CHAPTER 3

RESEARCH METHODOLOGY AND APPROACH

Pathological voice recognition has drawn much attention in the last decade. Speech processing is an outstanding tool for voice disorder detection and classification. The most interesting works are focused on the Parkinson's Disease (PD), Multiple Sclerosis (MS) and other diseases that belong to the class of neurodegenerative diseases affecting the patients' speech, motor and cognitive capabilities (Davis 1979).

Speech analysis is complex and has been ignored for an extended period of time. Voice pathologies are a consequence of improper use of voice, stress, inhalation of tobacco smoke, gastric reflux and hormonal problems. This diseases typically affect the vocal folds and are detected by direct laryngoscopy, which visualizes the vocal folds using a camera. This method is aggressive and painful for the patient depending on the equipment used, requiring a local anesthetic procedures. The Pathologist may perform miss diagnosis, due to tiredness or overworked and if there is a huge amount of data to examine. So, miss diagnosis is a major drawback and accuracy is very critical

The test can also be performed by indirect laryngoscopy using a mirror which is less aggressive, involving fewer amounts of local sedatives. But this equipment is expensive and involves high maintenance costs. Assessing the features of a voice disorder aids the pathologist to identify the voice disorder, decide the form of treatment and formulate a prognosis. An efficient, inoffensive, easy and low-priced method for pathologies recognition may help in initial evaluation and serve as a complementary method for the voice disorder diagnosis.

In order to overcome the above drawbacks, an attempt was made in this research and a model is developed to discriminate between normal and pathological voices and to carry out automatic screening of voice disorders. The research methodology of the proposed research work is designed to build a system that can automatically discriminate between voices as normal /pathological

3.1 Research Design

The research methodology proposes various techniques that are designed to find optimized solutions to meet the research objectives. The solutions provided in this research work are more feasible for discriminating the voices. Acoustic analysis is used to study the speech signals and help to extract the relevant features. In this research, for this purpose, a set of classification algorithms are analyzed and a method is proposed to find the optimal solution to the task undertaken. This research design in Figure 3.1 shows the proposed research contribution and enhanced techniques in Voice Pathological Identification System.

This research process works based on the phases, as following,

Phase 1: Preprocessing: To produce a parametric representation.

The First Phase focuses on silence and noise removal steps, here wiener and Discrete wavelet transformation filters are combined and proposed as Hybrid wiener and Discrete wavelet transformation HWFDWT to produce preprocessed speech

Phase 2: Feature Selection and Extraction: To select a subset of relevant variables and predictors, and to discard redundant and unwanted information.

The second phase deals with enhancing feature selection & Extraction process by the proposed method CSOMFCC, by applying Cat Swarm Optimization technique with MFCC coefficients to reduce dimensionality and execution time

Phase 3: Classification: To classify the mixed voiced data set into normal and pathological voices.

The Third phase of the work focuses on algorithms that improve classification accuracy and time with the proposed Modified BPNN algorithm

Phase 4: Performance Result: The ROC Curve plotted to distinguish between Pathological voices classification Accuracy and Pathology Diseases.

To evaluate the proposed model, various performance metrics used are Signal to noise ratio, Accuracy, specificity, sensitivity and time and finally, Roc Curve is plotted.

