CHAPTER 5

FEATURE SELECTION AND EXTRACTION

The Feature Selection and Extraction process is performed with the preprocessed voice signal as a primary step to discriminate between the normal and pathological voices. The feature subsets of relevant variables and predictors are selected to develop "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 and Real-Time data set. The results obtained are evaluated in terms of Accuracy, Sensitivity, and Specificity. Feature Selection process produces the best accuracy in extracting the dominant features such as InfoGain, Correlation and Principal Components Analysis evaluated by using feature selection methods. Feature selection methods consist of a selection of accuracy, sensitivity, specificity parameters extracted from the 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 features selected in the required time interval.

This research analyzes the Dimensionality Reduction by Wavelet Thresholding for features selection. It selects the normal and pathological voices in an accurate manner. In order to discriminate the voices, this research selects the original voice data collections such as patient's type, age, gender and treatment for pathology recognition. These pathology recognition constraints are estimated in acoustic analysis, such as the redundant, irrelevant data from the preprocessed input. The analysis helps to improve the working time, system memory consumption, speed at a processing time interval. The analysis works effectively against all private dataset. Especially the Info_Gain_Attribute is the main feature for calculating the information gain for each input voice signal. In a similar way, the Correlation for predictive analysis in data collection is used. Moreover, the most significant parameters were selected from the several features called Principal Components Analysis.

The Selected Feature voice input data is sent to the analysis of Feature Extraction. It helps to improve the dataset and the features extracted from the voice signal, and also, to these extracted feature signals are then applied to the classification phase. Extracting the exact pinch of data from the input raw voice signal is called Feature extraction. Feature extraction comprises getting Mel Frequency Cepstral Coefficients (MFCC) of the speech. This may help to extract the features like 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), and Residual of Original signal (Mean/STD). MFCC analysis involves the authentic procedure to analyze the unknown input voice signal (Hansen, J. H., et al 1998). It will happen by comparing extracted features like Signal Energy and Pitch from voice input signal.

5.1 Signal Energy and Pitch

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. The short-time energy of speech signals reflects the amplitude variation and is defined using equation 5.1. 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 whether 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.

The amplitude variation is defined as the short-time energy of speech signals given in equation 5.1.

$$\mathrm{E}(i) = \frac{1}{N} \sum_{n=1}^{N} |x_i(n)|^2 \tag{5.1}$$

The Feature Selection and Extraction Process of the preprocessed voice signal are carried out and it is presented in Figure 5.1.

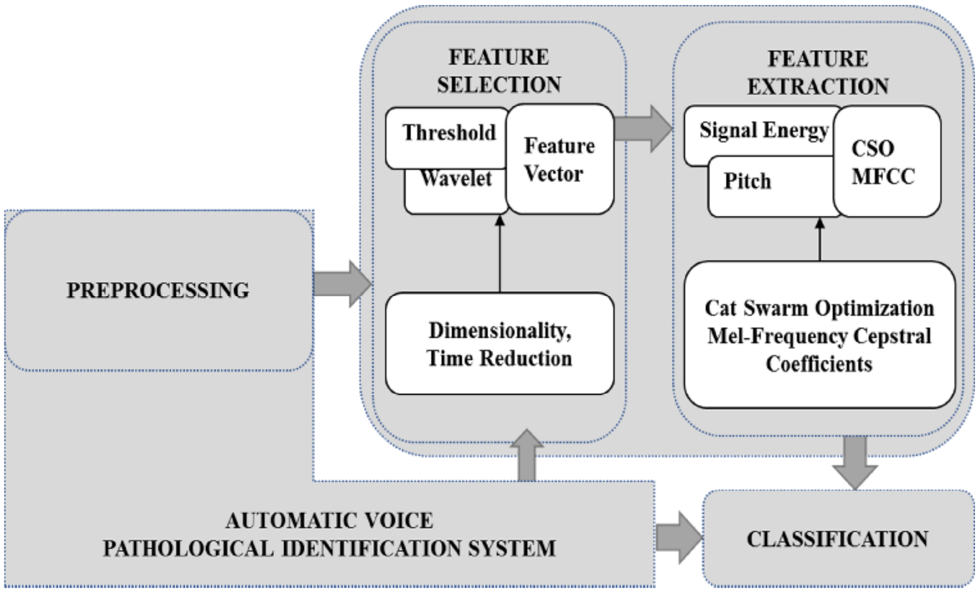

Figure 5.1.Feature Selection and Extraction Process

The effective Sensitivity, resourceful Specificity and a practical Accuracy of voice signal extraction have been aimed. This feature extraction process comprises of two aspects such as Mel Frequency Cepstral Coefficients (MFCC) and Cat Swarm Optimization (CSO). This research combines both aspects, namely Cat Swarm Optimization and Mel Frequency Cepstrum Coefficients called (CSOMFCC). It is used to extract the best features from the disorder voice signal. Feature extraction two processes such as

- Making structure of extracted disordered affected voice signal, and
- Application of algorithm trapped into the involvement of disordered affected voice signal.

To overcome the two Feature extraction process, this research introduces the CSOMFCC technique. It is replacing the existing feature selection method by the involvement of the neural network. Although the LPC feature extraction process performs fairly well, the speech signals are divided into frame blocks, windowing is performed by selecting a subset from a large dataset. Subsequently, autocorrelation is performed to find the correlation between signals and to find the repeating patterns such as periodical signal, occurred by noise than linear predictive analysis is done to store or transmit a series of values representing the voices, LPC analysis involves the decision making of voiced and unvoiced signals, and finally LPC feature vectors are generated.

5.1.1. Structure of Voice Signal

The Feature Selected voice input data is sent to the analysis of Feature Extraction. The selected voice signal is applied to the feature extraction process. Presence of MFCC analysis environments the Melcepts ( ) function which collects the feature selected voice input data into the Mel spectrum of the signal. Subsequently, the MFCC analysis outputs were 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.

5.1.2. Application of Algorithm to Voice Signal

The process of feature extraction of the speech signal using Mel Frequency Cepstral Coefficients (MFCC) vectors will produce an acoustic speech signal. Cat Swarm Optimization process is used to form the specific acoustic vector for each speaker. It helps to improve the dataset and features extracted from the voice signal and these features extracted signals are applied to classification. The procedure of feature extraction of input voice signal utilizing Mel Frequency Cepstral Coefficients (MFCC) vectors will deliver acoustic output voice signal. Cat Swarm Optimization process is utilized to shape the particular acoustic vector for every input signal. It is enhancing the input of private and Saarbruecken voice database and features extracted from the input.

5.2. Feature Selection Process

The Feature Selection is the process in Automatic Voice Pathological Identification System (AVPIS) to progress the amount of feature selected from the voice database. The progress is used to concentrate on the Accuracy, Sensitivity, and Specificity parameters for feature selection. The parameter analyses the signal energy by extremely discriminative features such as pitch, jitter, shimmer, and formants acoustic features (Alarifi, A., et al 2018). This acoustic feature analysis reflects the error rate minimization as well as independent Accuracy, Sensitivity, and Specificity diagnosis process. This will produce the automatic voice disorder examination process of voice quality.

