Assessment of Autism Spectrum Disorder in Toddlers using Speech Features

Thesis submitted in partial fulfillment of the requirement for M.Tech. Degree

Harshit Kumar Gupta
Entry No.: 2013EET2369
M.Tech.(Computer Technology)

Project Supervisor

Dr. Santanu Chaudhury
Electrical Engineering Department IIT Delhi

Dr. Nandini Chatterjee Singh
National Brain Research Center Manesar

Department of Electrical Engineering

Indian Institute of Technology Delhi

May 2015

CERTIFICATE

This is to certify that the M.Tech. Thesis titled "Assessment of Autism Spectrum Disorder in Toddlers using Speech Features" being submitted by Harshit Kumar Gupta, Entry Number 2013EET2369, is a bona fide work of him under my supervision. The matter submitted has not been replicated anywhere else for the fulfillment of any other objective.

Dr. Santanu Chaudhury
Department of Electrical Engineering
Indian Institute of Technology Delhi
Email: santanuc@ee.iitd.ernet.in.

Dr. Nandini Chatterjee Singh
National Brain Research Centre Manesar
Email: nandini@nbrc.ac.in.

May 2015

ACKNOWLEDGMENT

First of all, I would like to express my sincere gratitude towards Dr. Santanu Chaudhury for giving me an opportunity to work under him and guiding me at every juncture. His wide knowledge and logical way of thinking have been a great value for me. It is for his foresight, altruism and experience only that I was able to complete my major project in time.

I would also like to thank Dr. Nandini Chatterjee Singh of National Brain Research Centre Manesar for his valuable inputs regarding Autism Spectrum Disorder. During visits to NBRC Manesar,Gurgaon, she shared her knowledge about behavior of Autistic Children which helped me a lot.

I would also like to thanks National Brain Research Center at Manesar and UW Autism Center, University of Washington, Seattle for providing speech samples of Autistic and Typical Children for this project.

Harshit Kumar Gupta
2013EET2369

Contents

List of Figures

List of Abbreviations

ASD Autism Spectrum Disorder

TD Typical Development

DD Delayed Development

FT Fourier Transform

FFT Fast Fourier Transform

STFT Short Time Fourier Transform

DWT Discrete Wavelet Transform

DWPA Discrete Wavelet Packet Analysis

SVM Support Vector Machine

HMM Hidden Markov Model

CNN Convolutional Neural Networks

Chapter 1

Introduction

1.1 Motivation

This project focuses on early recognition of Autism Spectrum Disorder on children using speech based features. Basic idea behind this project was to use speech features to diagnose the neurological diseases. For this diagnosis, extraction of features from speech is currently emerging field. One such neurological disease is Autism Spectrum Disorder in which children lack communication and interaction ability with society. These children exhibit repetition in their activity, behavior and functioning. Children with ASD have following characteristics, delayed patterns of speech, missing or abnormal communication gestures, speech patterns diverging from normal, different speech qualities, reduced verbal communication.

It is very important to understand behavior of autistic children by acoustically monitoring them. It is an open problem for signal processing and machine learning to reliably identify autistic children by speech sample collected in acoustic environment. Some of the major challenges include simultaneously vocalizing multiple voices and background noise..

A technological solution was thus proposed by Dr. Santanu Chaudhury and Dr. Nandini Chatterjee Singh which could identify these characteristics in speech samples and also envisioned a machine Learning technique for classification of Autistic Children and Typical Children. If We could identify ASD characteristics early in children then treatment of those autistic children can be done at early age so that they can become normal children.

1.2 Outline of work done

The work done can be categorized as per the following objectives:

1.2.1 Audio Preprocessing

1. Description of dataset
2. Noise Reduction
3. Framing and Blocking
4. Windowing

1.2.2 Extraction of Acoustic Features from Speech

1. Fourier Transform
2. Gabor Transform
3. Discrete Wavelet Transform
4. Discrete Wavelet Packet Analysis

1.2.3 Classification of Autistic and Typical Children

1. Support Vector Machine
2. Random Forest
3. Hidden Markov Model
4. Convolutional Neural Network

Chapter 2

Speech Preprocessing

2.1 Description of Dataset

We performed the experiment on dataset provided from National Brain Research Centre, Manesar Gurgaon. Which was originally obtained from University of Washington, Seattle. At UW Autism Center, these speech samples were collected from ADOS (Autism Diagnostic Observation Schedule) and PCI (Parent Child Interaction). Dataset contains 40 speech samples,20 speech sample from autistic and typical children each. All audio files were 16-bit digitized and sampled at rate of 44.1 KHz. These samples are used to extract articulatory features.

2.2 Noise Reduction

Speech samples were collected in very noisy environment so we listen speech using Audacity, a free audio processing software. Then we apply wavelet based de-noising method to remove remaining noise. Then we amplify speech samples and also perform normalization on audio. audio normalization is performed to gain in amplitude of speech by stretching audio from average or peak amplitude to a target level.

2.3 Framing And Blocking

In framing we divide continuous 1 D signal into small frame of N samples, then we take next frame by shifting M samples, this causes adjacent samples to overlap by M-N samples.Choice of M and N is application specific maintaining constraint M $<$ N. We choose frame size

sufficiently small to capture significant information.

$$TotalFrames = \frac{(S - M)}{(N - M)} \tag{2.1}$$

where S is total number of samples and N is number of samples in frame and M is number of samples shift for next frame.
If we have 40 ms frame length then it carry pitch modulations information i.e it offers high frequency resolution and if we take 5 ms frame length then it carry information about change in timbre i.e. it offers high time resolution. So while working with any feature extraction technique like MFCC, STFT, DWT etc selection of these parameter is very common.

2.4 Windowing

To minimize spectral leakage at the start and end of frame it is necessary to apply windowing to frame.In this process we multiply window function and frame together. Window function is also called tapering function. There exists so many window function.

1. Rectangular Window
2. Triangular Window
3. **Hanning Window**
$$W_(n) = 0.5(1 - \cos(\frac{2\pi n}{N - 1})) \tag{2.2}$$
4. **Hamming Window** It optimizes maximum side lobe of signal.
$$W_(n) = \alpha - \beta \cos(\frac{2\pi n}{N - 1}) \tag{2.3}$$
where $\alpha = 0.54$ and $\beta = 1 - \alpha = 0.46$

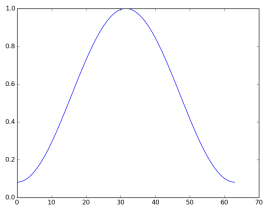

Figure 2.1: Hamming Window

Chapter 3

Description of Feature Extraction Techniques

3.1 Fourier Transform

Fourier Transform is used for conversion of one signal (function) from one basis to another like convert time domain speech signal to frequency domain signal which helps us to find contribution of each different frequency components. Sine and Cosine represent basis function in L^2 space. So we can express any periodic function in terms of sine and cosine.