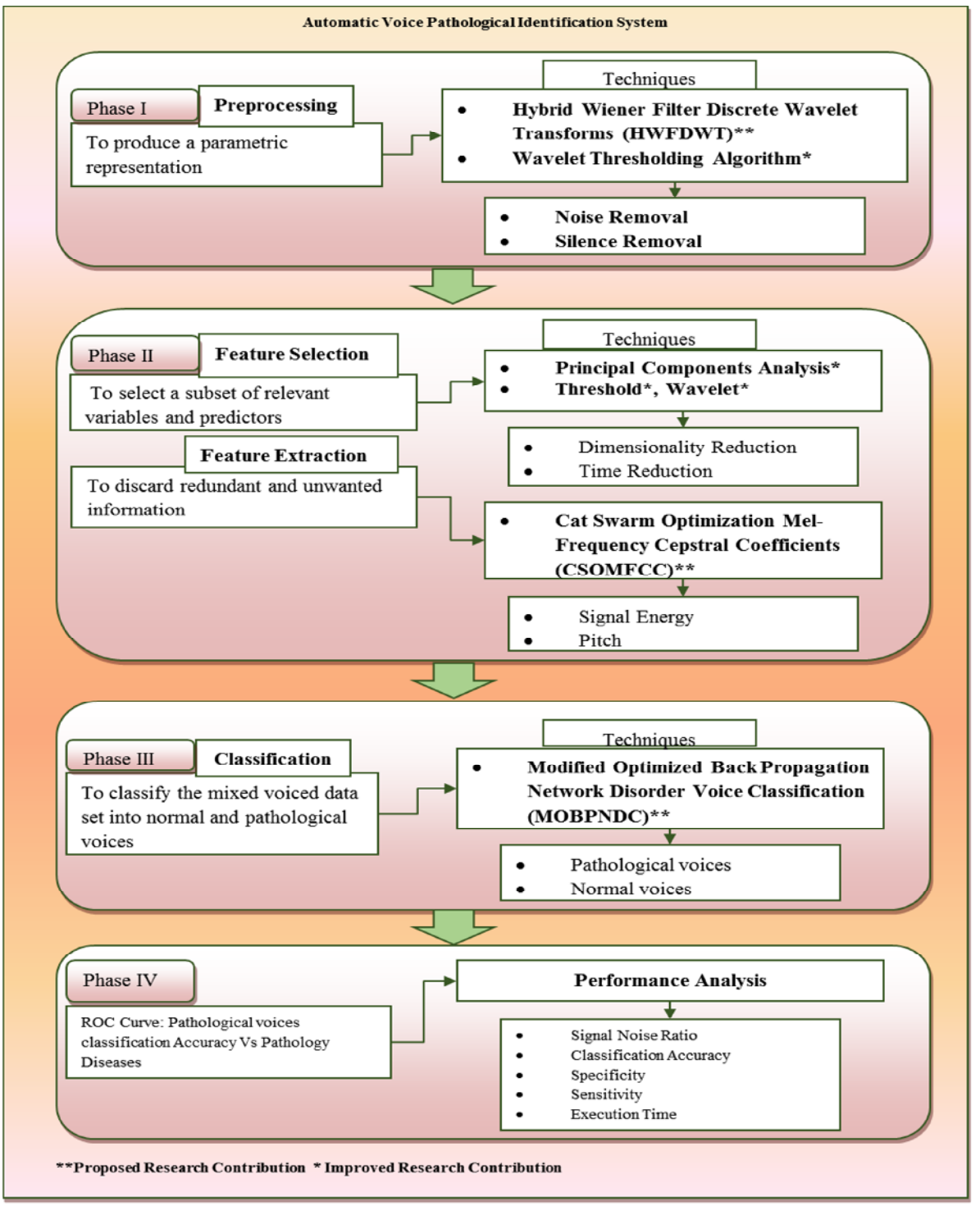

Figure 3.1 Proposed Research Design

3.2 Phase 1: Preprocessing

The voice pathology disorder analysis from the various input speech signal is carried out by preprocessing performances. The vital role of preprocessing is considered as using Noise Removal and Silence Removal techniques. This helps to find the acoustic parameters by Electro Glotto Graph (EGG) with the help of Wiener Filter and DWT Filters. The Wiener Filter and DWT Filters run the Electro Glotto Graph (EGG) of each input speech signal. For this concern, this research proposed the Hybrid Wiener Filter Discrete Wavelet Transforms (HWFDWT) measure is to calculate the enhanced speech signal from the input speech signal. This preprocessing progression is labeled as a denoising estimation in voice pathology disorder identification process. The proposed preprocessing method comprises more computations in comparison to the active system. But, in terms of Noise Removal and Silence Removal, it performs well. Perhaps the Wiener filter minimizes the Mean Square Error between the estimated random process and the desired process by Linear Time-Invariant (LTI) filtering. Similarly, the Discrete wavelet transform (DWT) produced the discrete wavelets sampled over Fourier transforms. This present work combines both the "Wiener filter minimization" as well as "discrete wavelets sample" called Hybrid Wiener Filter Discrete Wavelet Transforms (HWFDWT).