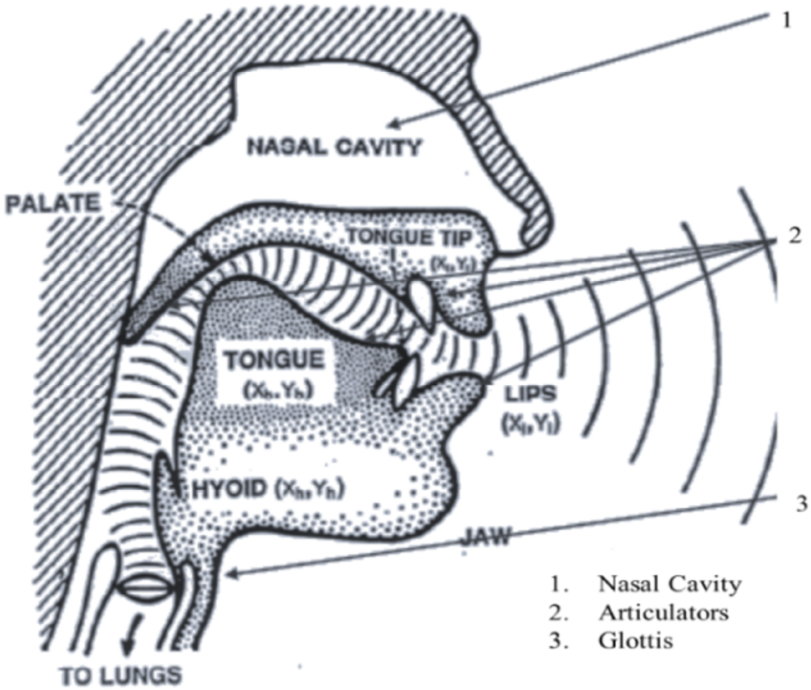

Figure 5.2. Voice Production System

Figure 5.2 shows the parts of the voice production system that can be affected by pathology on the nasal cavity, articulators, and glottis.

5.2.1. Voice Production System

The voice production system is a control for nasal cavity, articulators, and glottis spectral balance of the voice production system shown in Figure 5.2. The Nasal Cavity, Articulators and Glottis signal produce the vocal tract shape of the speaker. This can produce the result of resonance frequencies modification with respect to the voice signal. These resonance frequencies were very much useful in voice production with respect to the time duration. In the voice production system, the vocal tract speech analysis is evaluated by using the Linear Prediction Analysis (LPC). After filtering the vocal tract, the human signals are pointed for speech analysis. In this time, the resonance frequencies are evaluated by using the Linear Prediction Analysis algorithm. In filtering, the speech signal is separated from the infected signal. This process is called Inverse Filtering. The Inverse Filtering (residue) is used to finalize the glottal speech signal. This process may extend to many voice input signals with the fundamental frequencies/amplitudes the original vocal tract analyzed on a voice production system.

In pathological feature selection analysis, the human voice disorder selection procedure is useful for the feature selection process of vocal tract pathologies. In this case, the dimensionality, time reduction is analyzed by using the feature vector. Nevertheless, the research has been concentrating in dimensionality reduction and time reduction with respect to the feature vector. The pronouncement between two different voices such as healthy voice and pathological voice signals are analyzed in chapter 4. These input values are obtained from the preprocessed two-dimensional glottal wavelet of the voice reappearance input. The threshold value adopted is based on the maxima accuracy rate. The threshold value is adopted with respect to the acoustic features of amplitude perturbation, LPC, MFCC, pitch perturbation and so on.

In such a case, the voice pathology feature selection may not be analyzed by using highly discriminative acoustic features. In this case, the wavelet thresholding algorithm and the information gain were analyzed for dimensionality reduction. In this chapter Feature Selection Process has two estimation aspects such as dimensionality reduction and principal components analysis. The dimensionality reduction of pathological voices using wavelet thresholding algorithm features is analyzed with Information Gain. These features are extracted from a subset of the Saarbruecken Voice Database with sustained vowel /a/, /e/,/i/,/o/,/u/ samples. Form this Saarbruecken Voice Database and Private Database contribution, the selected features are evaluated with respect to the thresholding assignment on voice pathology detection. This system correspondingly identifies the least features that can significantly improve detection performance with respect to the Automated Voicing Analysis.

5.3. Automated Voicing Analysis (AVA)

At the time of phonation, the Automated Voicing Analysis (AVA) was achieved for each input voice searching. In this case, the Automated Voicing Analysis extracts and selects features such as long-term features and short-term features. This feature selection is used to recombine the voice quality for each signal.

Input: Input Voice
Read: Signal duration (SiD), fundamental frequency (f_o), standard deviation (SD)
Process: if SiD > 3ms
Select features: voice quality
Subprocess: long-term features
Calculate fundamental frequency (f_o) for input voice,
Calculate standard deviation (SD) for fundamental frequency (f_o)
Select features: voice quality ($f_{o,}$ SD)
voice quality ($f_{o,}$ SD) > threshold (SiD)
return result of Subprocess: long-term features
else
if SiD <3ms
Subprocess: Short-term features
Calculate fundamental frequency (f_o) for input voice,
Calculate standard deviation (SD) for fundamental frequency (f_o)
Select features: voice quality ($f_{o,}$ SD)
voice quality ($f_{o,}$ SD) < threshold (SiD)
return result of Subprocess: short -term features
Result: voicing occurred
Return: voiced signal, glottal harmonics, and the glottal noise

Figure 5.3 Algorithm for Automated Voicing Analysis (AVA)

From Figure 5.3, the algorithm for automated voicing analysis is worked based on the long-term features and short -term features selection with respect to the result of occurred voice/non-occurred voice. The input voice signal was processed to select the long-term features and short-term features for selection depend on voice quality. From this analysis, the fundamental frequency (f_0) for input voice is calculated with respect to the Standard Deviation (SD). In that context, the fundamental frequency (f_0) were compared with the voice quality ($f_{0,}$ SD) and threshold (SiD). If the voice quality ($f_{0,}$ SD) is a greater value of fixed threshold (SiD) the features were selected on the origin of voice quality ($f_{0,}$ SD). This result is called the long-term features selection.

Meanwhile, the fundamental frequency (f_0) context is compared with the voice quality (f_0, SD) and threshold (SiD). If the voice quality ($f_{0,}$ SD) is a lesser value of fixed threshold (SiD), the features were not selected on the origin of voice quality (f_0, SD). This result is called the short-term features selection. This spectral envelope of the first few glottal harmonics and the glottal noise were detected with respect to the Automated

Voicing Analysis (AVA) to identify the voiced frames. The Automated Voicing Analysis (AVA) is used to remove the drifting signal amplitude and autocovariance difference. This is also used to increase the normalized spectra of the voice signal.

Automated Voicing Analysis (AVA) deselect the higher inter-patient variability based fundamental frequency (f_0) and its harmonics. So, the glottal noise is considered to be a voice quality section, a parameter based on Voice Quality Energy (VQE), but derived from the AVA, VQE, the energy was calculated for input voice.

The feature selection process provides information on the examination of removed features of the input voice for selection of pathology (Castellanos, G., et al 2006). The selection of pathology conversation includes the following process steps:

Step 1: To check selection features when performing Automated Voicing Analysis manually versus automatically using measured pre-defined input voice for selection of pathology

Step 2: To check recorded replication voice modulation, whereby foundations of input dataset reduced data were traced from the pre-defined dataset to the transmitted set using simulation

Step 3: The comparison of voice quality (f_0, SD) and threshold (SiD) methodology with the previously read voice signal

Step 4: The improved system performance for pathologic categorization with respect to the homogenous aspects.

