ISSN 2347 - 3983
Volume 8. No. 6, June 2020
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter101862020.pdf
https://doi.org/10.30534/ijeter/2020/101862020

Functional Analysis and Hybrid Optimal Cepstrum Approach for Gender Classification Using Machine Learning

Anil Kumar Maddali[1], Habibulla Khan[2],
[1] Assistant Professor, Department of Electronics & Communication Engineering, Koneru Lakshmaiah Education Foundation; Vaddeswaram, A.P., India; mak71282@gmail.com
[2] Professor, Department of Electronics & Communication Engineering, Koneru Lakshmaiah Education Foundation; Vaddeswaram, A.P., India; habibulla@kluniversity.in

ABSTRACT

Speech recognition improvises the real time analytical need and importance for different application from daily activities to automated world where each operations are controlled either by voice or sound. The current design for gender and its iterative algorithms improvise a system that would generate different scenarios engulfing the real time aspects of the algorithm on machine modelling and functional characteristics on the databases chosen. Our proposed model provides two-step process for audio data either to consider the different voices from database (KAGGLE) improvising the different sections of the design model to ensure the data parameters are tabulated. These parameters would suggest the difference between the SNR from the noised or recorded samples from mobiles or any other recorded devices. Database Acquisition of the voices samples are assigned with male and female pitch signal depending upon the algorithm utilized. Our proposed model utilizes ensemble feature of SVM and KNN promising the better performance results on each database chosen and mentioned in section V.

With 10 samples of MALE and FEMALE voice speech data have been evaluated to recognize the gender as MALE or FEMALE. The results of the current design with SVM, KNN and HOC approach have been tabulated to depict that the proposed design model has overall 3% accuracy improvement on KNN and SVM as shown in section VI.

Key words*:* Audacity, Short time autocorrelation, Center clipping, average magnitude difference function (AMDF), cepstrum, harmonic product spectrum, normalized cross correlation(NCCF), Harmonic product spectrum, Logistic predictive coefficients, Spectral Entropy ($fspent$), Spectral Flatness($fsfm$), Mode frequency($fmode$), Frequency centroid($fcentroid$), Mean fundamental frequency($Fmean$), Minimum fundamental frequency($Fmin$),Maximum fundamental frequency($Fmax$) etc.

1.INTRODUCTION

In accordance with the design capabilities we have seen different subsystem that would provide the current features that would suffice the requirements from the customer. One such system of the design would be the Gender Classification and recognition approach to analyse the speech or voice samples. Existing scenarios improvises the usage of MFCC and SVM, MFCC and KNN models to ensure the voice data recognition. Machine leaning approaches and its implementation with respect classification models and its deep learning architectures utilized for the work in [2]. For improvising such model and its implementation characteristics on the current design platform on MATLAB, we have ensure to build the train and test set models improvising 10 different samples on each male and female data set from (KAGGLE). Existing work on the differences between male and female speech have been examined with different parametric criteria such as (e.g vocal tract length) [1], phonetic [3], and quality of voice differences [4]. The perception of the pitches of the different samples would provide the fundamental frequency will be major asset to speech recognition.

The outline of the current paper presents about six stages of the design on Gender classification and recognitions. Section I describes about the importance on the how classification is done and its current relations on ASR. Section II-III imparts the literature survey and design modelling on the existing scenarios for each set of algorithms utilized. Section IV-V improvises on the detail modelling and classification scenario on noise and speeches for male and female characteristics. Section VI, provisions on results and descriptions on each set of algorithms and current proposed algorithm comparisons. Finally we conclude the design with its scope of the work.

2.LITERATURE SURVEY

Jisung Wang, Sangki Kim, Yeha Lee, [1] has augmented the data analysis and its overfitting scenario avoiding problem have been crucial to the current learning algorithms considered. ASR with modelling with WAVENET proposal

would diverse the pitch related pattern observations to reduce the usage of acoustic features. Speed perturbation and WORLD –voice based conversion models would enhance the design performance.

Samik Sadhu, Ruizhi Li, Hynek Hermansky, [2], improvised ASR with a novel scenario on frequency bands and frame level features with M-vectors and Mel-Frequency-Cepstral Coefficient models ensuring the boosting of the performance with 5% with compared on the algorithms HMM-GMM ASR about 18 percent over traditional MFCC features.

Andrew Rosenberg, Yu Zhang, Bhuvana Ramabhadran ,Ye Jia, Pedro Moreno, Yonghui Wu, Zelin Wu [3] have modelled and design a system for human speeches, which are trained to establish the correctness of the speeches and its real time importance with tacotron speech analysis and its architectures proving naturalized synthesized voices. The speech enhancing capabilities are augmented with robust speech recognitions ensuring performance gap on synthesized and trained speeches.

Hainan Xu, Shuoyang Ding, Shinji Watanabe, [4] have implemented on speeches for End to end recognition model, requiring sub-words of extraction of the data processed. These data with its frequencies are might be inferior or lead to erroneous data at speech output. To implement such problem PASM technique is implemented with sub word extraction from the current word leverages of pronunciations.

Burnett, M., & Kulesza, T [5] have utilized the functionality of the communication model for each set of data sending from the sender to the receiver at each phase on data stream. The data optimized with current machine learning models for E-E connection on IOT systems improvising the sender and receiver frequency data.

Childers and Bae [6] have initiated the speaker and its performance characteristics with recognition model improvising the gender and its features on the real time scale in 1988.