Analysis formula for Discrete Fourier transform can be written as follows:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-j\frac{2\pi}{N}nk} \tag{3.1}$$

We evaluate above formula for k=0,1,... upto N-1 when we convert time domain signal to frequency domain.

According to **Parseval's Theorem**, if we change the underlying basis, the energy of signal will not change.

$$\sum_{n=0}^{N-1} |x_n|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |X_k|^2 \tag{3.2}$$

We can also get original signal back using **Synthesis formula** or **Inverse Fourier Transform** which can be written as follows for k=0,1,... upto N-1:

$$x_k = \frac{1}{N} \sum_{n=0}^{N-1} x_n e^{-j\frac{2\pi}{N}nk} \tag{3.3}$$

Calculation of Fourier Transform has $O(N^2)$ complexity which is very high computational cost when we have long signal so one faster algorithm is used for this transform which is known as **Fast Fourier Transform(FFT)**.FFT has $O(N \log_2 N)$ cost for calculation of N point signal.Points given by FFT are symmetrical so only $\frac{N}{2} + 1$ points are unique and sufficient to represent FFT, so we keep only first $\frac{N}{2} + 1$ points and discard the rest.

3.2 Mel Frequency Cepstral Coefficient

These coefficients are mainly used in speech recognition, music information recognition and genre classification. MFCC also represent state of human auditory system. Mel Frequency Cepstrum (MFC) is a power spectrum representation of signal at each window which is linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency. Group of MFC coefficient is known as Mel-Frequency Cepstral Coefficients (MFCC).

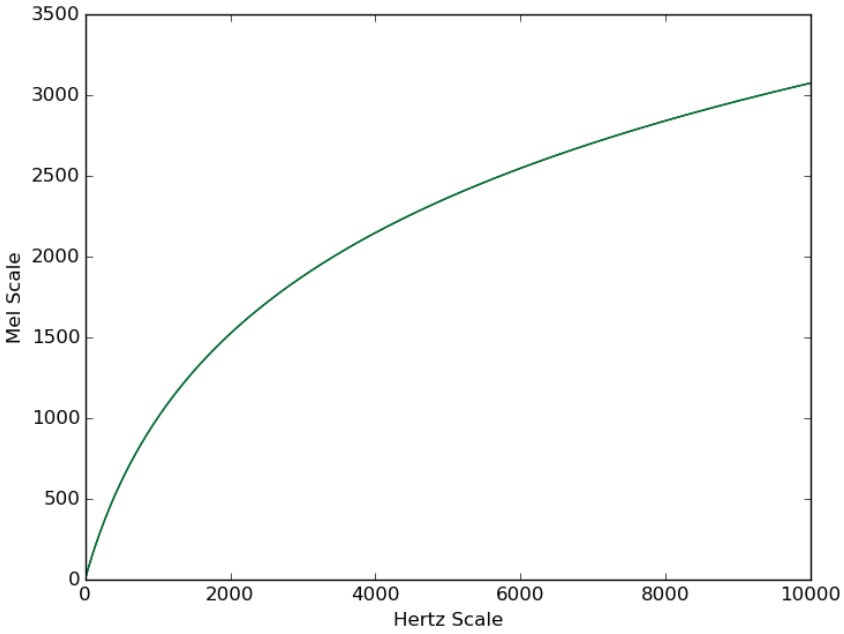

Figure 3.1: Hertz Scale vs Mel Scale

The formula to convert from frequency to Mel scale is:

$$M_f = 1125 \ln(1 + \frac{f}{700}) \tag{3.4}$$

Following is a detailed procedure of calculation of MFCC:

1. Divide the signal in short windowed frame.
2. Calculate periodogram estimate of power spectrum for each frame.
3. map power spectrum on mel frequency scale
4. take log of all mel power bands
5. take DCT of mel log power spectrum
6. keep only some first DCT coefficients, discard the rest.

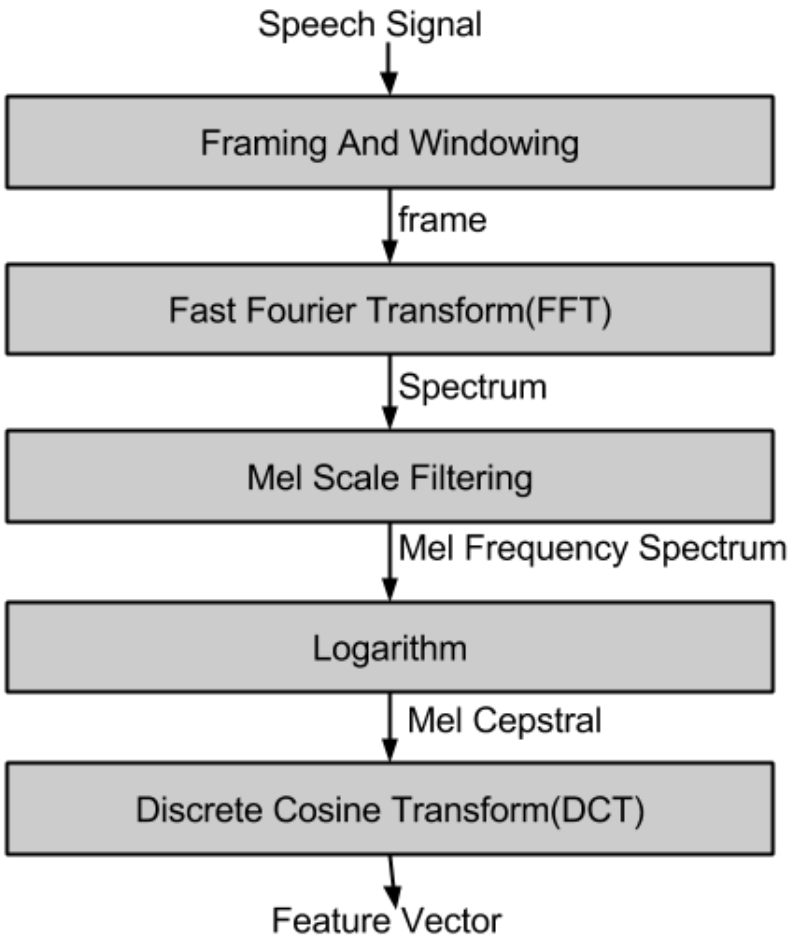

Figure 3.2: MFCC Feature Extraction

3.2.1 Discrete Cosine Transform (DCT)

DCT calculates real part of Fourier transform.First DCT coefficient is called DC coefficient which is lowest frequency in block. If we are given a list of N values in vector I then their DCT coefficients can be given by:

$$F(u) = \sqrt{\frac{2}{N}} C(u) \sum_{x=0}^{N-1} \cos(\frac{(2x+1)u\pi}{2n}) I(x) \tag{3.5}$$

for u=0,1,.....N-1

$$C(u) = \begin{cases} \frac{1}{\sqrt{2}} & \text{if } u = 0 \\ 0 & \text{otherwise} \end{cases}$$