Step 5: Practical performance for pathology assessment voices and the input signal.

5.4. Dimensionality Reduction

The dimensionality reduction or dimension reduction is the process of reducing the number of extracted features by obtaining a set of principal features (Godino-Llorente, J. I., et al 2006). It can be divided into feature selection and extraction.

5.4.1. Acoustical Analysis

An acoustical analysis is taken manually for dimensionality reduction. There are many processing levels available in the Acoustical analysis are discussed below, Level 1: Manually produce the standard loudness level to produce the vowel |a| with a constant pitch of patient.

Level 2: Vowel |a| was recorded with a standard mouth-to-microphone distance of 5 cm and in quiet room recording were used.

Level 3: Instruct to the patient to maintain the pitch level at least 3 seconds

Level 4: Repeat the process for at least 4 times and finalize the test sample

Level 5: Also, consider the following 19 parameters for acoustical analysis such as absolute jitter, amplitude perturbation quotient, amplitude tremor intensity index, degree of sub-harmonics, degree of voiceless, degree of voice breaks, fo-tremor intensity index, fundamental frequency variation, jitter percent, noise harmonic ratio, peak-to-peak, amplitude variation, pitch perturbation quotient, relative average perturbation, shimmer in decibels, shimmer percent, smoothed amplitude, perturbation quotient, smoothed pitch perturbation quotient, soft phonation index and voice turbulence index. The above parameters might be extracted from the acoustical analysis. All those parameters were constantly suitable for all samples.

5.5. Pre-Feature Extraction Stage

Additionally, the benefit of an acoustical analysis system is used to find voice disorders and evaluate the treatment process with respect to the Level 5 acoustical analysis parameters. In this stage, the wavelets are used with wavelet packets such as continuous and discrete forms of the voice signal. The voice signal is named the Discrete Wavelet Transform (DWT) and Continuous Wavelet Transform (CWT).

In feature-extraction perspective, the extraction efficiency is varied with respect to the suitability extracted features and the Feature selection availability. The selection of the Discrete Wavelet Transform (DWT) and Continuous Wavelet Transform (CWT) significantly affects the output features selection and feature-extraction.

5.5.1. Motivation of Feature Selection Process

In feature selection and feature-extraction cases, the Discrete Wavelet Transform (DWT) is used in the acoustical analysis of self-determining of the voice signal itself. The similarly in human voice is obvious so, the Discrete Wavelet Transform (DWT) is used in processing the voices. The proposed Feature Selection and Feature Extraction process is shown in Figure 5.4

Feature-Extraction Stage and Acoustical analysis are combined each other and forming the comprehensive approach called the Discrete Wavelet Transform (DWT) specification. This specification was used to find wavelet that is designed to be optimally used to distinguish pathological voices from normal voices.

To manage the acoustical analysis and Feature-Extraction Stage of this phase, in the proposed system feature selection and Extraction process is performed in order to collect only the relevant features. This helps us in dimensionality and time reduction. In the Feature Selection phase, from the denoised signals, the number of specific wavelets and threshold type is applied in order to select the most dominant features. Several families of wavelets have proven to be useful in MATLAB toolbox; Haar family is the first and simplest wavelet type, Haar wavelet is continuous. So, it is used in the research. Thresholding can be applied as a hard threshold or soft threshold. In a hard threshold, all coefficients greater than a given threshold value are retained, and the remaining coefficients are made zero. In soft thresholding, if a coefficient is greater than the given threshold, then the coefficient is subtracted from the threshold, which was discussed in the automated voicing analysis (AVA) algorithm. This operation shrinks coefficients to an absolute value. The most important thing is the research should focus on is selecting the right threshold method and calculation. This research has taken soft thresholding, because, the signal and noise exist in similar frequency levels and similar amplitudes. For not losing much of the information, soft thresholding is applied. Out of the available selection rules, the threshold calculation method is picked as heursure for the same reasoning with respect to the Rigrsure (adaptive threshold selection using the principles of steins unbiased risk estimate (sure), Sqtwolog – (fixed form threshold sqrt(2*log(length(x)), Heursure (heuristic variant of riqrsure and sqtwolog), and Minimaxi (minimax thresholding).

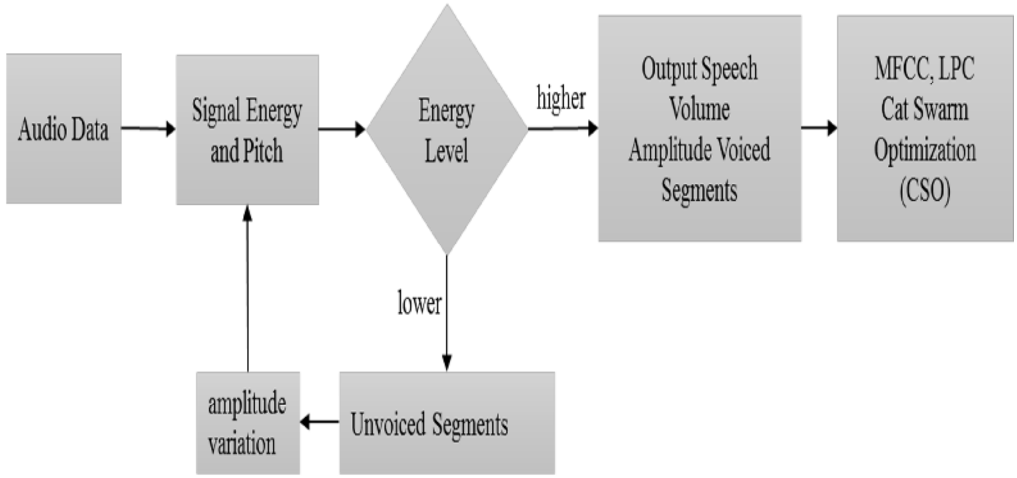

Figure 5.4. Proposed Feature Selection and Extraction

The process of selecting a subset of relevant features for use in model construction is called feature selection. The process of retaining useful information from the signal and discarding the unwanted information is feature extraction. finally, the Proposed Feature Selection and Feature Extraction framework selects and extracts the most relevant features and are given as input to the classification phase.

5.5.2. Info Gain Attribute Evolution

Information Gain attribute computes the data gain for each element. The outcomes can fluctuate from 0 (no data) to 1 (most relevant data). The information gain attribute evolution got for each considered component. The best esteem is accomplished by the age, trailed by two MFCC coefficients, the jitter, the second subordinate, and others. For the exploratory tests, the diminished features in data gain are avoided, this is equivalent to 0, while those highlights with a data increase more noteworthy than 0 were considered. The info gain attribute evolution the prescient capacity of each characteristic, giving us the possibility of inclining toward sets of properties that are very corresponded with the feature. Also, utilized the feature selection and feature extraction strategy are utilized to choose the most noteworthy parameters. The foremost segments which have acquired in any event half of the positioning is chosen (Teixeira, J. P., et al 2013).

5.5.3. Pathology Threshold for Proposed Feature Selection and Extraction

In this section the feature contracts with the pathology threshold for proposed feature selection and extraction determination with respect to Automated Voicing Analysis (AVA). In the pathology threshold for proposed feature selection and extraction, the two databases will be critical to find threshold values using each database to determine the threshold separately and then applying the result to each datasets.

For taking the two different databases it will allow the efficient voice pathology detection for taking into account the feature selection and feature extraction result for the AVA, distinctly for each dataset. Therefore, by providing separate thresholds, one for Saarbruecken Dataset and another for Real-time Dataset the efficiency will be improved.

