac{\sqrt{\sum_{k=0}^{\frac{\mathrm{K}}{2}-1} \big(\big|X\big(k, \, n\big)\big|\big) - (|X(k, \, n-1)|)^2}}{\mathrm{K}/2} \qquad (25)$$

According to Zwicker and Fasti, a quasi-periodic transition or a variation of the excitation pattern level can be used to represent the spectral flux. [12]. It can be thought of as a crude approximation of the sensory roughness. The resulting spectrum flux value falls between 0 to SF(n) A, where A is the largest spectral magnitude that may be achieved. The output range of the vibration signal is determined by its normalisation and frequency transformation.

3.4.3.3 Spectral Mean Feature of Magnitude of CWT

The nature and structure of magnitude spectrum of the CWT has been determined by the spectral mean feature [12]. The magnitude spectrum for each bin is averaged to produce the term "spectral mean."

$$v_{SM}(n) = \frac{2}{\kappa}\sum_{k=0}^{\frac{\kappa}{2}-1} | X\left(k,\ n\right) | \qquad (26)$$

When compared and scaled with neighbouring windows of the CWT, empirical observations show that these metrics demonstrate a great deal of resemblance. As a result, although if their values are noticeably different, the spectral mean plots for a signal have a very similar shape, with the exception of a few windows. As a result, if we use find two features from spectral mean that are its arithmetic mean and its variance, then we can discriminate between abnormal and normal signals accurately.

3.5 Feature Selection using Correlation Coefficient

We have extracted different type's features from VAG signals and combined together. Some of the retrieved characteristics are connected with one another due to their spread (variance) in the same direction, implying statistical interdependence to some extent. To remove the overlapping features, we use the correlation coefficients between the features and remove those features which have high correlation. In this way, only unique and independent features left behind, which decreases the complexity and also improves the classification accuracy.

The total number of features are 34 in which selected features are 20. In selected feature set, 4 features are statistical features, 12 are Auto-Encoder based features and 4 features are CWT based features. The remaining unnecessary features have been removed from the feature set as if they employed, they might drastically obstruct the findings of classifier.

The four selected statistical features are Approximate Entropy, Shannon Entropy, Tsallis Entropy and VAG signal Standard deviation. The four selected CWT based features are Spectral Mean (variance), Spectral Flux (mean), Spectral Flux (variance) and Spectral Slope (variance).

3.6 Classification of VAG Signals

The VAG signals are classified into normal and abnormal signals by using different types of classifiers. In this research, we have used Decision Tree and Support Vector Machine (SVM) classifiers to predict the class of each VAG signals. The combination of different types of selected features are combined together and fed into the classifier models for the classification purposes.

3.6.1 Decision Tree Classifier

Decision tree is a type of classifier that is used to classify the datasets. Decision Tree is a machine learning algorithm in which a model is trained by using a decision rules. These decision rules are formed by using the dataset for the training of model. It is like a graphical tree like structure having nodes to predict the class or value of the target variable. The first node is called root node and all the other branches are originated from it like a normal tree growth in real life. This root node expand into different branches, which are based on decision rules. These branches ends on leaf nodes and each leaf node represent the outcome.

Decision Tree is an algorithm of supervised machine learning. It is a graphical tree like structure that is used for classification and Regression problems. It starts with a root node and expand into a different branches. Its branches represent the decision rules and each leaf node represent the outcome.

There are two types of nodes in the typical Decision Tree, first is a “Decision Node” and the second one is a “Leaf Node”. The decision nodes are taken by the Decision nodes and that is why they have several branches. Leaf nodes are the end nodes that are the result of those decisions and have no more branches. The other nodes present in the internal structure of decision tree represent the characteristics of a dataset and the judgments are made based on the characteristics of the provided dataset for training.

3.6.2 Support vector Machine (SVM)

The Support Vector Machine (SVM) is used for classification of data into different labels. It is a machine learning algorithm trained by the features of given training dataset. These optimized features help SVM in classification of binary classes. In our case, SVM classifier used the selected features for the segregation of given abnormal and normal signals using a 2D hyper-plane. The hyperplane is adjusted in such a way that it maximizes the distance to nearest data-points from the decision boundary. These nearest data-points are called support vectors. If there is more distance between the support vectors and the decision boundary, the more accurate results can be achieved.

3.6.3 Ten-fold cross-validation

Machine learning models involves the overfitting problems when the dataset quantity is small. This would lead us to inaccurate results. To overcome this problem, k-fold cross-validation is employed while applying the machine learning algorithms for training and testing.

In this research, we used 10-fold cross validation on our data set to get better quality results from the machine learning models. In this process, the overall dataset breaks into 10 equal parts for training of algorithm and its testing purpose [22]. In which one part is used for testing and the other nine parts are used for training of model in each iteration. This process in repeated ten times, in which every time the different testing data is chosen among the parts of data set. The final result is an average of ten iteration, which is the reliable and high quality result [23].

Figure 7: Working of 10-Fold Cross Validation

CHAPTER 04

RESULTS AND DISCUSSION

In this chapter, we will discuss the results of our proposed methodology. The results include the images of signals and the classifications results. The outputs we achieved by employing various techniques and algorithms are shown and described in this chapter.

4.1 A Raw Abnormal and Normal VAG Signal

4.2 Original Signal Vs Smoothed Signal

4.3 Wavelet Decomposed Sub-bands and Reconstructed Signal

4.4 Performance Evaluation Parameters (Accuracy, Sensitivity, Specifity, PPV, NPV, MCC Score)

4.5 Classification Results (Includes Confusion matrices and ROC Curves)

4.1 A Raw Abnormal and Normal VAG Signal

The dataset we have used in this research has two types of signals. One is Abnormal VAG signals that show the knee pathology and the second is Normal VAG signals that show the normal knee. All the computations that involved in signal preprocessing, extraction of features, selection of suitable features, and classification has been implemented using MATLAB (R2020a).

4.2 Original Signal Vs Smoothed Signal

The Raw VAG signals contains various interference of noises and baseline wanders. These interferences reduce the quality of VAG signals and these VAG signal cannot be used to get high accurate results. In order to solve this problem, we have first do pre-processing of signals that includes smoothing the VAG signals with an appropriate Moving Average Filter. Smoothing of VAG signals helps to capture the important data patterns and noise removal in signals. The smoothing filter is used in MATLAB named as 'movmean'.

The following figure show the step-wise pre-processing of normal VAG signals. Fig. 6 shows a sample raw VAG signal, and Fig. 7 represents the smoothed VAG signal for a healthy subject.