3.3 Gabor Transform

When we analyse signal in frequency domain then information in time domain is lost. we can't know which frequency component occur at which time. As **Heisenberg's Uncertainty Principle** suggests that we can not have good resolution both in time and frequency domain. Gabor transform allow us to have frequency resolution within a time window. If we use skinny time window for analysis then we capture high frequencies with time localization but if we use fat window then we capture low frequencies which are not localized in time. So we take a time window like hanning or hamming and find Fourier transform of signal only in that window. That's why this gabor transform is also known as **Short Time Fourier Transform**.

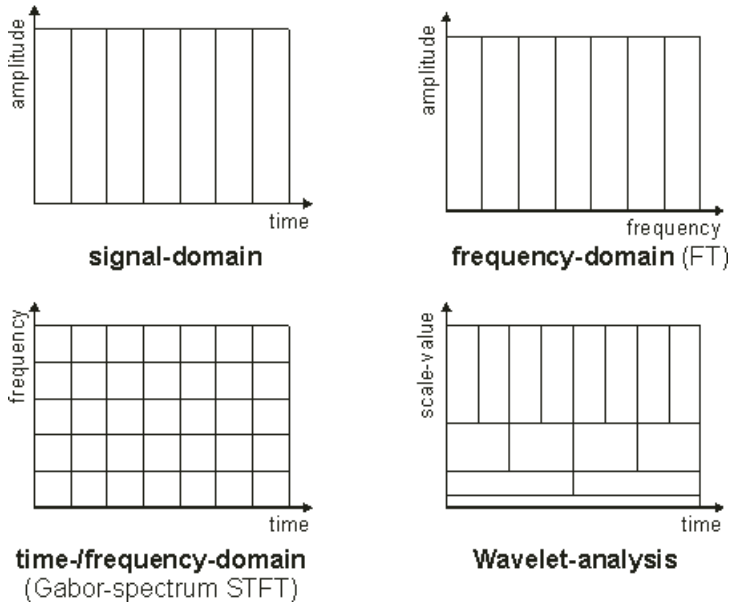

Figure 3.3: Signal resolution at different domains

3.3.1 Spectrogram

Spectrogram offers us visualization of spectral components w.r.t. time. It is actually a time vs frequency representation of signal. but this is not accurate resolution instead we use time window to calculate STFT of signal. Then we plot frequencies in each time window. In this color depth of each point indicate no of times frequency occurring in that time window.

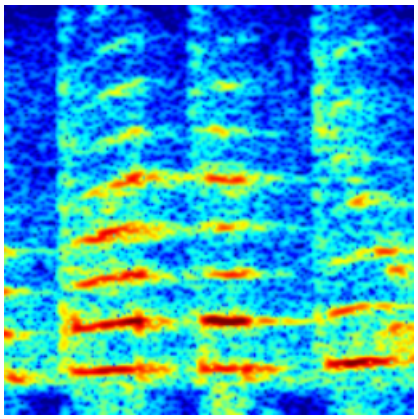

Figure 3.4: Spectrogram of autistic child

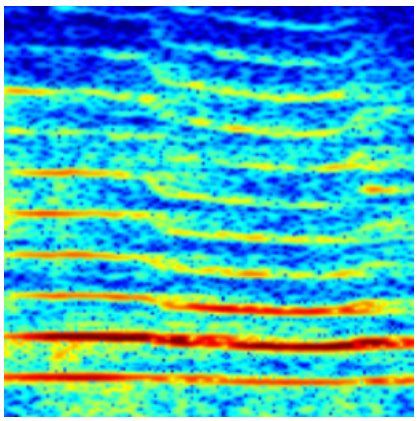

Figure 3.5: Spectrogram of typical child

3.4 Wavelet Transform

Wavelet are functions which provide an orthonormal basis for function in L^2 space like sine and cosine are orthonormal basis in Fourier transform. Wavelet provides us multi-resolution analysis.

All wavelets are derived from mother wavelet, so wavelet with scale s and time τ is equal to mother wavelet normalized by square root of scale s and shifted in time by τ and change in scale s

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}}\psi(\frac{t-\tau}{s}) \tag{3.7}$$

where s represents scale and τ represents time.

Formula of **wavelet transform** can be written as follows:

$$\gamma(s,\tau) = \int f(t)\psi^*_{s,\tau}(t)dt \tag{3.8}$$

where $\psi^*_{s,\tau}$ represent complex conjugate of mother wavelet.

Formula of **inverse wavelet transform** can be written as follows:

$$f(t) = \int \int \gamma(s,\tau)\psi_{s,\tau}(t)d\tau ds \tag{3.9}$$

3.4.1 Properties of Wavelets

Wavelet basis function has following properties:

1. Orthogonal (basis function are perpendicular to each other)
2. Each vector is normal vector
3. Output of transform will have same energy as input.
4. Sparsity (so many coefficients are zero)
5. Linear Time Complexity for transformation.

3.4.2 Wavelet Families

Wavelets are organized into group called Wavelet Family. Most commonly used families are:

1. **Haar** : These are used for extracting image feature for object recognition and real-time face detector.

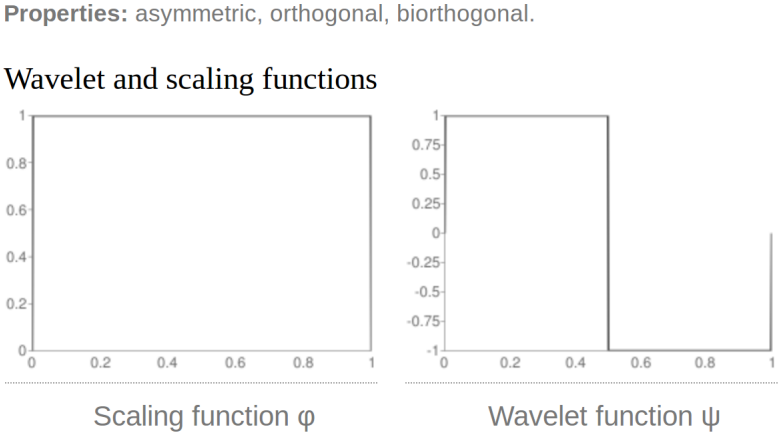

Figure 3.6: Haar Wavelet

2. **Daubechies** : These are used for image or speech denoising and feature extraction for speech classification.