This research scrutinizes the input voice signals from other noises and silence. Many researchers tried to scrutinize the Noise and Silence removal from the original voice. Previously, the speech and noise model has been proposed. But it displayed incorrect results for some infected voice's noisy conditions. To overcome this problem, the Wiener Filter and Discrete Wavelet Transforms are combined for preprocessing called Hybrid Wiener Filter Discrete Wavelet Transforms (HWFDWT). With the help of Linear Time-Invariant (LTI) filtering and Voice Enhancing Threshold (VETh), the Voice Denoising is carried out by the Wiener filter. The input noisy voice signal is always having a high "Mean Square Error". This occurrence is also called voice noise spectrum. In this research, the infected voice's noisy conditions were removed by thresholding the "Mean Square Error". The Mean Square Error of noise signals is relatively large compared to silence signals. For this, a smaller coefficient is considered as the zero-Mean Square Error. So it is the finest elimination of Silence signal while protecting the Original input signal. From this

consideration, the Wiener filter minimizes the Mean Square Error between the input noisy voice signal and output Denoised Voice.

The main goal of the HWFDWT is to estimate of a silence signal using related noisy voice signal. The Mean Square Error is calculated for each input signal. The Mean Square Error is less neglectable for some input signal called silence filtered signal. This statistical estimation is carried out to all signal filtering. Meanwhile, systematic pitch period and approximation pitch periods were calculated based on the acoustic indices. With respect to the acoustic indices, ElectroGlottoGraph (EGG) is plotted with vocal folds. This EGG is plotted for Constant voice signal glottal period denoised voice signal strength. In this preprocessing analysis of input noisy voice signal, the silence and noisy signals were removed with respect to the ElectroGlottoGraph (EGG) plotted.

3.3 Phase 2: Feature Selection

To select a subset of relevant variables and predictors are used to make the accurate "Voice Pathological Identification System" with the help of voice pathology detection. All relevant variables and prediction are performed on a dataset of voices selected from the Saarbruecken Voice Database. Feature Selection process produced the best accuracy in voice pathology disorder is achieved by selecting the relevant features such as InfoGain, Correlation, and Principal Components Analysis and are evaluated by using feature selection methods. Feature selection methods consist of a selection of Accuracy, Sensitivity, Specificity parameters extracted from an infected input voice signal to evaluate the normal voices and pathological voices, according to the Wavelet Thresholding Algorithm for Dimensionality Reduction. This observation practice analyzes the vocal folds in an efficient manner and improves the overall quality of the selected features in the required time interval. This research discussed the Dimensionality Reduction by Wavelet Thresholding for features selection, and it selects the normal and pathological voices with an accurate manner. From this consideration, this research selects the original voice data collections based on the patient's type, age, gender and treatment for pathology recognition. The pathology recognition constraint is estimated in acoustic analysis, such as the redundant, irrelevant data from the preprocessed data.

3.3.1 Feature Extraction

Extracting the exact pinch of data from the input raw voice signal is called Feature extraction. Feature extraction is performed to extract the Mel Frequency Cepstral Coefficients (MFCC) of the speech. The most dominant features extracted are, Perturbation (pitch, amplitude), Frequency, Cepstrum (Cepstral Energy), Shape of Signal Envelope, Degree of Voicing (DOV), Amplitude Distribution/ Periodic Features, HNR Spectral/ Cepstral, RR (Harmonic Model Signature), Residual of Orig.signal (Mean/STD). MFCC analysis involves that authentic procedure to analyze the unknown input voice signal. It is achieved by comparing extracted features like Signal Energy and Pitch from disorder voice input signal.