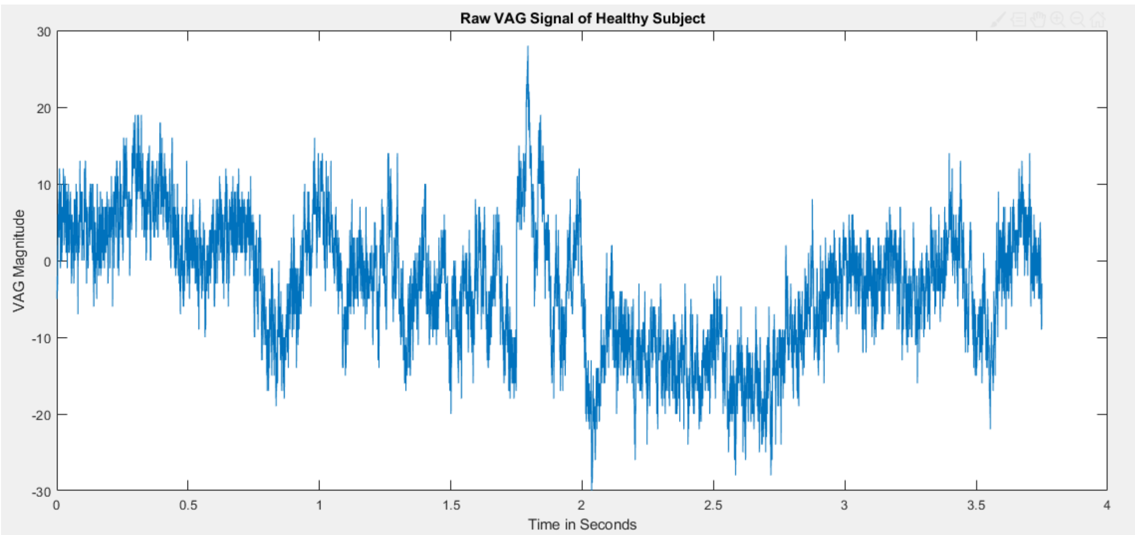

Figure 8: Raw VAG signal of a Healthy subject

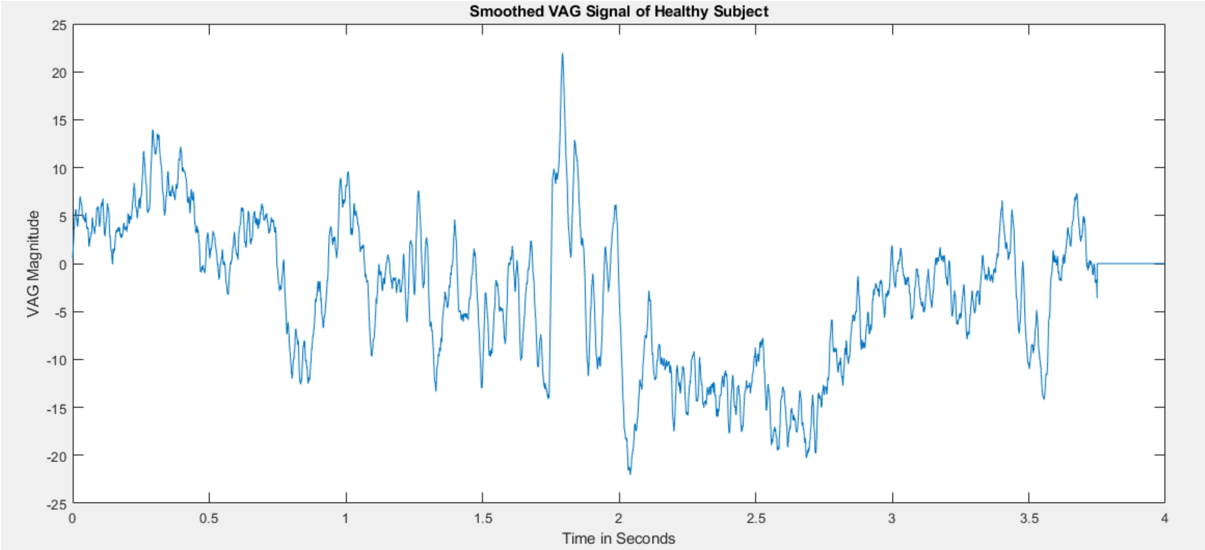

Figure 9: Smoothed VAG signal of a Healthy subject

In the same way, Fig. 8 shows a sample raw VAG signal, and Fig. 9 represents the smoothed VAG signal for an Unhealthy subject.