Given the feature selection and feature extraction result, the discrimination instance provides the best selection. Meanwhile, the feature selection and feature extraction result with regard to AVA for the Saarbruecken dataset suggests a higher threshold. At the same time, the feature selection and feature extraction result, the discrimination instance provides the best selection. And, the feature selection and feature extraction result regarding AVA for the Real-time dataset suggests a relatively lower threshold (Vasilakis, M., et al 2009). To associate the thresholds values of Real-time dataset and Saarbruecken dataset, this research achieved continuous experiments on these databases.

5.5.4. Pseudocode for Pathology Thresholding

```
Initialize: selection rate extraction rate
Call: Extraction Rate (ER)
ER_threshold = Count (correct detections)_threshold
               Total (detections)_threshold
Collect: Automated Voicing Analysis (AVA)
Repeat: four features
End: with Extraction Rate (ER)
```

Since Automated Voicing Analysis (AVA) provides a short-term sequence of jitter values for each input voice signal, this calculation will repeat the features that make use of the Extraction Rate (ER) is presented in pathology thresholding using Equation 5.2.

$$ER_{threshold} = \frac{Count\ (correct\ detections)_{threshold}}{Total\ (detections)_{threshold}} \tag{5.2}$$

Considering, that each short-term value corresponds to an analysis frame, the five features are defined as

(1) Calculate the frames percentage over the threshold (Over)

(2) Find the consecutive frames that are over the threshold (Max Over)

(3) Find the consecutive frames that are under the threshold (Max Under)

(4) Analysis window per signal was determined by the size of (Max Under) Time | size of (Max Over) Time

(5) signal (hop size) = signal (average pitch period)

The five features are based on frames rather than time since for each signal all frames were equal in size because the analysis window per signal was determined by the average pitch period of the signal, and a fixed hop size (Tavares, R., et al 2010). As expected, the threshold which was defined in a specific database provides the best selection for each taken dataset. Finally, the Pathology Thresholding establishes the efficient selection and extraction for the Real-time dataset and Saarbruecken dataset.

5.6. Feature Extraction Process

The feature extraction deals with the task of extracting the best feature and redundant set which reduces the extraction time and as well as increases the feature selection accuracy. From a given input voice signal, the feature extraction algorithm selects a subset of size *m* which increases the extraction accuracy. If the size of the feature set is *m*, then there will be *2m* possible feature subsets. The extraction of the best feature subset can be viewed as a combinatorial optimization problem and is solved using Mel Frequency Cepstral Coefficients (MFCC) with respect to the Signal Energy and Pitch. Each voice signal is processed by the Mel Frequency Cepstral Coefficients (MFCC) with respect to the

Signal Energy and Pitch techniques. The acoustic features are extracted with respect to the feature subset.

The aim of using the feature extraction algorithm is to use a smaller amount of features to achieve more extraction accuracy. Therefore, the thresholding pathology contains

(i) Accuracy, and

(ii) Features used.

Feature extraction is a voice recognition step which reduces the number of features, removes irrelevant, noisy and redundant data, and results in acceptable recognition accuracy. In this work, a Mel Frequency Cepstral Coefficients (MFCC) with respect to signal energy and pitch is used to extract the most significant features for the discrimination of healthy and pathological voices.

5.6.1. Cat Swarm Optimization Mel Frequency Cepstrum Coefficients (CSOMFCC)

Mel Frequency Cepstral Coefficients (MFCC) is used to extract features in the vocal tract and the waveform. This acoustic signal identifies (12 extracted features) which are prominent and having encountered them which contains data about the vocal tract and the excitation waveform.

The research identifies the pathologies on vocal folds and vocal tract. From this extraction, the vocal folds are the main source of voice disorder because of its emphasis on pathologies voice disorder. Mel Frequency Cepstral Coefficients (MFCC) can be predicted by using a Feature Extraction Process with respect to the respective parameters such as Jitter, Shimmer. This is derived from Linear Prediction Coefficients (LPC).

5.6.2. Linear Predictive Coding (LPC)

The basic idea behind the Linear Predictive Coding (LPC) analysis is that a speech sample can be approximated as a linear combination of past speech samples. LPC is a frame-based analysis of the speech signal which is performed to provide observation vectors of speech. LPC feature extraction process is explained in Figure 5.5.

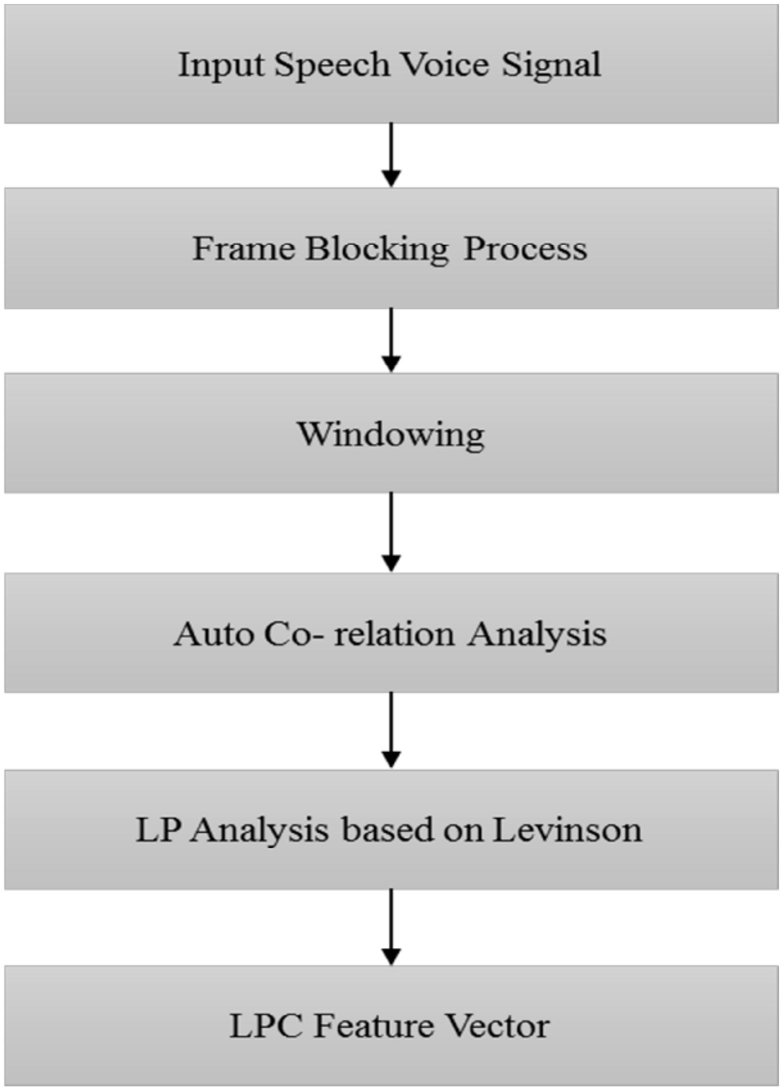

Figure 5.5. Linear Predictive Coding (LPC) Feature Extraction Process

Though, Frame Blocking-based LP Analysis on Levinson MFCCs typically translate more information on feature extraction process, whereas, LPC-based MFCCs remove Jitter, Shimmer for getting the LPC Feature Vector.