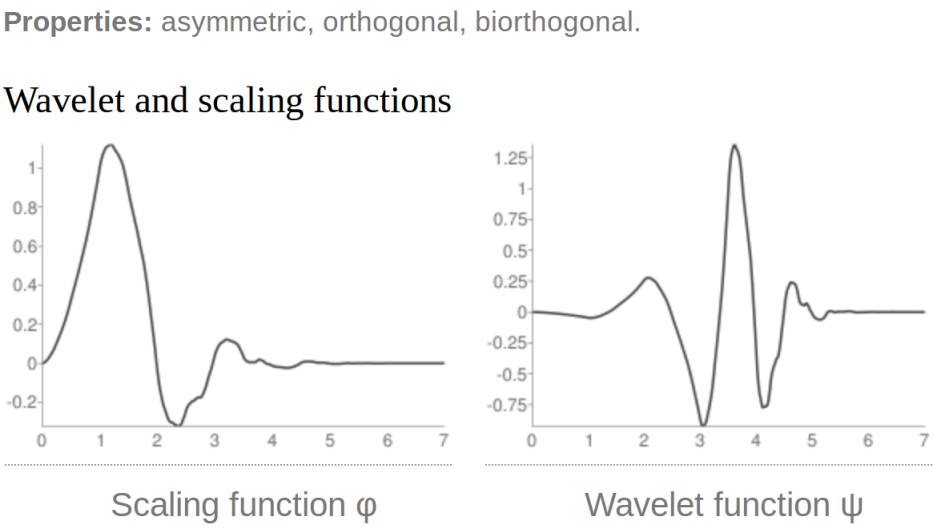

Figure 3.7: Daubechies 4 Wavelet

3. Symlets
4. Coiflets
5. Biorthogonal
6. Reverse Biorthogonal
7. "Discrete" Meyer

3.4.3 Types of Wavelet Coefficient

After wavelet Transform we find two types of coefficients:

1. **Approximation Coefficients** : these coefficient represent high scale, low frequency component of signal. We calculate these coefficients by applying low decomposition filter (LDF) on signal and then applying down-sampling by factor of 2.
2. **Detail Coefficients** : these coefficient represent low scale, high frequency component of signal. We calculate these coefficients by applying high decomposition filter (HDF) on signal and then applying down-sampling by factor of 2.

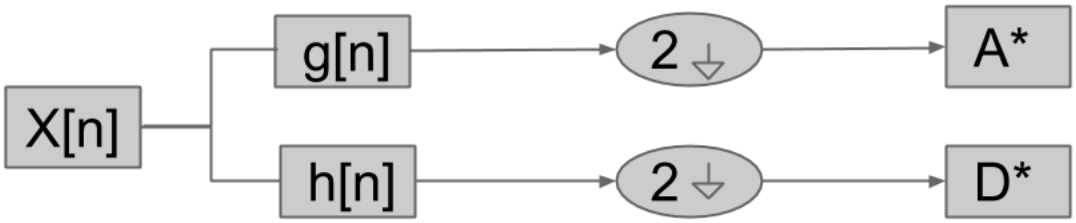

Figure 3.8: Decomposition filer on signal

3.4.4 Significance of Wavelet Coefficients

Wavelet transform allows us exceptional localization in both time domain via translation of mother wavelet and in scale (frequency) domain via dilations. Translation and dilation operations are applied to mother wavelet to calculate wavelet coefficient which represent similarity (co-relation) between wavelet and localized section of signal. So wavelet transform better resolves high frequency component in temporal resolution and low frequency component in scale resolution.

3.4.5 Scalogram

Scalogram is a time vs scale representation of signal. For this we do cross-correlation of chosen wavelet and our signal then plot result, this is scale-1. Next we dilate wavelet (stretch it by some factor) then again do cross-correlation of this new waveform with our signal then we get scale-2 plot. So scalogram shows result of performing a cross-correlation of signal with wavelet at different scale (dilation and stretch factor) In this plot bright spot means we get a good cross-correlation score between stretched wavelet and signal. Bright spot indicate where peaks and valleys of stretched and shifted signal align best with peaks and valley of signal. Dark means no alignment, dimmer means some peaks and valleys line up but brightest where all peaks and valley align.

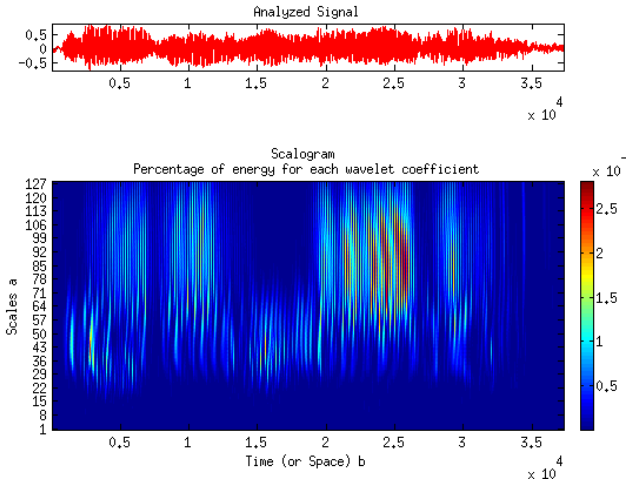

Figure 3.9: Scalogram of autistic child

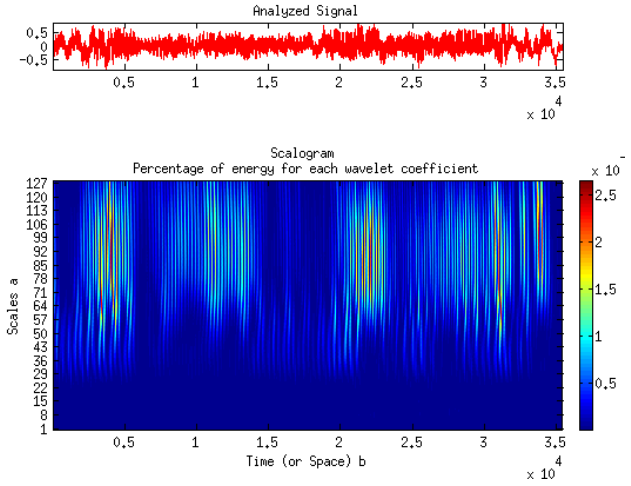

Figure 3.10: Scalogram of typical child

3.4.6 Discrete Wavelet Transform

Discrete Wavelet Transform utilizes low frequency component (approximation coefficient). In this transform we iteratively perform low pass filtering and high pass filtering at each level on approximation coefficient only. This multiple level decomposition is also called Wavelet decomposition tree.