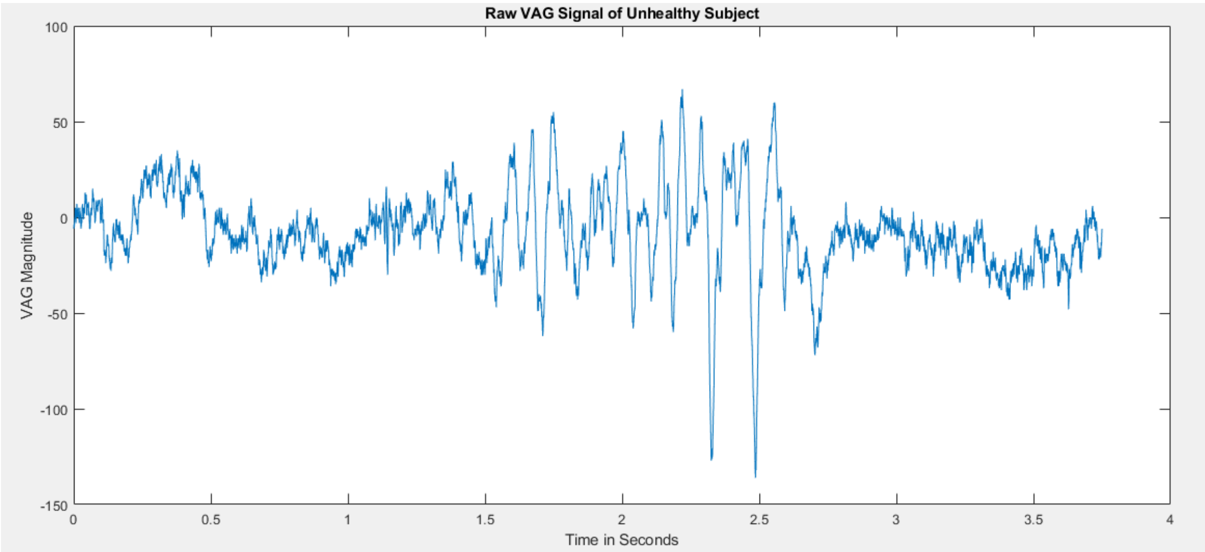

Figure 10: Raw VAG Signal of an Unhealthy Subject

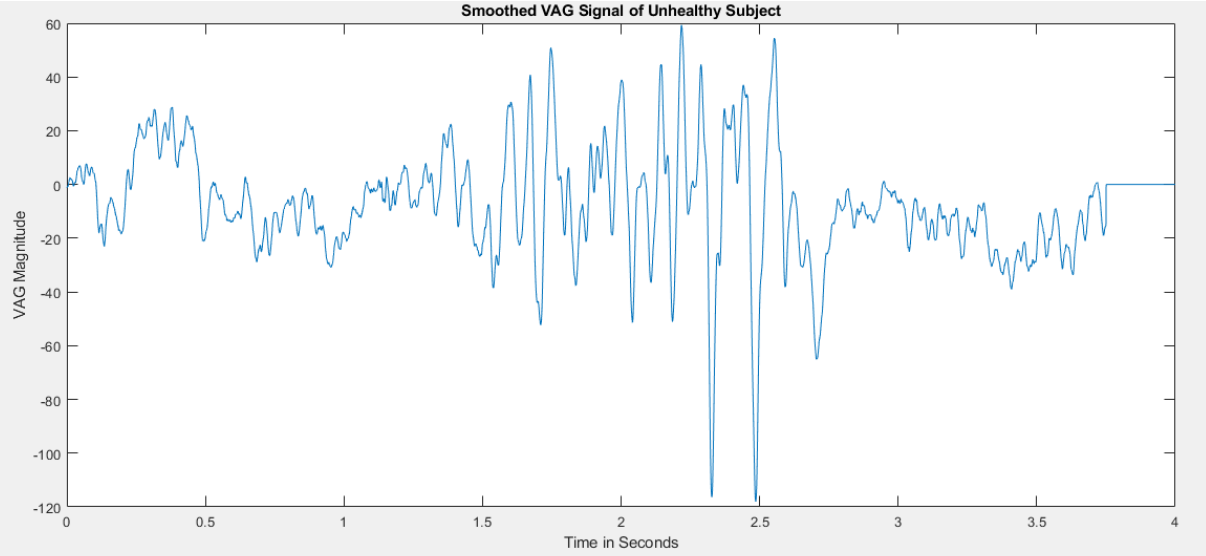

Figure 11: Smoothed VAG Signal of an Unhealthy subject

4.3 Wavelet Decomposed Sub-bands and Reconstructed Signal

The smoothed VAG signals are further processed so that we can remove any type of noise and baseline wanders as much as possible. Wavelet Decomposition is the best method to get the high quality VAG signals. This algorithm decomposes the VAG signal into the detail coefficients and approximations [24]. These coefficients are then further split into approximation and coefficients. As a result, the signal disintegrates like a tree. These coefficients belong to a specific frequency at each stage.

Figure 12, represents the sub-band signals obtained from Wavelet Decomposition of a VAG signal for a healthy subject.