The Linear Predictive Coding (LPC) based feature extraction processes are established in Figure 5.5, where Frame Blocking-based LP Analysis on Levinson MFCCs are found to be more dependent on Auto Co-relation Analysis. It is producing the high-pitched speech outcome for all disordered voices except normal voices. Frame Blocking - based LP Analysis on Levinson MFCCs, were found to be more penetrating to protective noise in voice signal recognition. This is the situation since Frame Blocking -based LP Analysis on Levinson MFCCs ignore the pitch-based vocal tract seen in Frame Blocking – based LP Analysis on Levinson MFCCs. In terms of vocal folds, this produced pathological voice might be having a higher mass, lower closure and elasticity variation on the pathological voice. This procedure is not accomplishing the glottal cycles and, so, it is rectified based on the vocal folds closure and complete closure on glottal cycles. For this reason, voice harmonic structure changed on the production of disordered voices. Due to this reason the voices are identified as the disordered voices. The voice has change in harmonic energy and fundamental frequency perturbation.

Frame Blocking -based LP Analysis on Levinson MFCCs were measured to be suitable for the persistence, because, in the presence of voice disorders with respect to the irregular movement of the closure. Feature selection plays a major role in selecting the features, which is appropriate for Feature Extraction. Feature selection can be an essential one for automated voicing analysis System. MFCC, i.e, Mel Frequency Cepstral Coefficients, is one of the most favored component extraction strategies. In this investigation, Mel Rate of Occurrence Cepstrum is representations of a brief - term control otherworldly scope of a voice transmission, which delivered and predicated on the direct cosine change of the log control go over a nonlinear Mel-scale rate of the event. Mel Rate of repeat Cepstral coefficients turned into an individual from joined with as evacuate a consolidated gathering to comprise an MFCC. The Mel Frequency Cepstrum utilizes divided recurrence groups on the Mel Scale similarly, which appraises the human sound-related framework's reaction more intently than the standard Cepstrum which utilizes straightly dispersed recurrence groups.

The Mel-frequency Cepstral Coefficient (MFCC) technique is often used to create the pattern of the sound files. The MFCC is based on the known variation of the human ear’s critical bandwidth frequencies with filters spaced linearly at low frequencies and logarithmically at high frequencies used to capture the important characteristics of speech. The signal is divided into overlapping frames to compute MFCC coefficients. Let each frame consist of N samples and let adjacent frames being separated by M samples where M<N. Each frame is multiplied by a Hamming window. The MFCC feature extraction process is shown in Figure 5.6.

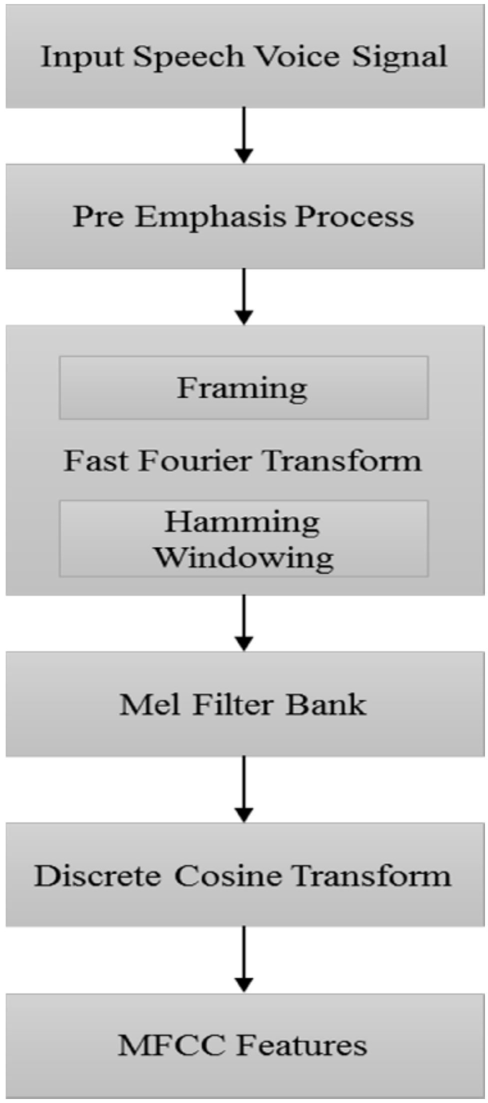

Figure 5.6. Formation of MFCC

The speech waveform is cropped to remove silence or acoustical interference that may be present in the beginning or end of the sound file. The windowing block minimizes the discontinuities of the signal by tapering the beginning and end of each frame to zero. The FFT block converts each frame from the time domain to the frequency domain. In the Mel-frequency wrapping block, the signal is plotted against the Mel spectrum to mimic human hearing. In this section, the voice pathology identification system was designed and enforced in Matlab. For the experimental study, ten samples (6 Pathological and 4 Normal) voice files in the .wav format are givens as input to the system, that was recorded in an exceedingly noise free atmosphere employing a mike array. The reading of the designed Voice Pathology Identification System is delineated in Figure 5.7.

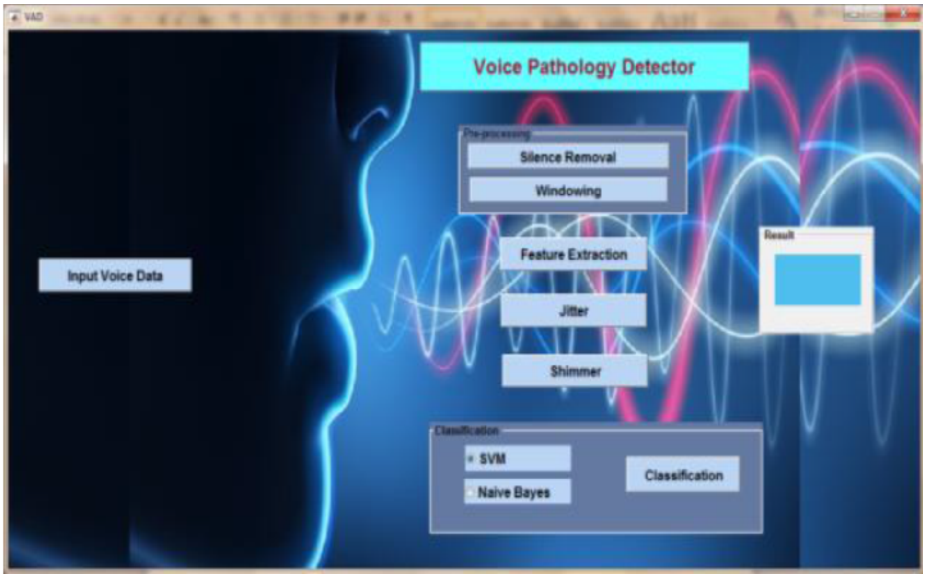

Figure 5.7. Voice Pathology Identification System

The first section shows the estimation of short time energy present in the voice signal. The second section shows the threshold estimation based on Spectral Centroid. The third section detects the silence present in the voice signal and the fourth section removes the silence detected in the third section. The third section includes Mel Frequency Cepstral Coefficients and the Mel Frequency Cepstral, which equalize the input voice frequency information measure. It describes the ability spectral envelope of one frame. Here, Twelve Mel Frequency Cepstral Coefficients were generated from the Log filter bank. Figure 5.8 and Figure 5.9 represents the Mel – Frequency Cepstral coefficients for the input.