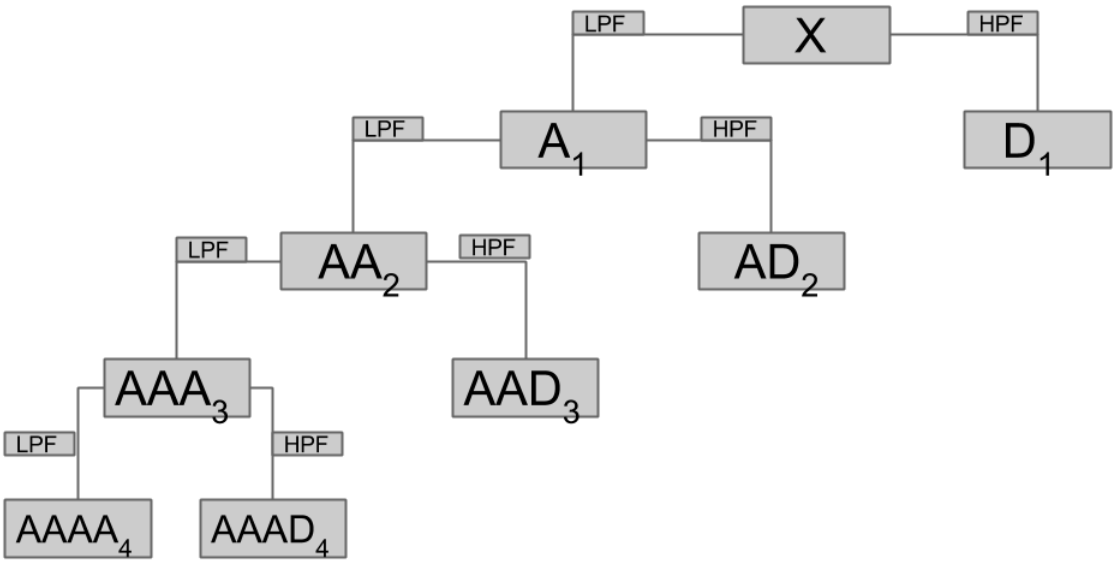

Figure 3.11: Discrete Wavelet Transform

3.4.7 Discrete Wavelet Packet Analysis

Discrete Wavelet Packet Analysis utilizes both low frequency component (approximation coefficient) and high frequency component (details coefficient). In this transform we iteratively perform low pass filtering and high pass filtering at each level.

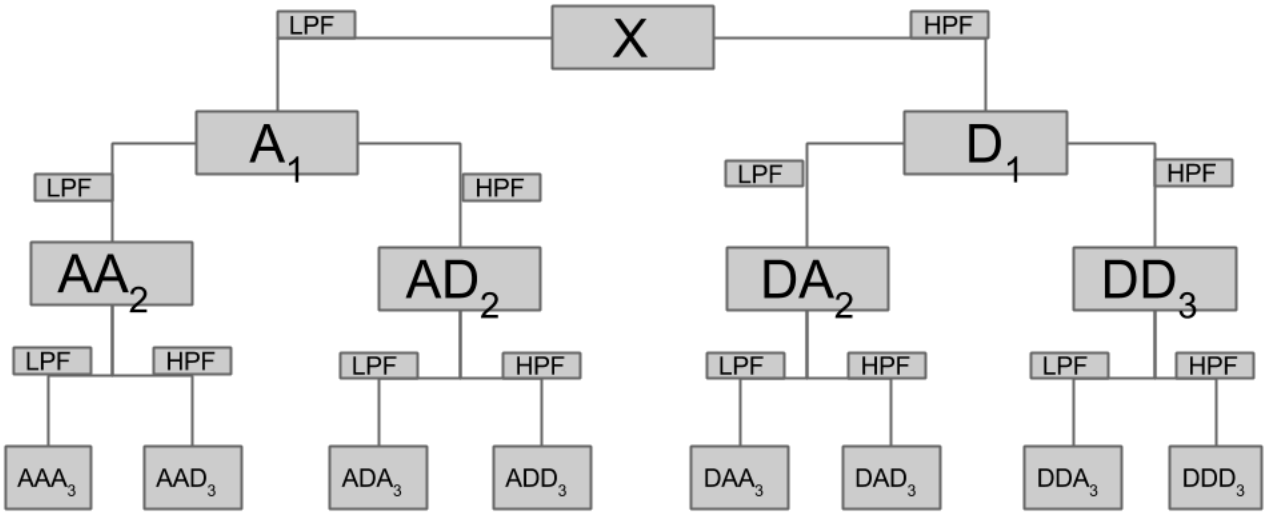

Figure 3.12: Discrete Wavelet Packet Analysis

Chapter 4

Classification Algorithms

4.1 Support Vector Machine (SVM)

This classifier constructs a hyperplane or a set of hyperplane in high or infinite dimensional space. This classifier minimize empirical classification error and maximize geometric margin. this is also known as maximum margin classifier. It can also implement non-linear classifiers by applying kernel trick to maximum margin classifier. There are two types of SVM classifiers:

1. C-SVM
2. nu-SVM

4.1.1 C-SVM

In this SVM we try to minimize following error function during training.

$$E(w) = \frac{1}{2}w^T w + C\sum_{i=0}^{N}\xi_i \tag{4.1}$$

Error function E(w) is subjected to following constraints: $y_i(w^T\phi(x_i + b) \leq 1 - \xi_i$ and $\xi_i \leq 0$

4.1.2 nu-SVM

In this SVM we try to minimize following error function during training.

$$E(w) = \frac{1}{2}w^T w - \upsilon\rho + \frac{1}{N}\sum_{i=0}^{N}\xi_i \tag{4.2}$$

Error function E(w) is subjected to following constraints: $y_i(w^T\phi(x_i + b) \leq \rho - \xi_i$ and $\xi_i \leq 0$ and $\rho \leq 0$

4.1.3 Advantages of SVM

Following are main advantage of SVM.