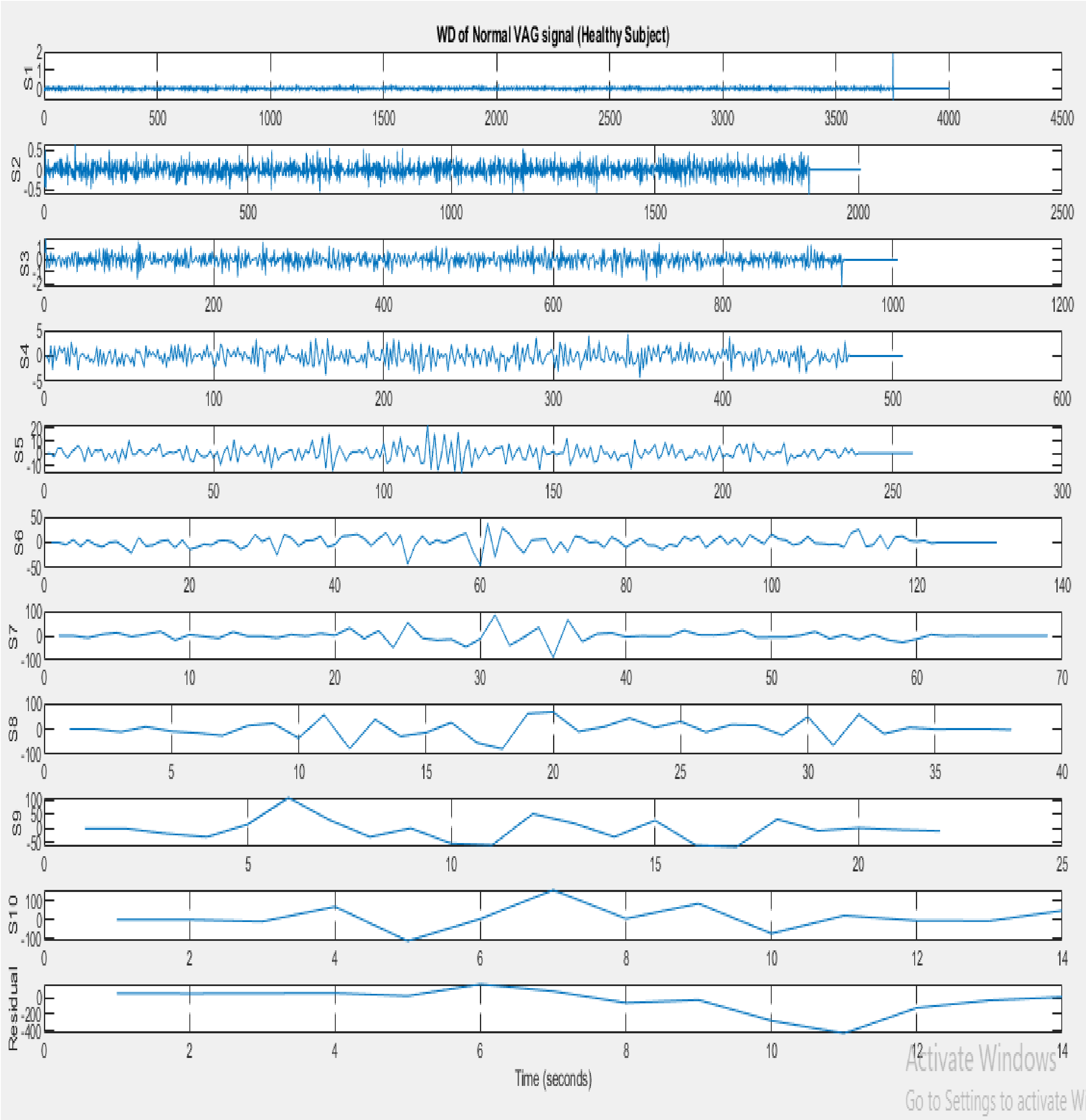

Figure 12: WD of a Healthy VAG subject

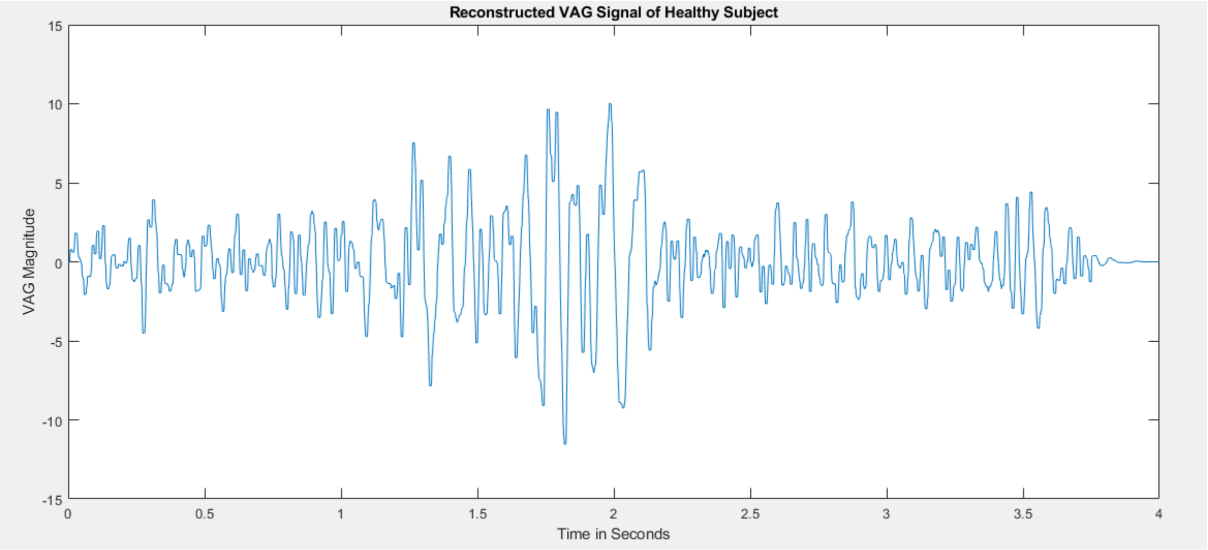

Figure 13: Reconstructed VAG Signal of a Healthy Subject

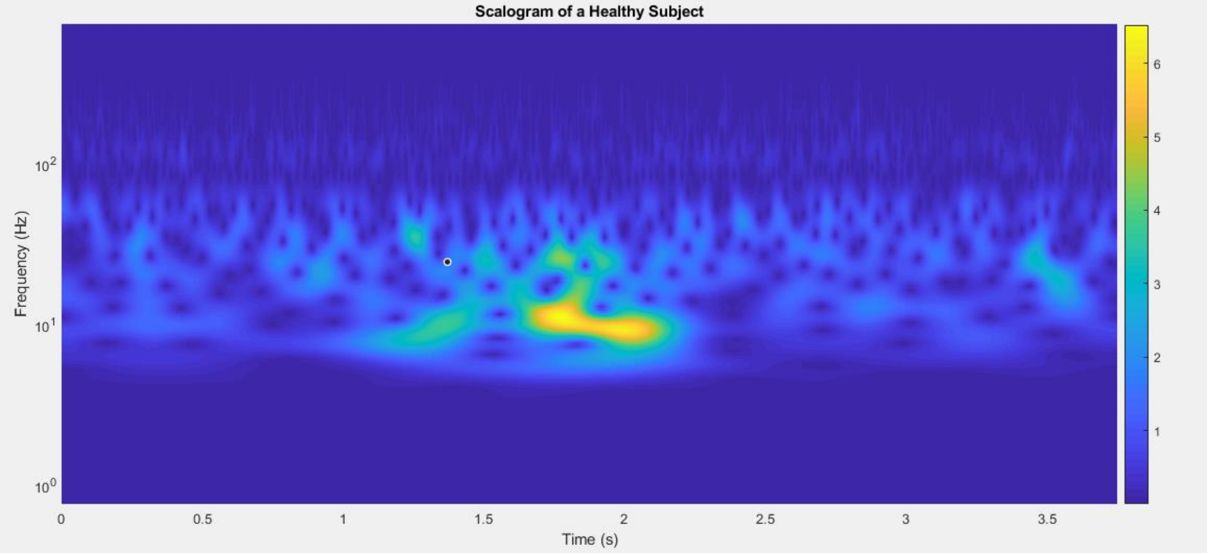

Figure 14: CWT of a Healthy Subject

Figure 15 represents the sub-band signals obtained from Wavelet Decomposition of a VAG signal for an Unhealthy subject.