First and foremost, the signal is framed rapid into short frames. Usually, the signals are framed into 20-40 ms frames. 25ms is standard (If the frame is much shorter it does not have enough samples to get a reliable spectral estimate, and if it is longer the signal changes too much throughout the frame). So, 25 ms frames are taken. Each frame periodogram estimate of the power spectrum is calculated (to identify which frequencies are present in the frame). Mel-filterbank is applied to the power spectra, and energy in each filter is added. The log value of all filter bank energies is taken. It allows us to use of cepstral mean subtraction, which is a channel normalization technique. Then the DCT of the log filter bank energies is considered. Because the filter banks are all overlapping, the filter bank energies are quite correlated with each other. The DCT decorrelates the energies. Finally, 2-13 DCT coefficients are retained, and the rest is discarded. This is because the higher DCT coefficients represent fast changes in the filter bank energies that actually degrade the performance, and a small improvement is achieved by dropping them. In this research, the 12 Mel Frequency Cepstral Coefficients are taken in order to extract the dominant features from the speech signals are displayed in Figure 5.10.

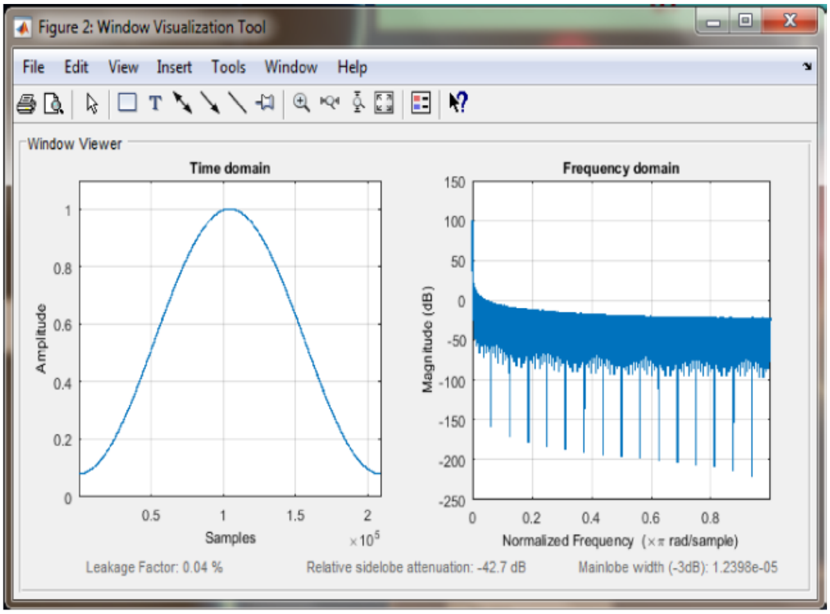

Figure 5.8. Windowing

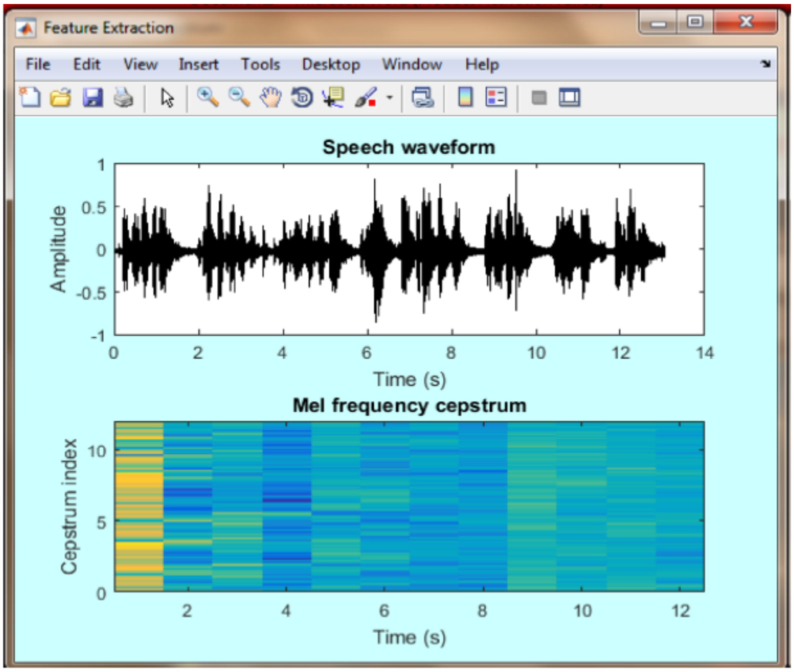

Figure 5.9. Traditional Speech Signal Vs. Mel Frequency Cepstrum

The normal input speech waveform after removing the silence is displayed in the first section and The Mel Frequency Cepstrum is displayed in the second section.

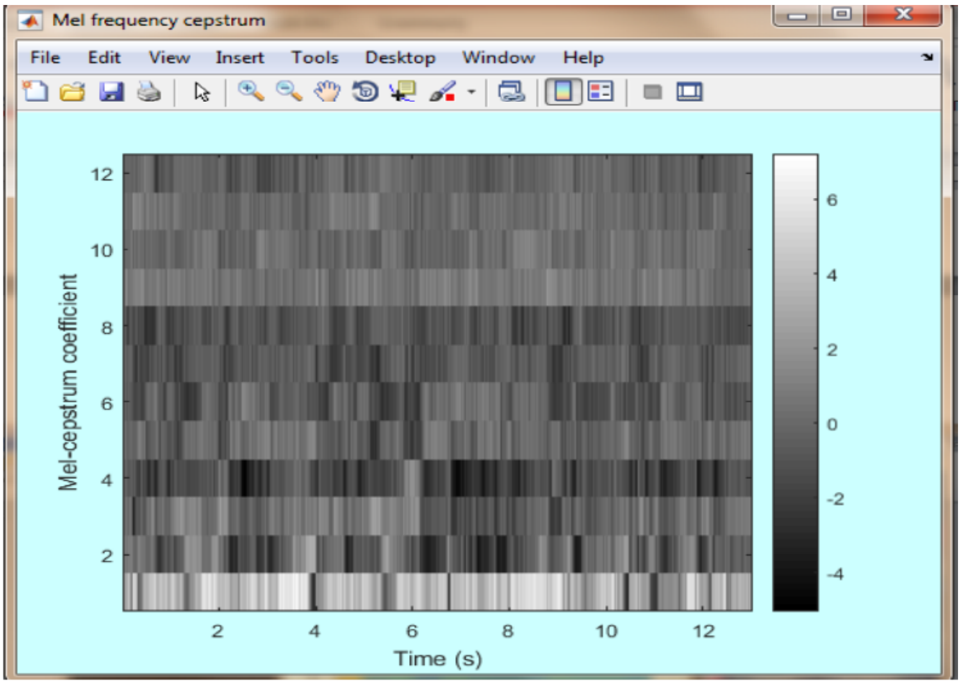

Figure 5.10. Twelve MFC Coefficients

The proposed Mel Frequency Cepstral technique is dependent on parametric cepstral coefficients in Mel and Bark scales. The most applicable features are consequently chosen to utilize two calculations. One depends on Principal Components Analysis (PCA) and other depends on Features Selection. So as to choose whether a voice recording is sound or obsessive, four distinct classifiers are executed. All investigations of voices are chosen from the Saarbruecken Voice Database. The outcomes are assessed as far as accuracy, sensitivity, specificity, and ROC area are concerned. The MFCC coefficients endeavor to examine the vocal tract freely of the vocal folds that can be harmed because of voice pathologies.