The energy level of unvoiced segments is noticeably lower than that of the voiced segments. The higher the energy, the higher the volume of the output speech signals and the higher the amplitude. In addition to this, Pitch is the more prominent feature to find the voice quality, using the pitch range as high and low. To find out whether it is a human voice or not, and if it is a voice of male or female, the pitch is used as an important feature. Pitch is a major auditory attribute of musical tones, along with duration, loudness, and timbre. Probably, Cat Swarm Optimization (CSO) is used to cross-validate the extracted features. The seeking mode and tracing mode are the phases where the cats take rest and seek to the most important features and travel over the path to find the relevant features in an optimized manner and in less time is called Cat Swarm Optimization technique. It has been aimed the effective sensitivity, resourceful specificity and a practical accuracy of disordered voice signal extraction.

The feature extraction technique comprises of two aspects such as Mel Frequency Cepstral Coefficients (MFCC) and Cat Swarm Optimization (CSO). This research combines both aspect namely Cat Swarm Optimization Mel Frequency Cepstrum Coefficients called (CSOMFCC). It is used to extract the best features from the disorder voice signal. Feature extraction has three processes such as making structure of extracted disordered affected voice signal and Application of algorithm trapped into the involvement of disordered affected voice signal. The three stage process takes more time to extract the features and also the output produced will not be upto the mark. To overcome, the issues in the existing Feature extraction process, this research introduces the CSOMFCC technique.

It is replacing the existing feature extraction method by the involvement of the neural network. Although the LPC feature extraction process, the speech signals are divided into frame blocks, windowing is performed by selecting a subset from a large dataset and subsequently autocorrelation is performed to find the correlation between signals and to find the repeating patterns such as periodical signal, occurred by noise then linear predictive analysis is done to store or transmit a series of values representing the voices, LPC analysis involves in the decision making of voiced and unvoiced signals, and finally LPC feature vectors are generated.

The Feature Selected as voice input data is sent to the analysis of Feature Extraction. The voice signal applied to the feature extraction process. Presence of MFCC analysis environments the Melcepts( ) function are collected the Selected Feature as voice input data is converted into the Mel spectrum of the signal. Subsequently, the MFCC analysis output was sent to the Cat Swarm Optimization process. In this situation, the most relevant features of the MFCC analyzed signal was extracted in the feature selection process.

3.4 Phase 3: Classification

The classification is defined as an absolute grouping of mixed input voice data set into normal voices and pathological voices. The proposed methodology of classification processed the extracted features are implied to different classifiers to discriminate between voices. Classification algorithms considered for the research are Support Vector Machine (SVM) Back Propagation Neural Network (BPNN). The BPNN algorithm is optimized to produce the classified output. And proposed a new method called Modified Optimized Back Propagation Network Disorder Voice Classification (MOBPNDVC). The proposed optimization process is a modified back propagation learning algorithm which involves functional constraints. Although Support Vector Machine (SVM) classifier has gained wide acceptance because of the high generalization ability for a wide range of applications in various domains and also uses the quadratic programming for identifying the supports vectors. Whenever the number of samples in the training set is high, identifying potential supports vectors is difficult. Reducing the number of supports vectors used during classification has a direct impact on the speed of Support Vector Machine (SVM). However, the Back Propagation Neural Network (BPNN) always works depending on the

Artificial Neural Network (ANN) because of storage and optimization purpose of classified documents.

The Support Vector Machine (SVM) classifiers were preferred because the quadratic programming and High generalization ability proficiencies are present in the feature extracted input voice signal. The quadratic programming process is supported on the Support Vector Machine classification because of generalization maximization. To get a better result, this research chose the different basis functions like feature extracted male input voice signal and female input voice signal separately. The (.wav) audio formatted input voice signal is extracted and separates the features like gender (male, female) depending upon the Pitch and speed to classify the features as Male or Female.

The proposed method, Modified Optimized Back Propagation Network Disorder Voice Classification (MOBPNDVC) performs efficiently and optimizes the best classification result. The Optimization classification accuracy was achieved in classifying input voice signal as normal or pathological against the disorder types as Laryngitis, Laryngoceles, Dysphonia, Diplophonia, and Chorditis.