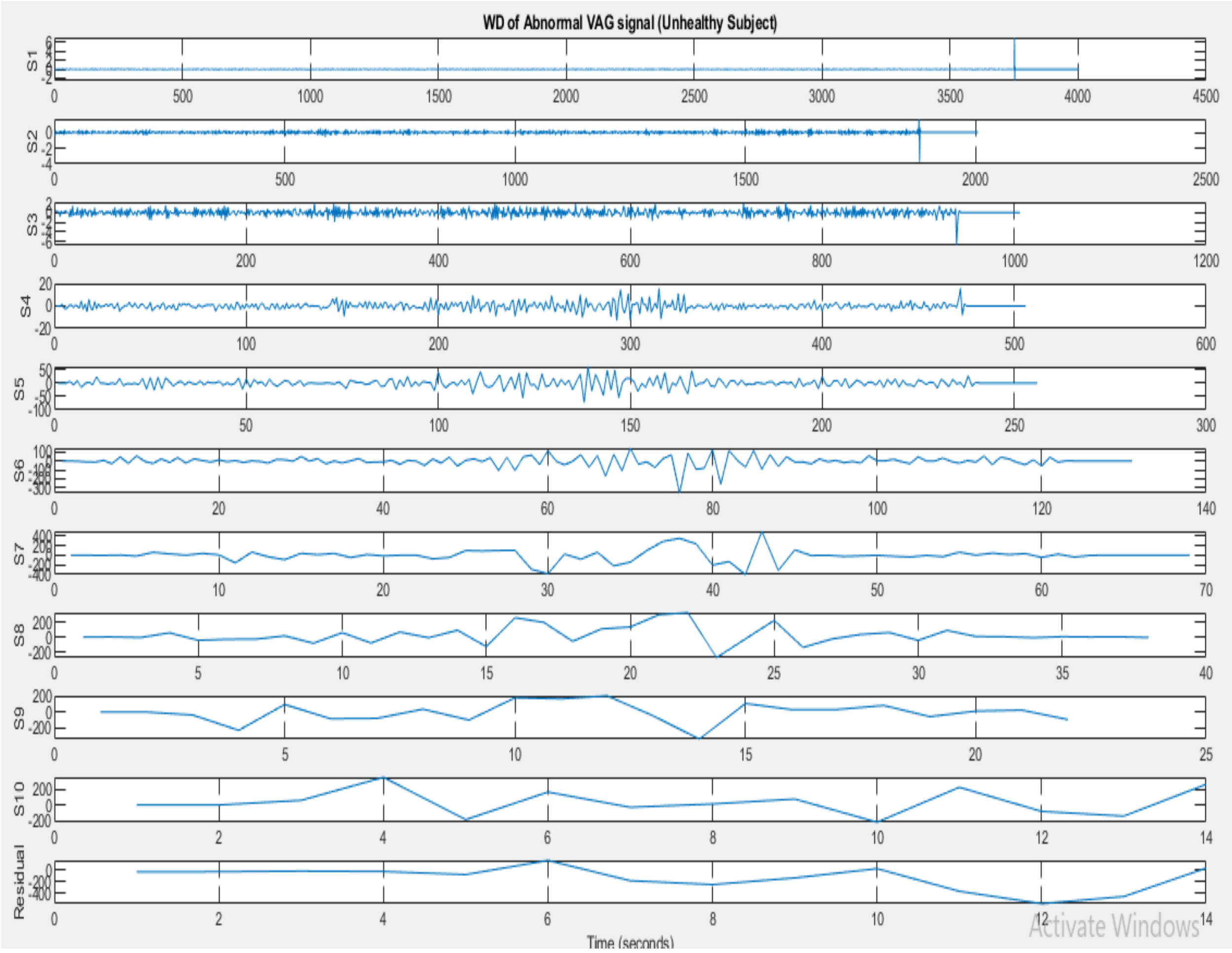

Figure 15: WD of an Unhealthy VAG subject

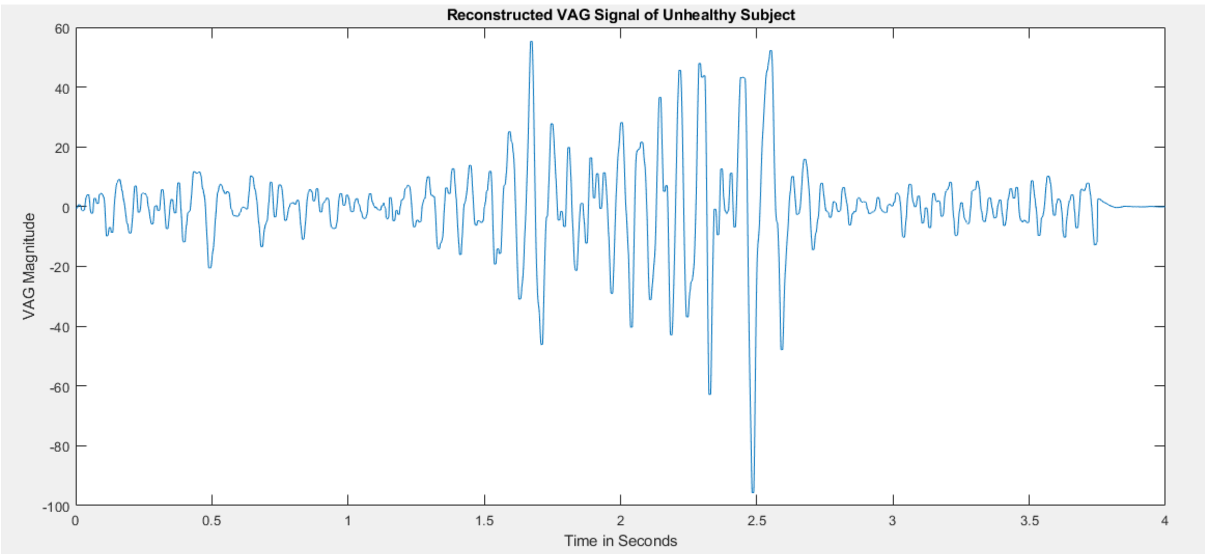

Figure 16: Reconstructed VAG Signal of an Unhealthy Subject

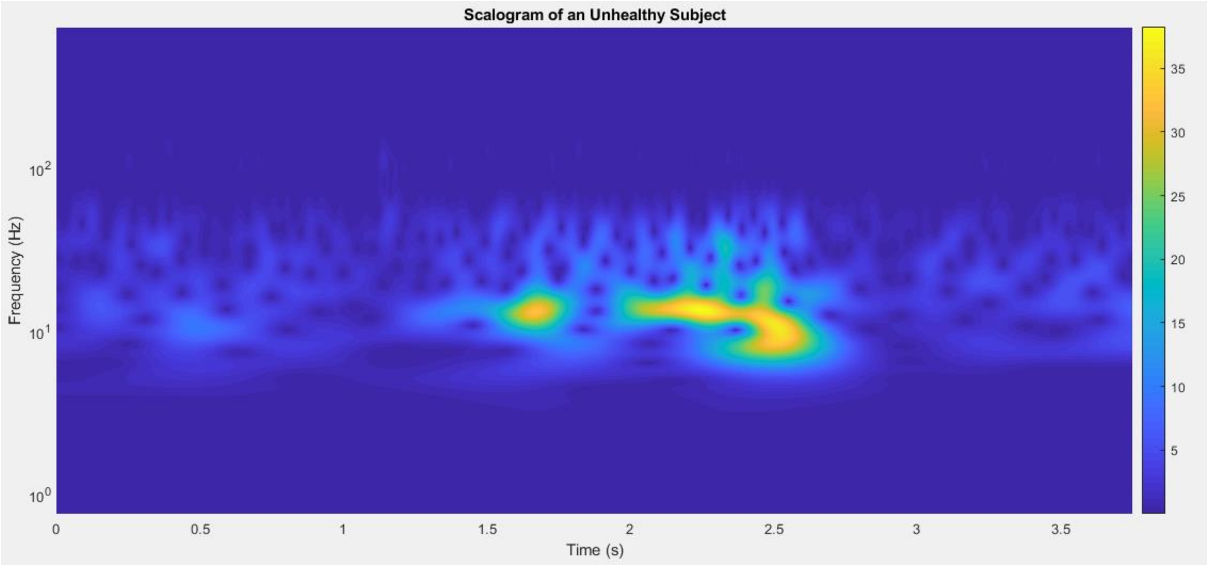

Figure 17: CWT of an Unhealthy Subject

4.4 Performance Evaluation

The classification findings are provided in terms of various parameters to assess how well the suggested approach for classifying VAG signals performed. The performance of the classification model is judged by using these parameters. These parameters tell the efficiency of algorithm and evaluate the performance of proposed methodology. Number of performance parameters are acquired from the Confusion Matrix obtained from Decision Tree and SVM classifiers.

These parameters are Accuracy, Sensitivity, Specificity, Negative Predicted Values (NPV), Positive Predicted Values (NPV), and Matthews Correlation Coefficient (MCC). By contrasting actual values with the predictions made by the model, confusion matrix is used to assess a model's performance. Because our program needs binary categorization, which has two classes—normal and abnormal—we employed a 2x2 confusion matrix. The first row of the 2x2 confusion matrix shows the true positives (TP) and false negatives (FN), and their sum shows the overall number of anomalous signals we have taken into consideration. Similar to the first row, the second row tabulates the false positives (FP) and true negatives (TN), the sum of which represents all of the normal signals we have employed.

The subjects represented by TP and TN are those that the classifier correctly identifies as abnormal and normal, respectively. In contrast, FP stands for normal participants that were mistakenly classed as abnormal, while FN stands for abnormal subjects that were mistakenly labelled as normal. There are no FP and FN in an ideal model (where accuracy is 100%). In addition, other model assessment parameters that are provided in the result tables after the confusion matrices are computed using the TP, TN, FP, and FN values. The specificity measure shows how many subjects are incorrectly labelled as abnormal when they are truly normal. The sensitivity metric, on the other hand, enables us to determine the proportion of abnormal participants who are accurately identified as having abnormal knee joints.