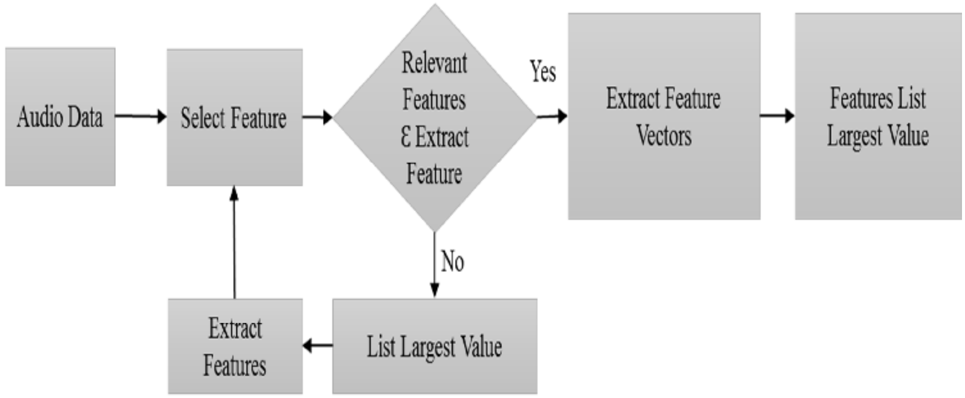

Figure 5.11.Feature Selection System

The feature selection system is shown in the above Figure 5.11. For a given input Audio data, compute the effectiveness of Relevant Features measure and Extract Feature measure. And this was compared with each Features measure. If the measures are giving the largest list value the Features are extracted for each term of the vocabulary. Else, this is applied to the feature vectors calculation by the feature vectors and selects the relevant features that have the highest values of feature vectors. All other terms are discarded and not used in feature selection and extraction. The utility measures such as mutual information and frequency are taken for feature selection.

5.7. Features Validation - Cat Swarm Optimization (CSO)

Cat Swarm Optimization (CSO) is an advancement subset of feature calculation. This calculation reflects the creation of cats for the conduct of feature selection. This calculation will be isolated into two modes, Seeking Mode and Tracing Mode. Cats while looking for mode will alarm to its environment, cats in the following mode will discover and follow the target. CSO utilize two modes to locate the best features (Muthusamy, H., et al 2015).

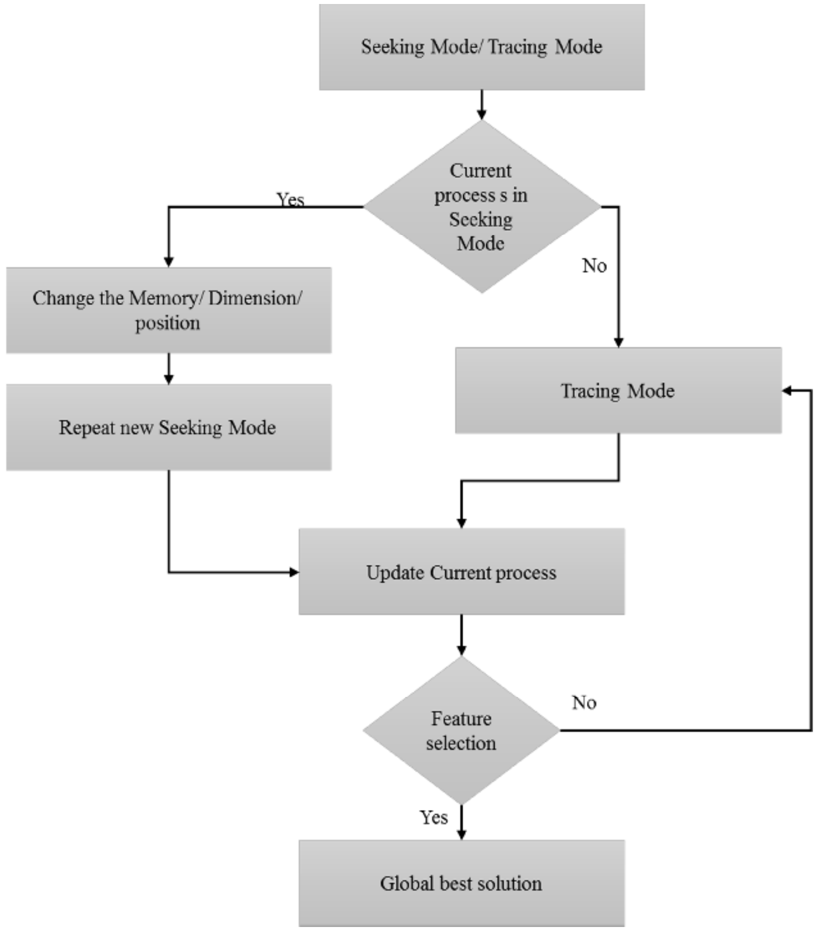

Figure 5.12. Features Validation - Cat Swarm Optimization (CSO)

The main goal of the Features Validation - Cat Swarm Optimization (CSO) study is to afford a Cat Swarm Optimization based system for select the Global best solution Feature selection with an optimized feature on input Audio Data. Compared to other Optimization techniques, Features Validation - Cat Swarm Optimization (CSO) offers Global best solution over new Seeking Mode and Tracing Mode.

This feature selection and extraction have been combined with the Mel Frequency Cepstral Coefficients (MFCC) and Cat Swarm Optimization (CSO) for best features extraction because it uses the wavelet-energy and significant Cat Swarm Optimization (CSO) approach. This approach introduced a new feature descriptor of local feature vector Mel Frequency Cepstral Coefficients. However, Cat Swarm Optimization (CSO), Mel Frequency Cepstral Coefficients (MFCC) and Automated Voicing Analysis (AVA) still may be unreliable during feature extraction. To improve the best features extraction, the combined Cat Swarm Optimization Mel Frequency Cepstral Coefficients called CSOMFCC method is introduced in this research. This represents the behaviors of the cat as well as Mel Frequency Cepstral Coefficients (MFCC) signal energy and pitch.

The Cat Swarm Optimization Mel Frequency Cepstral Coefficients called (CSOMFCC) function is defined as

$$\max_{f(x)} \stackrel{\text{def}}{=} (x_2 - x_1^2)^2 + (1 - x_1)^2$$

(5.3)

Features Validation function due to Automated Voicing Analysis (AVA) is calculated using Equation 5.3. Many optimization procedures are displaying to deliberate the separation of normal voices with the pathological voices for solving Features Validation function. Nevertheless, in the case, the CSO has been successfully applied to the best feature's extraction.

As per Figure 5.12 Cat Swarm Optimization (CSO) has two modes of operation such as seeking mode and tracing mode. In seeking mode, the cat is always in the rest mode. Each time the cats are awake from the rest mode for seeking the destination. Meanwhile, in the tracing mode, the cat takes in the iteration modes running after the results got from the selection. This helps to improve the optimization in terms of sensitivity, specificity, and accuracy.

5.7.1 Seeking Mode

In the seeking mode, the cat takes rest, looks around and finds the next position to move on. the essential factors considered are namely, Seeking Memory Pool (SMP). It is the size of the seeking memory for each cat, Seeking Range of Selected Dimension (SRD). It gives the mutative ratio, the difference between the new and the old position of the cats' that will be within the range. Counts of Dimension to Change (CDC). It deals with the number of dimensions that will be varied. Self-Position Considering (SPC), SPC is a Boolean variable that decides whether the cat is in seeking or tracing mode, be it TRUE or FALSE.