1. works well in very high dimensional data.
2. if number of dimensions is greater than number of samples, then still it works fine.
3. only a subset of points contribute in decision making, so it takes very less space on memory and those points are called support vector.
4. supports a variety of kernel functions.

4.1.4 Disadvantages of SVM

Following are main disadvantage of SVM.

1. if number of features is greater than number of samples, then model is likely to give poor performance.
2. no direct estimation of probability

4.2 Random Forest

Random Forest is an ensemble classifier that consists of many decision trees and its output is the mode of classes that is output of individual decision trees. Random forest combines random selection of feature and bagging.**Bagging (Bootstrap Aggregation)** of classification of decision trees reduce the bias of single tree. It uses permutation to determine variable importance. We assume all trees are drawn from identical distribution which minimize loss function at each node in given tree.

4.2.1 Terminology used in Random Forest

Out of Bag (OOB) Error Rate: there is no need for separate testing or cross validation in random forest. We calculate how many times predicted class from it is not equal to true class from votes of each built tree and when this error is averaged over all cases called OOB error.
Variable Importance: Subtract votes-count for correct in the variable-m-permuted oob data from votes-count for correct class in the untouched oob data. Importance score for variable m is then obtained local importance score averaged over all tree.

4.2.2 Algorithm of Random Forest

Algorithm of random forest can be written as follows:
For each tree perform following 1 to 3 tasks

1. Fit decision tress minimizing loss function using 2/3 of samples.
2. Predict classes of remaining samples and calculate the misclassification rate which is equal to out of bag error rate.
3. we permute the variables and calculate the out of bag error. An increase in out of bag error from original indicate the importance of variable.
4. Aggregate oob error and importance measures from all trees to determine overall oob error rate and Variable Importance measure.

4.2.3 Advantages of Random Forest

Advantages of random forest can be written as follows:

1. Very accurate
2. Efficient for large databases
3. Handles so many features without feature deletion.
4. Estimation of variable importance in classification
5. Reports internal unbiased estimate of the generalization error
6. Balance between high bias and high variance.
7. Estimate missing data
8. Accurate even if large amount of data is missing

4.2.4 Disadvantages of Random Forest

Disadvantages of Random forest can be written as follows:

1. Over-fit for noisy classification dataset
2. Biased in of categorical variable with different number of levels.
3. Variable importance score for categorical variable is not reliable.

4.3 Hidden Markov Model

Hidden Markov model is a generative probabilistic model. In this model sequence of internal hidden state generate a sequence of observations.We can not observe hidden states directly. We assume transitions between hidden states is first order Markov Chain. The hidden states can not be observed directly. The transitions between hidden states are assumed to have the form of a first-order Markov chain.

4.3.1 Notation of HMM

Notations used for HMM are as follows:

1. **N**: number of possible states
2. **M**: number of possible observations
3. **X**: $\{q_1,...q_N\}$(finite set of states)
4. **O**: $\{v_1,...,v_M\}$ (finite set of observations)
5. X_t: random variable denoting the state at time t (state variable)
6. O_t: random variable denoting the observation at time t (output variable)
7. σ: $o_1,...,o_T$ (sequence of actual observations)

4.3.2 Parameters of HMM

Parameters for hidden Markov model can be given as follows:

1. **transition probabilities** $A = [a_{ij}]$

$$a_{ij} = Pr(X_{t+1} = q_j | X_t = q_i) \tag{4.3}$$

2. **observation probabilities** $B = [b_i]$

$$b_i(k) = Pr(O_t = v_k | X_t = q_i) \tag{4.4}$$

3. **initial state distribution** $\pi = [\pi_i]$

$$\pi_i = Pr(X_0 = q_i) \tag{4.5}$$

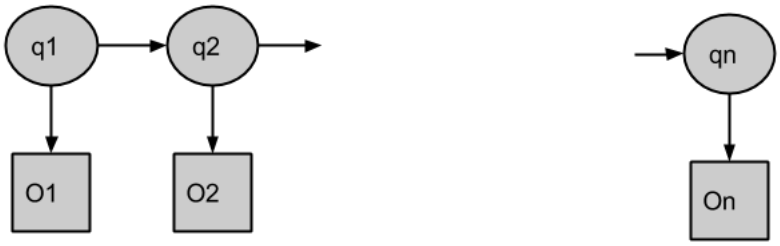

Figure 4.1: Hidden Markov Model

4.3.3 Fundamental Problems of HMM

Fundamental problems for hidden Markov model are as follows:

1. **Evaluation Problem:** Estimate optimal sequence of hidden states, given model parameters and observed data

2. **Decoding Problem:**Calculate the likelihood of data, given the model parameters and observed data i.e. it uncovers the hidden part

3. **Learning Problem:** Estimate the model parameters, given just observed data. Maximize $Pr(O|\lambda)$ given model parameter $= \{A, B, \pi\}$

Following algorithms are used for solution of HMM problems.

1. **Forward-Backward algorithm** Evaluation problem can be solved using dynamic programming technique known as Forward-Backward algorithm. This gives us most likely obervation sequence

2. **Viterbi Algorithm** Decoding problem can be solved using dynamic programming technique known as Viterbi algorithm.

3. **Baum-Welch algorithm** Learning problem can be solved by an iterative Expectation-Maximization (EM) algorithm, known as the Baum-Welch algorithm.

4.4 Convolutional Neural Network

These are also called as convnet. These are supervised deep neural networks. This approach is inspired from working of human visual perspective feeds.

4.4.1 Convolution

This is an operation on two function f and g which produce filtered version of f using filter g. Convolution perform smoothing or sharpening. When it is applied to 2-D functions it is used for feature detection, edge finding, image matching, motion detection etc.

$$f(x) * g(x) = \int_{-\infty}^{\infty} f(\tau)g(x-\tau)d\tau \tag{4.6}$$

$$f(x,y) * g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\tau_1,\tau_2)g(x-\tau_1,y-\tau_2)d\tau_1 d\tau 2 \tag{4.7}$$

4.4.2 Characteristics of CNN

This approach is currently dominating on other deep learning approach because:

1. **Sparse Connectivity**: this allow us to use different copies of same feature detector with different positions.
2. **Weight Sharing**: This is also known as **Feature Maps** which allows us to use several different feature types, each with its own map of replicated weights.
3. **Sub Sampling**: This is achieved with the help of averaging or max pooling feature maps.