Sensitivity:

It is the proportion of abnormal subjects successfully diagnosed by the algorithm. It is also known as Recall.

$$\text{Sensitivity} = \frac{TP}{TP+FN} \tag{27}$$

Specificity:

It is the percentage of normal participants successfully identified by the algorithm.

$$\text{Specificity} = \frac{TN}{TN+FP} \tag{28}$$

Accuracy:

It is the proportion of correctly classified subjects as a percentage of all subjects considered for classification.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{29}$$

Positive Predicted Value (PPV)

It is the proportion of True Positives to all positive test findings. This includes healthy participants who were misdiagnosed as patients. It also tells us the precision of the model.

$$\text{Positive Predicted Value (PPV)} = \frac{TP}{TP+FP} \tag{30}$$

Negative Predicted Value (NPV)

The percentage of predicted negatives that actually occur is the negative predictive value in machine learning. It displays the likelihood that a predicted negative really is a negative.

$$\text{Negative Predicted Value (NPV)} = \frac{TN}{TN+FN} \tag{31}$$

Mathews Correlation Coefficient (MCC)

The biochemist Brian W. Matthews first proposed the Matthews correlation coefficient (MCC) in 1975. It is now often referred to as the MCC in machine learning, as a metric for the accuracy of binary (two-class) classifications. The Matthews correlation coefficient (MCC), on the other hand, is a more trustworthy statistical indicator. It only produces a high score if the prediction performed well in each of the four confusion matrix categories (true positives, false negatives, true negatives, and false positives), in proportion to the size of the dataset's positive and negative components.

$$\text{MCC} = \frac{(TP*TN-FP*FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{32}$$

Receiver Operating Characteristics (ROC) Curve

Another crucial measure for assessing performance is the receiver operating characteristic (ROC) curve, which is frequently employed for validating classification models. The ROC curve [25] is a graph that contrasts the proportion of true positives and false positives. We can compare the performance of several classification models, including Decision Tree and SVM, using the area under the ROC curve. The ability of a feature or trained model to distinguish between two or more diagnostic categories is visually represented by the area under the ROC curve. A wider region encircled by the curve is typically linked to higher classification accuracy [26].

4.5 Classification Results

The classification model evaluation parameters values are tabulated in Tables 1. Table 1 provides the classification performance of Decision Tree and Support vector machine (SVM) using different selected features as a feature set.

From the Table 1, we can see that the Decision Tree model that is trained with the feature set selected by the Correlation Coefficient gives the highest level of accuracy of 93.26%. Decision Tree also have high value Sensitivity of 86.84%, Specificity of 98.04%, PPV of 97.06%, NPV of 90.91% and a Matthews Correlation Coefficient (MCC) score of 0.8641 as compared to

SVM model. The SVM Model gives us the accuracy of 87.64%, Sensitivity of 84.21%, Specificity of 90.19%, PPV of 86.49%, NPV of 88.46% and a Matthews Correlation Coefficient (MCC) score of 0.7468. The AUC-ROC curve for Decision Tree has a value of a 0.91 and for SVM AUC-ROC value 0.93. This shows that the proposed methodology outperformed in maximum performance evaluation parameters when compared to other research works.

Performance Parameters	Decision Tree Results	SVM Result
Accuracy	0.9326	0.8764
Sensitivity (Recall)	0.8684	0.8421
Specificity	0.9804	0.9019
PPV (Precision)	0.9706	0.8649
NPV	0.9091	0.8846
MCC	0.8641	0.7468
AUC-ROC	0.91	0.93

Table 1: Classification Results of DT & SVM

.

The ROC charts of Decision Tree and SVM classifiers are shown in Figure 18 and Figure 19

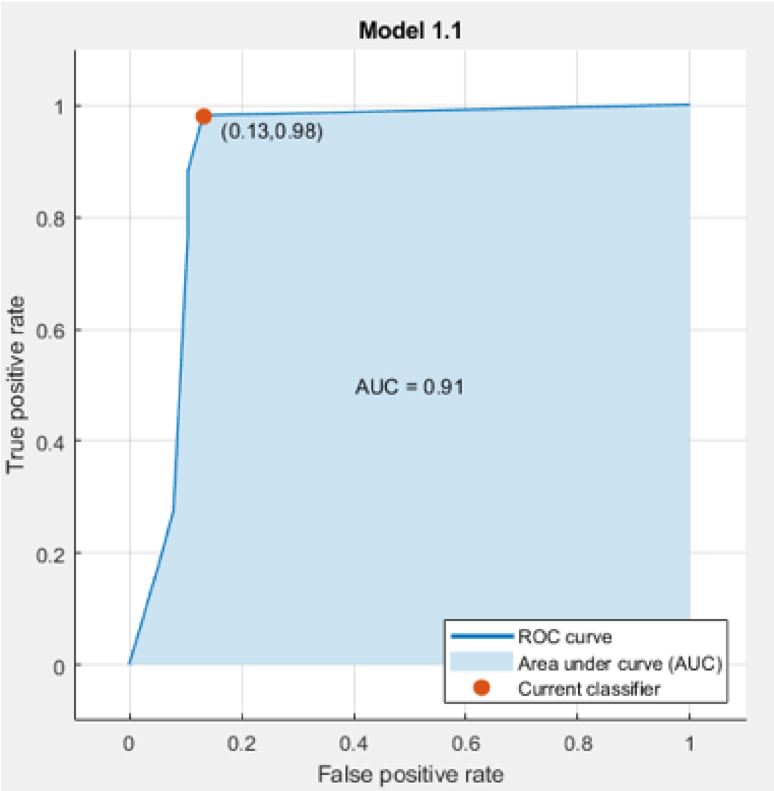

Figure 18: ROC Curve for the DT Classifier

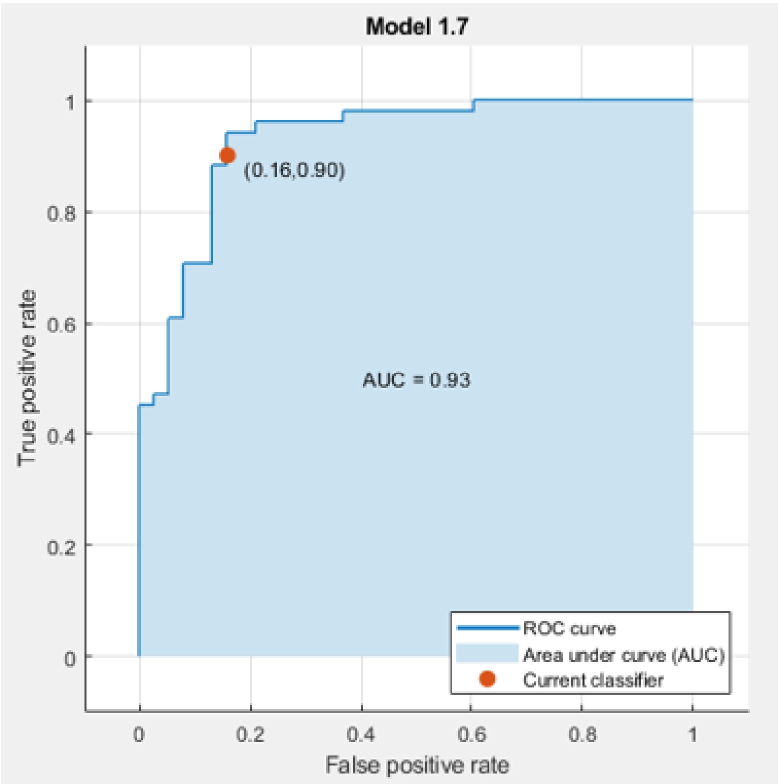

Figure 19: ROC Curve for the SVM Classifier

The confusion matrices of Decision Tree and SVM classifiers are shown in Fig. 20 and Fig. 21.