This Seeking Mode show is utilized to display the cat among time of resting yet, but being alarm checking out its condition for its best these of action. Looking for Seeking Mode has the fundamental elements, which are planned as pursues like memory, dimension, and self-position by following steps.

Step 1: Collect the present position of the cat. If the value of the present position is reflecting on the same dataset, let check the originality, and then retain the present position as one of the candidates.

Step 2: For each cat, according to dimension, allocate randomly plus or minus memory allocation on the present values and replace the self-position for each memory allocation.

Step 3: Calculate the fitness values of all position of a cat.

Step 4: If all fitness is not exactly equal, calculate the selecting probability of each candidate point by equation 5.3. Otherwise, set all the selecting probability of each candidate, the point is 1.

Step 5: Randomly pick the point to move to from the candidate points and replace the position of a cat.

5.7.2 Tracing Mode

The activity of following Tracing Mode concurring can be depicted as follows,

Step 1: Update the velocities for every dimension

Step 2: Check if the velocities are in the range of maximum velocity.

Step 2.1: In case the new velocity is over-range, it is set equal to the limit.

Step 3: Update the position of a cat.

Step 4: These three steps are used into tracing mode for finding the Mixture Ratio (MR) with respect to the seeking mode and tracing mode.

5.7.3 Pseudocode of Cat Swarm Optimization Mel Frequency Cepstrum Coefficients (CSOMFCC)

The Selected Feature is given as voice input data and is sent to the analysis of Feature Extraction. The voice signals applied to the feature extraction process. Presence of MFCC analysis environments the Melcepts () function are collect the Feature Selected voice input data 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.

Input: Feature Extracted Input Voice Signal, Discrete Wavelet Transforms
Aim: extract wavelet coefficients by CSOMFCC
For features reduction
Apply Principal Component Analysis (PCA)
Cat Swarm Optimization: for the train the weights and biases of the PCA
Do: cross-validation
Seeking mode:
Find: Memory Pool, Dimension, Dimension Change, and Random Number
else Goto Tracing mode;
Tracing mode:
Find: random variable (mixture ratio)
else Goto Seeking mode;
best cat (memory) ← optimization (fitness value)
Else Cat Swarm Optimization:
Display: Cat Swarm Optimization: achieved
return: overall accuracy

In the proposed methodology, Cat Swarm Optimization with Mel Frequency Cepstrum Coefficients is combined together to extract the best features from the signals. The main aim of the feature selection and extraction process is to reduce the time and dimensionality and also to extract more useful/dominant information hidden in the signals by circumventing needless or redundant information. As the CSO algorithm converges at high speed to bring out the optimal solution, MFCC features are extracted at a minimum time using this technique. The 12 MFC coefficients extracted are listed here. DOV and Pitch are the dominant features, DOV – Degree of Voicing, to detect whether it is the voice of a human being not…and Pitch – According to the pitch frequency level, the human voice as male/female and as normal/abnormal is decided. These extracted features are stored in the database and the voice samples are trained and tested accordingly.

A comparison analysis is processed for sampled Voice data for extracting the relevant features from the speech signals. The CSOMFCC method outperforms well when compared to MFCC and LPC analysis in extracting the features shown in Table 5.1. From this consideration, the CSOMFCC method yields better performance and also to converge the speed shown in Table 5.2 to bring out the optimal solution and the average time taken to extract the features are calculated and the graph is plotted.

Table 5.1. Feature Extraction Values

LPC	MFCC	CSOMFCC
(12x25) 300	(12x21) 252	(12x17) 204

Table 5.2. Execution Time (Sec)

Datasets	LPC	MFCC	CSOMFCC
Saarbruecken	0.72	0.54	0.17
Real-Time	0.69	0.61	0.26

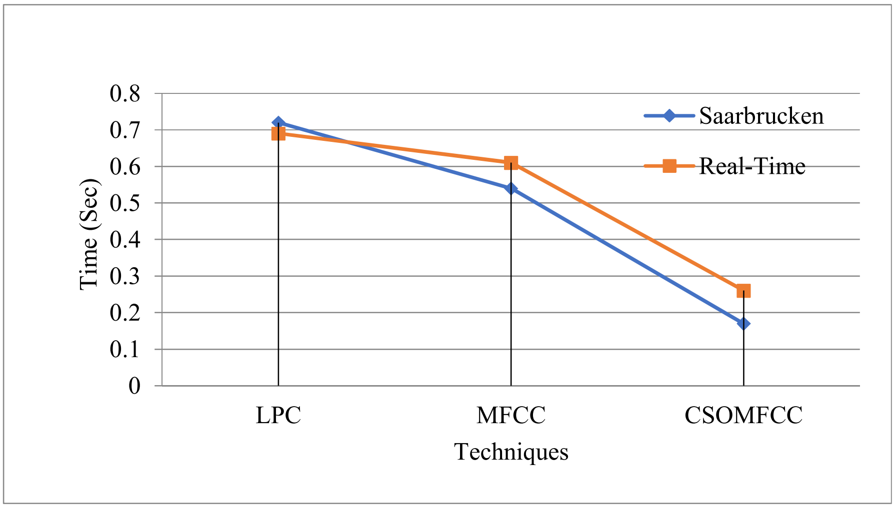

Figure 5.13. Proposed CSOMFCC Execution Time

Figure 5.13 clearly shows that the proposed technique CSOMFCC has extracted the features in reduced dimensionality and in less time. This Pseudocode of CSOMFCC presents the plotting of optimization and accuracy achievement. The first part describes the concepts of the MFCC problem and the last part describes the concepts of the CSO algorithm to the optimization problem. The proposed classification method is presented in the next chapter. A comparison analysis is done for sampled Voice data in extracting the relevant features from the speech signals. Here, The CSOMFCC method outperforms well when compared to MFCC and LPC analysis in extracting the features, The optimization technique applied to the MFCC method in order to yield better performance and also to converge the speed to bring out the optimal solution. The average time taken to extract the features is calculated and the graph is plotted. It clearly shows that the proposed technique CSOMFCC has extracted the features in reduced dimensionality and in less time.

5.8 Chapter Summary

In this research, the proposed Cat Swarm Optimization with Mel Frequency Cepstral Coefficients (CSOMFCC) methodology are combined together to extract the best features from the input voice signals. The feature selection and extraction process are performed to reduce time and dimensionality and also to extract more useful/dominant information hidden in the signals by circumventing needless or redundant information. As the CSO algorithm converges at high speed to bring out the optimal solution, MFCC features are extracted at a minimum time using CSO technique. Here, the 12 MFC coefficients extracted these extracted features are stored in the database and the voice samples are trained and tested accordingly. A comparison analysis is done for sampled Voice data in extracting the relevant features from the speech signals. Here, The CSOMFCC method outperforms well when compared to MFCC and LPC analysis in extracting the features, and also the applied optimization technique to the MFCC method in order to yield better performance and also to converge the speed to bring out the optimal solution. The average time taken to extract the features is calculated and the graph is plotted. The above graph clearly shows that the proposed technique CSOMFCC has extracted the features in reduced dimensionality and in less time. This feature set is used as input by Classification Phase of the research design, which performs voice disorder detection. The details of Classification algorithm is presented in Chapter 6, **Classification of Voiced Data into Normal and Pathological Voices.**