4.4.3 Types of layers in CNN

1. Convolution layer
2. Dropout layer
3. Pooling layer
4. ReLU layer
5. Drop layer

4.4.4 Advantages of CNN

1. **Equivalent activities**: Replicated features do not make the neural activities invariant to translation, They make activities to be equivalent.

2. **Invariant knowledge**: During training if some feature is useful at some position then during testing also feature at that position will be useful.

Chapter 5

Methods used for classification

5.1 Classification using SVM and Random Forest with MFCC

5.1.1 Feature Construction from MFCC coefficients

We apply MFCC feature extraction technique speech frames to feature. Then we divide MFCC Feature into histogram bins. We use frequency of each bin as feature. MFCC features are state of art for classification tasks. Depending on scenario we have to use feature construction technique from MFCC features.

5.1.2 Selection of Parameters of Classification Model

5.1.3 Support vector Machine

For selection of parameters we build different models with different parameters and select the model which gives best results. Following are different parameters which we tuned by exhaustive search.

1. **kernel :** linear,polynomial, radial basis function (RBF) kernel
2. **C :** Penalty Parameter
3. **gamma :** Kernel coefficient for poly and RBF kernel
4. **degree :** Degree of polynomial kernel

5.1.4 Random Forest

For selection of parameters we build different models with different parameters and select the model which gives best results. Following are different parameters which we tuned by exhaustive search.

1. **No of Estimators :** No of trees to be build for classification

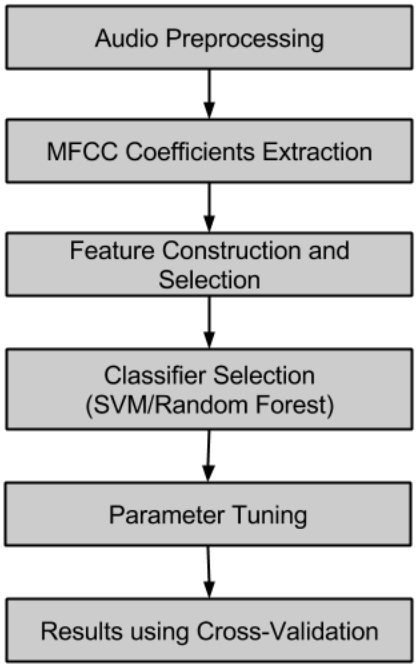

Figure 5.1: Classification with MFCC features

5.2 Classification using SVM and Random Forest with DWT and DWPA

5.2.1 Feature Construction from DWT or DWPA coefficients

We apply DWT or DWPA on speech frames to get approximation and detail coefficients on several level of decomposition. Then we take sum and variance of coefficients of frames of each level. Then we take difference between these sum and variance of successive frames.

if F is all approximation or detail coefficients after LPF and HPF and i represent frame number then feature is constructed from following:

1. mean(F_i)
2. std(F_i)
3. mean(F_i) - mean(F_{i-1})
4. std(F_i) - std(F_{i-1})
5. mean(F_{i+1}) - $2 * mean(F_i)$ + mean(F_{i+1})
6. std(F_{i+1}) - $2 * std(F_i)$ + std(F_{i+1})

5.2.2 Selection of Number of decomposition levels and Mother wavelet

We should choose mother wavelet which should be computationally efficient and provide best distinguishing characteristics. we perform experiments from haar, Symlet, Daubechies wavelet using different levels from 1 to 8. Then number of decomposition level and mother wavelet is decided by their performance. We experimented a lot get best number of level of decomposition and best mother wavelet. We find out Daubechies-8 (db8) with 5 level of decomposition gives best results with classifiers.

5.2.3 Selection of Parameters of Classification Model

5.2.4 Support vector Machine

For selection of parameters we build different models with different parameters and select the model which gives best results. Following are different parameters which we tuned by

exhaustive search.

1. **kernel :** linear,polynomial, radial basis function (RBF) kernel
2. **C :** Penalty Parameter
3. **gamma :** Kernel coefficient for poly and RBF kernel
4. **degree :** Degree of polynomial kernel

5.2.5 Random Forest

For selection of parameters we build different models with different parameters and select the model which gives best results. Following are different parameters which we tuned by exhaustive search.

1. **No of Estimators :** No of trees to be build for classification

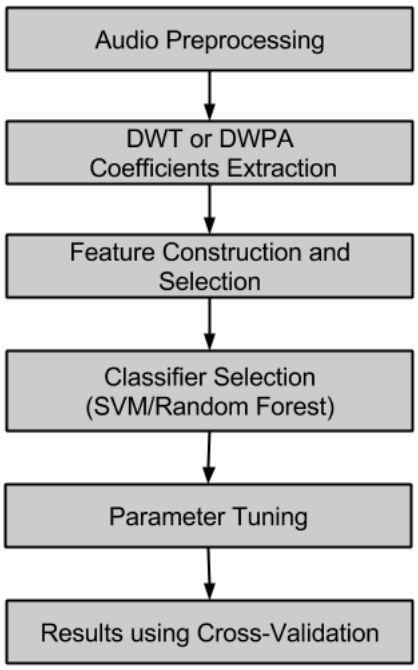

Figure 5.2: Classification with DWT/DWPA features

5.3 Classification using HMM

5.3.1 Feature Construction using Peaks Detection

I have extracted features by detecting peaks in frequency for each frame of signal. For feature extraction from speech has moved current art of state to Deep learning but still these feature works fine with audio. If we use STFT for finding peaks in frequency. For STFT we apply FFT to each frame so we have used FFTWFFTW05 routines which are very fast for calculation of FFT.

5.3.2 Algorithm for Peak Detection

1. Let X be the length of window on each frame which we use on signal.
2. Divide the window in three parts like left,right and center.
3. Use Max function on each of 3 parts of window.
4. if function value on center is greater than function value on left and right parts then goto next condition otherwise goto 6.
5. we have find peak if function value on center is actually on middle position.
6. move to the next frame and repeat whole process.
7. when we process whole data then we sort all obtained peaks in decreasing order by their amplitude.
8. Output only N=6 peaks from each frame.

5.3.3 Gaussian Mixture Model

Gaussian (Normal) distribution is widely used continuous distribution. It is represented by 2 parameters mean (μ) and variance (σ^2). Probability distribution function with Gaussian can be written as follows:

$$Pr(x|\mu,\sigma) = N(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{-(x-\mu)^2}{2\sigma^2} (5.1)$$

5.3.4 Working of HMM model

We will use Maximum Likelihood estimation (MLE) to maximize probabilities of estimations and learn HMM parameters. In this classification task only Forward Backward Algorithm and Baum-Welch Algorithm is needed.