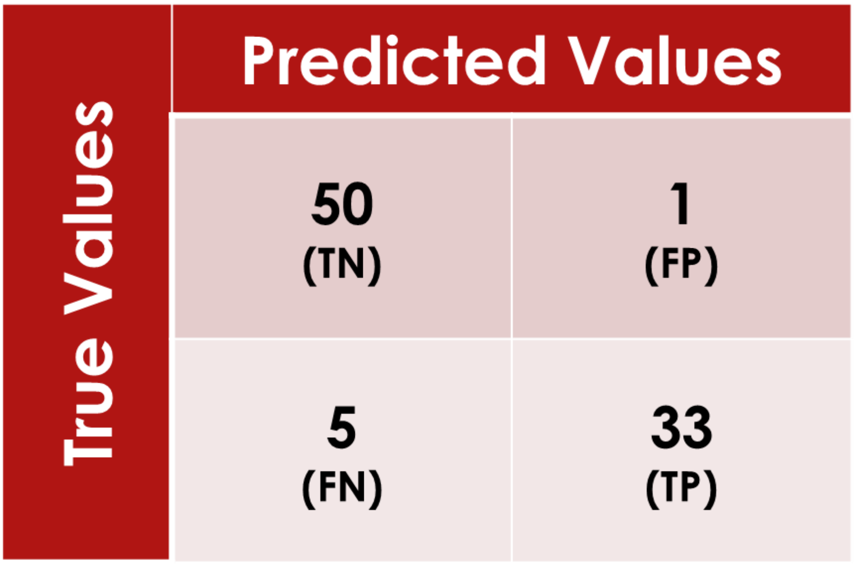

Figure 20: Confusion Matrix of the DT Classifier

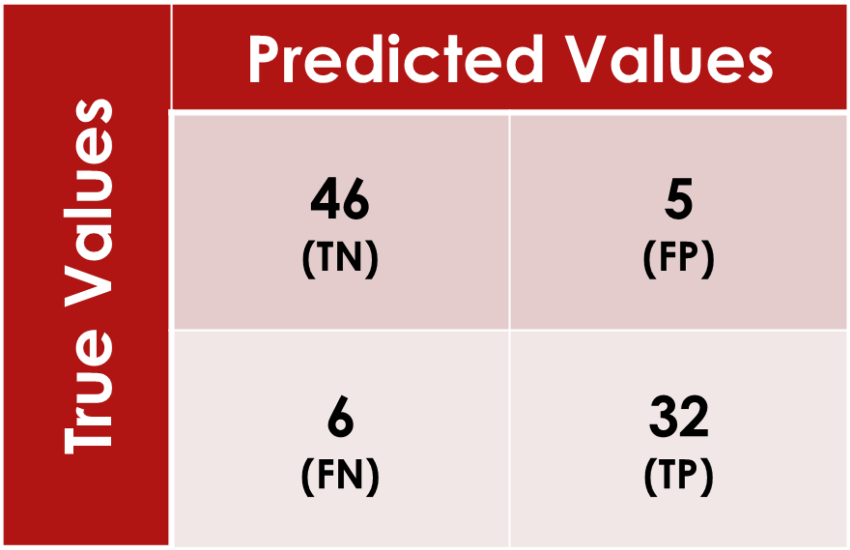

Figure 21: Confusion Matrix of the SVM Classifier

4.6 Comparison of proposed methodology with previous literature

In previous researches, mostly research bound to use specific type of features like Entropy Based Features or Spectral Domain Features. Due to this reason, the accuracy of classifier is limited to specific range like from 84 to 89 percentage. In order to increase the accuracy of classification model, researchers used Deep Learning Methods that includes Convolutional

Neural Networks to extract thousands of features that are used to classify abnormal and normal signals. This method gives accurate results upto 91 or 92 percent but with a heavy cost. It takes more time and computational power for training of Convolutional Networks. It is also tedious to work with thousands of features and used them for classification.

In [12], the highest accuracy achieved without using Deep Learning Methods is about 89.23%. But here the researchers reduced the size of dataset from 89 VAG signals to 65 VAG signals. This has been done by using Standard Deviation thresholding. In this way, all those signals are removed that show anomalous behavior and very hard to classify them.

In our research, we have used all the datasets of 89 signals including the difficult signals. The accuracy achieved in our case is 93.26%. The reason behind the high accuracy is that we combine different types of features, which includes Statistical Features, Auto-Encoder Based features and CWT based features. Auto-Encoder Based features were not used previously in any known research and play an important role in achieving a high accuracy. The Combination of these features help to analyze the signals from every possible way whether it's a time domain or frequency domain.

Correlation Coefficients between features help us to remove those features that show strong correlation between them. After checking all the features, the feature set is reduced from 34 features to 20 features. These selected features quantify the irregular behavior of abnormal signals and separate them from normal signals. The reduced feature set is fed into different classifiers for classification of abnormal and normal VAG signals. The most accurate results are obtained by Decision Tree and SVM. In this way, we have achieved the high classification accuracy without using the Deep Learning Methods for feature extraction.

The comparison of our proposed methodology and previous researches are present in Table 2. This table briefly describe what type of Methodologies they have used for Signal Pre-processing, what types of features they are extracted, type of classifiers models along with their publishing year and source.

Source	Methodology	Features	Classifier	Accuracy	Publishing Year
[4]	EMD & DFA	6 Features (EA-Mean, EA-SD, EA-RMS, Symbolic, Approximate % Fuzzy Entropy)	QDA, SVM, GLRA	QDA: 82.2% SVM: 0.83.7% GLRA: 0.86.3%	2016
[5]	EEMD & DFA	Spectral, Tsallis & Permutation Entropy	RF	RF:86.52%	2016
[6]	CEEMDAN & DFA	Sample, Approximate, Renyi, Tsallis Permutation & Shannon Entropy	SVM	SVM: 86.61%	2017
[9]	EMD & DFA	4 CWT Coefficients, Kurtosis & Skewness of ACF	LS-SVM	LS-SVM:86.67%	2020
[12]	CMAF & SD Thresholding (65 Signals)	Time & Spectral Domain Features	RF, SVM, GNB	RF: 89.23%	2021
Proposed System	Smoothing & WD (89 Signals)	Statistical Features, Auto-Encoder Features, CWT based Features	Decision Tree, SVM	D. Tree: 93.26% SVM: 87.64%	-

Table 2: Comparison of Proposed Methodology with Previous Literatures

CONCLUSION

This study proposed a new methodology that uses different features for the classification of Vibroarthrographic (VAG) signals into abnormal and normal signals. The different types of statistical features are combined with Auto-Encoder based features and Continuous Wavelet Transform Based Features. The features are selected using the Correlation Coefficient to reduce the feature dataset. It gives us the highly optimized features that distinguished normal and abnormal VAG signals with a greater accuracy. The two classifiers are used for classification purposes that are Decision Tree and Support Vector Machine.