3.5 Phase 4: Performance Analysis

The performance quality of the output produced by the proposed system is measured using the various parameters are analyzed. It clearly shows that the system developed is accurately processed and efficiently classified the voice samples into Normal and pathology for both the datasets

3.6 Dataset Specification

In this research, taken two databases, namely, Saarbruecken Voice Database (2000 persons) and Private Real-Time Dataset (80 samples). The voice samples are recorded in (.wav) audio format. Also, the length of the audio clips with sustained vowels is 2 seconds and all recordings are sampled at 50 kHz with 16-bit resolution. Voice samples of people, whose age in the range of 30-35 are considered. These samples are mainly concentrated on the five different diseases like Laryngitis, Laryngoceles, Dysphonia, Diplophonia, and Chorditis. Since these are the major voice disorder types which affect the voice parameters and the voice organs like vocal cord, vocal fold, vocal box, bulge appearance in the neck.

In the Saarbruecken dataset This sample contains the recordings of vowels /a/, /i/, /u/ and phrases like “good morning”, “how are you”, “hello”, “where are you” and “welcome” are produced at different pitch levels like low, normal, and high. The Real-time Dataset was collected from the Department of Pathology, Karpagam University, Faculty of Medical Sciences and Research, Coimbatore. This sample contains the recordings of vowels /a/,/e/, /i/ /o/, /u/ and phrases.

3.7 Performance Metrics

The performance evaluation was conducted using various quality metrics that estimate the algorithms in terms of its efficiency with respect to accuracy and speed. All the proposed methods were compared with the existing and/or conventional counterparts.

- **Parametric Representation by Preprocessing**

The performance metrics used to extract the speech signals separately after silence and noise removal is proposed in Phase I of the research work, are described below

i. Signal to Noise Ratio Calculation

A speech distortion index is defined to measure the degree to which the speech signal is deformed. The noise reduction to quantify the amount of noise being attenuated, analytically examined the performance behavior of the proposed Hybrid Multichannel Wiener Filter Discrete Wavelet Transform (HWFDWT) by SNR value using the equation (3.1)

$$SNR = 20 \square log10\left(\frac{S}{N}\right) \tag{3.1}$$

where the Signal to Noise ratio is calculated for the sampled data of both normal and pathological persons. SNR graph is plotted for different filters.

- **Feature Selection and Extraction**

The performance metrics used for feature selection and extraction of the speech signals is execution time.

i. Execution Time

The Execution time is defined as the total time taken to extract the redundant features from the input speech signals.

- **Classification of Voiced Data**

The performance quality of the output produced is measured using the following parameters, Accuracy, Specificity, Sensitivity and Execution Time. Based on these parameters the efficiency of the developed system is measured.

i. Accuracy

The Classification Accuracy is calculated using the equation (3.2). Most classification algorithms seek models that attain the highest accuracy when applied to the test dataset

$$Accuracy=(TP+TN) / (TP+TN+FP+FN) \quad (3.2)$$

ii. Specificity

The Specificity calculation is done using equation (3.3). The ratio between classified normal and total normal samples is obtained.

$$Specificity = TN/TN+FP \quad (3.3)$$

iii. Sensitivity

The Specificity calculation is done using equation (3.4). The ratio between identified pathological and total pathological samples is obtained.

$$Sensitivity = TP/TP+FN \quad (3.4)$$

- TP - positive and predicted by the model as positive.
- FN - positive but predicted by the model as negative.
- TN - negative and predicted by the model as negative.
- FP - negative but predicted by the model as positive.

iv. Execution Time

The Total time is taken to perform classification and is calculated using the equation (3.5)

$$T = 1/f \text{ or as: } f = 1/T \tag{3.5}$$

3.8 Chapter Summary

The Proposed Automatic Voice Identification System consists of three main steps, namely, preprocessing, feature selection and extraction and classification. This chapter presented the research design and explained, in brief, the various contributions made to enhance the algorithms proposed in each step. A detailed description of the working of the proposed preprocessing methods is presented in the next Chapter, Chapter 4 (**Parametric Representation by Preprocessing**).