The required tasks have been accomplished through implementation of various combination of feature extraction algorithm from signal processing and classification algorithms of machine learning. We have found peaks in frequency but we need probabilities to train HMM , so we will normalize the observations. Each frame has a set of peaks. We can obtain state probabilities by dividing each frame peaks by sum of all frame peaks. These probabilities will show distribution of peaks. Equal peaks will cause equal probabilities, and unique distribution will lead to unique set of attributes.

HMM learns transition probabilities between frames using state probabilities obtained from peaks. Now we will train 2 separate HMM-GMM model one for autistic class and one for typical class. When we need to predict the class, we give features of test samples to both model, the model which give higher probabilities for maximum likelihood is the predicted class.

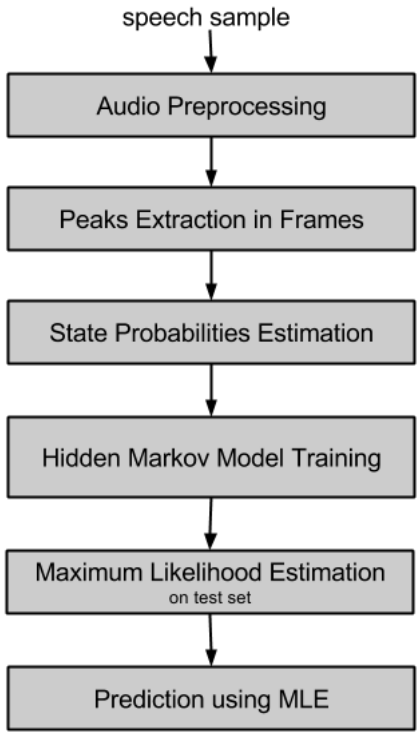

Figure 5.3: Classification with HMM Model

5.4 Classification with CNN

1. **Convolutional layer** In this type of neural network, convolution kernel are trained by back-proagation algorithm while previously kernel or filter were fixed like Gaussian blur,sober etc. We have so many kernels in each layer and each of them works on entire image with weight sharing. Convolution operation extract necessary information from image. The capacity of a neural net varies, depending on the number of layers. Initial layers extract information like edges, line. Next layers combine those low level features into curves, shapes like feature. At last these layers represent object level features. The first convolution layers will obtain the low-level features, like edges, lines and corners.

2. **Pooling layer** This ensure activity invariance by producing same result even if oject has some translation. This layer either compute maximum or average to reduce variance. When we take maximum then this is also called "MAX Pooling Layer".

3. **Dropout layer** Fully connected Layers has large number of parameters which causes over-fitting. Dropout improves training speed. It is like dropping out neurons which in turn will not contribute to forward pass and back propagation.

4. **Rectified Linear Units(ReLU) layer** This layer introduces non-linearity into model without affecting receptive fields of convolution layer otherwise without these output of model with many layers will be just a linear combination of inputs. There are so many function activation function which introduce non-linearity such as-

   (a) **ReLU** $f(x) = max(0, x)$

   (b) **hyperbolic tangent** $f(x) = |tanh(x)|$

   (c) **sigmoid function** $f(x) = \frac{1}{1+e^{-x}}$

5. **Loss layer** We have different loss functions for different scenario such as-

   (a) **Softmax loss :** for predicting single class out of K mutually exclusive class.

   (b) **Sigmoid cross-entropy loss** for predicting single class out of K independent class.

   (c) **Euclidean loss** for regression to real-valued labels.

5.5 Results

Results produced by any classification model depends on quality of features and numbers of samples. As we have 40 samples of speech to cross validate any of above model. So we can not strongly suggest that these are best classification results. Proposed classification model and their accuracy on data-set with 5-fold cross validation is given below:

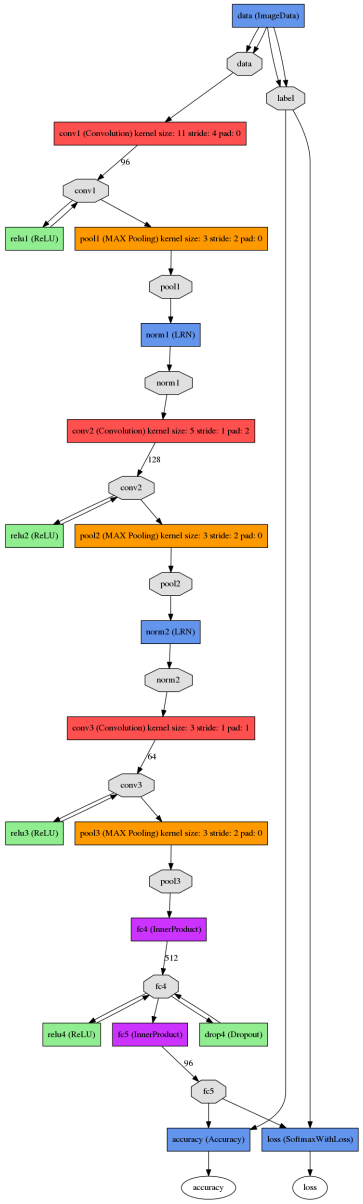

Figure 5.4: Classification with CNN

Proposed Model	Accuracy
MFCC features with SVM	60%
MFCC features with Random Forest	65%
DWT features with SVM	80%
DWT features with Random Forest	77.5%
DWPA features with SVM	85%
DWPA features with Random Forest	80%
STFT Features with HMM	75%
Spectrogram with CNN*	55%

Figure 5.5: Classification Results
* As Convolution Neural Network is deep network which needs large data set to be trained, so this model will surely perform better if we use more number of samples to train.

Chapter 6

Future Work

The future work in this project can be summarized as follows:

1. Collect more data set using cloud based android application which is released at www.diseaseprediction.url.ph

2. Implement sequential deep learning approach for classification such as Recurrent Neural Network. Emerging state of art for classification in this field id RNN-HMM model which is combination of Recurrent Neural Network with Hidden Markov Model.

Bibliography

harshit

ABOUT THE AUTHOR

The author is a student of M.Tech (Computer Technology) under Electrical Engineering Department at Indian Institute of Technology Delhi at the time of this writing. He completed his B.Tech with honors in Computer Science of Engineering from Kamla Nehru Institute of Technology Sultanpur Uttar Pradesh Technical University. The author is interested to work in the areas of Machine Learning, Data Analytics and Deep Learning. He is going to join McAfee (Intel Security) as a Sr. Software Engineer.