Results shown that the Decision Tree Classifier performs best as compared to SVM Classifier that gives us the accuracy value of 93.26%. The proposed methodology also performed very well as compared to previous researches discussed in Introduction. The results of our methodology shows that we can achieve high accurate results by using the combination of Statistical features, Auto-Encoder based features and CWT based features without going into the Deep Learning Methods, which are complex and time consuming. This model will help the doctor to give us best results in a very less time and help them in early treatment of knee disorders.

REFERENCES

[1] A., Balajee & Rajagopal, Venkatesh. (2021). Machine Learning based Identification and Classification of Disorders in human Knee Joint - Computational Approach. 10.21203/rs.3.rs591767/v1.

[2] Chen J-C, Tung P-C, Huang S-F, Wu S-W, Lin S-L, Tu K-L. Extraction and screening of knee joint vibroarthrographic signals using the empirical mode decomposition method. Int J Innovative Comput Inf Control 2013;9(6):2689–700.

[3] Y. Wu, S. Cai, S. Yang, F. Zheng, N. Xiang, Classification of knee joint vibration signals using bivariate feature distribution estimation and maximal posterior probability decision criterion, Entropy 15 (4) (2013) 1375–1387.

[4] Wu, Yunfeng & Chen, Pinnan & Luo, Xin & Huang, Hui & Liao, Lifang & Yao, Yuchen & Wu, Meihong & Rangayyan, Rangaraj. Quantification of Knee Vibroarthrographic Signal Irregularity Associated with Patellofemoral Joint Cartilage Pathology Based on Entropy and Envelope Amplitude Measures. Computer Methods and Programs in Biomedicine (2016). 130. 1-12. 10.1016/j.cmpb.2016.03.021.

[5] S. Nalband, R. Sreekrishna, A.A. Prince, Analysis of knee joint vibration signals using ensemble empirical mode decomposition, Proc. Comput. Sci. 89 (2016) 820–827.

[6] S. Nalband, A. Prince, A. Agrawal, Entropy-based feature extraction and classification of vibroarthographic signal using complete ensemble empirical mode decomposition with adaptive noise, IET Sci. Meas. Technol. 12 (3) (2017) 350–359.

[7] S. Nalband, A. Sundar, A.A. Prince, A. Agarwal, Feature selection and classification methodology for the detection of knee-joint disorders, Comput. Methods Programs Biomed. 127 (2016) 94–104.

[8] Rangayyan, R. M. and Wu, Y. (2007). Analysis of knee-joint vibroarthrographic signals using statistical measures. In Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07), pages 377–382. IEEE.

[9] R. Gong, K. Hase, H. Goto, K. Yoshioka, S. Ota, Knee osteoarthritis detection based on the combination of empirical mode decomposition and wavelet analysis, J. Biomech. Sci. Eng. 15 (3) (2020), 20-00017.

[10] Wu, Y. and Krishnan, S. (2011). Combining least-squares support vector machines for classification of biomedical signals: a case study with knee-joint vibroarthrographic signals. Journal of Experimental & Theoretical Artificial Intelligence, 23(1):63–77.

[11] Maussavi, Z. M., Rangayyan, R. M., Bell, G. D., Frank, C. B., and Ladly, K. O. (1996). Screening of vibroarthrographic signals via adaptive segmentation and linear prediction modeling. IEEE transactions on biomedical engineering, 43(1):15.

[12] Shidore, Mrunal & Athreya, Shreeram & Deshpande, Shantanu & Jalnekar, Rajesh. (2021). Screening of knee-joint vibroarthrographic signals using time and spectral domain features. Biomedical Signal Processing and Control. 68. 102808. 10.1016/j.bspc.2021.102808.

[13] S. Nalband, C. Valliappan, A.A. Prince, A. Agrawal, Time-frequency based feature extraction for the analysis of vibroarthographic signals, Comput. Electr. Eng. 69 (2018) 720–731.

[14] A., Balajee & Rajagopal, Venkatesh. (2021). Machine Learning based Identification and Classification of Disorders in human Knee Joint - Computational Approach. 10.21203/rs.3.rs-591767/v1.

[15] Y. Athavale, S. Krishnan, A telehealth system framework for assessing knee-joint conditions using vibroarthrographic signals, Biomed. Signal Process. Control 55 (2020) 101580.

[16] S. Nalband, C. Valliappan, R.G.A. Prince, A.A. Agrawal, Feature extraction and classification of knee joint disorders using hilbert huang transform, in 2017 14th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), IEEE, 2017, pp. 266–269.

[17] S. Nalband, C. Valliappan, A.A. Prince, A. Agrawal, Time-frequency based feature extraction for the analysis of vibroarthographic signals, Comput. Electr. Eng. 69 (2018) 720–731.

[18] Y. Wu, S. Krishnan, R.M. Rangayyan, Computer-aided diagnosis of knee-joint disorders via vibroarthrographic signal analysis: a review, Crit. Rev.™ Biomed. Eng. 38 (2) (2010).

[19] Y. Wu, S. Yang, F. Zheng, S. Cai, M. Lu, M. Wu, Removal of artifacts in knee joint vibroarthrographic signals using ensemble empirical mode decomposition and detrended fluctuation analysis, Physiol. Meas. 35 (3) (2014) 429.

[20] Parzen, E. (1962). On estimation of a probability density function and mode. The annals of mathematical statistics, 33(3):1065–1076.

[21] Mallat S. A wavelet tour of signal processing the sparse way. Academic press; 2008. 9780123743701.

[22] Y. Song and P. Liò, "A new approach for epileptic seizure detection: sample entropy-based features extraction and extreme learning machine," Journal of Biomedical Science and Engineering, vol. 3, pp. 556-567, 2010.

[23] https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623

[24] F. Hlawatsch and G. F. Boudreaux-Bartels, “Linear and Quadratic Time- Frequency Signal Representations,” IEEE Signal Processing Magazine, pp. 21- 67, 1992.

[25] M.H. Zweig, G. Campbell, Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine, Clin. Chem. 39 (4) (1993) 561–577.

[26] M. Greiner, D. Pfeiffer, R. Smith, Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests, Prev. Vet. Med. 45 (1–2) (2000) 23–41.

[27] M.C. Eguia, M.I. Rabinovich, H.D.I. Abarbanel, Informationtransmission and recovery in neural communicationchannels, Phys. Rev. E 62 (5) (2000) 7111–7122.