Sigmund, M. [7] with improvements on the digital world, additional functionalities for the speeches with Cepstral Coefficient (CC), prosodic feature, formant (F1, F2, F3) have been proven for better functional capabilities.

Tolba, H. [8] have presented CHM model to impart with different recognition capabilities on the words or syllable for the continuous signal speech applied to the design environment in recent decade.

Chisaki, Y., Nakashima, H., Shiroshita, S., Usagawa, T., & Ebata, M, [9] have implemented the pitch analysis and its configurational parameters to estimate speech data.

Sigmund, M, [10] has modelled and implemented the effective recognition model with greater than 90% of accuracy with short pitches and speeches. For larger scenarios 93% have been estimated as per the design requirement.

Ali, M. S., Islam, M. S., and Hossain, [11] with and accuracy of 90% in small and quit environment recognized the voice samples with use of the SVM and GMM model in speech analysis. The obtained accuracy have been improved with respect to HMM as GMM is sub-module and is convenient with better robustness and computational capabilities. Wavelet, TF, and NN modelling have been improvised with designed model.

Bissell, C.,[12]. M. A.in 2012 accuracy of 80% is observed for the modelling frequency spectrum analysis ensuring the correct frequency values observed.

Erokyar, H,[13] in 2014 obtained the results data and recognized speeches with 64.2% accuracy on 108 experiments.

3.DESIGN MODEL PURPOSE

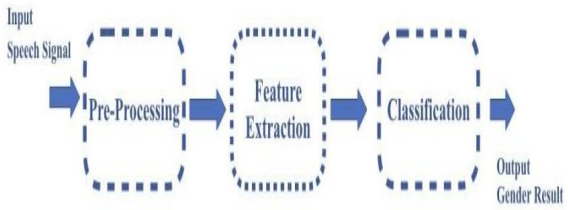

Figure 1: Representing the existing model design for gender recognition

Our design model scenario with accordance with the figure-1 comprises three stage models and its implementation algorithms to enhance specific criteria's of the user or designer choices. Each stage could provide different set and capabilities of the important part of the research work especially when we deal with audio signals. Improvising the specific futuristic abilities on the design model would suffice the implementation stage for each sub stage occurrence.

As per the above figure we have seen that PRE-POCESSING is the first key stage on which each database signals are being analysed and implement the related algorithms depending on user choices. Noise avoidance is key step for the pre-processing stage indulging the different noise models and its relational capabilities ensuring the correct feature of noise reductions based on the adaptive model chosen. To ensure such system changes we have approach a FIR filter with Hybrid Adaptive filtering on the noise model chosen in mathematical modelling in section five. Feature Extraction is other modelling area where different set of parametric features such as:

i. Mean frequency ($fmean$),
ii. Mode frequency($fmode$),

iii. Frequency centroid($fcentroid$),
iv. Mean fundamental frequency($Fmean$),
v. Minimum fundamental frequency($Fmin$),
vi. Maximum fundamental frequency ($Fmax$).

4.AUTOMATIC SPEECH RECOGNITION SYSTEMS USING MFCC AND FFT

4.1 Design Block Diagram

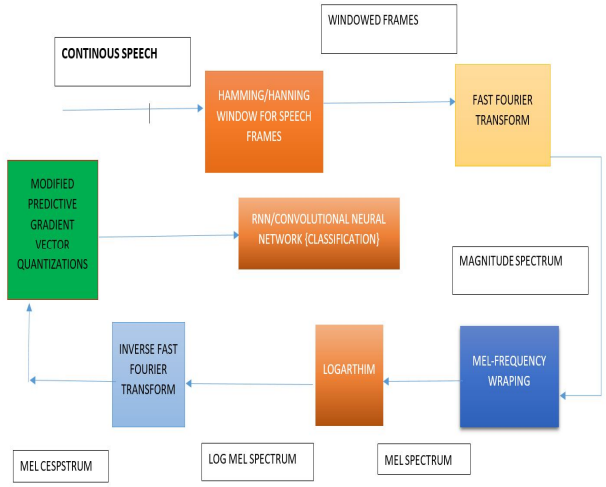

Figure 2: Representing the Existing block diagram for audio speech recognition

Firstly, we initiate the preprocessing phase consisting of Windowing techniques resulting better frames for each design scenario. Acknowledging the concept and observed the figure-2, we have modelled the existing design and its implementation analysis which has effects on the current design scenario specifically leading to analyze the different block involving the transforms, mel-spectrum, and finally optimization and clustering. Now, in our design we have proposed specific changes in the algorithms utilized in current design scenario to improve the capabilities of the optimizations and clustering modules.

One of such designed optimization technique utilized in our work is stated to be Gradient predictive Vector Quantization and Modified weight based Euclidian distance calculation.

Here, these optimization technique GPVQ would affect the current design based on the weights generated from the distances calculated for the each dataset sample and the user speech sample to be observed. These weights are to be predictive in nature which would affect the design characteristics so that for every iteration we could analyze the weights observed and weights predicted so to reach better classification of the samples. The detailed explanations of each such designed block have been explained below.

A) Feature Extraction Using MFCC (Mel Frequency Cepstral Coefficients)

The Mel scale parametric values are obtained from the application of cosine change of log range from the scale utilized. The observed coefficients with current design Mel scale would be called as MFCC. Removing MFCCs from the speech signal which comprises of a couple of procedures as showed in.

B) Fast Fourier Transform (FFT)

The initial phase in the element extraction stage is to change over the speech signals in frequency domain modelling with the usage of the Fourier transforms. The confined and windowed signal from the pre-handling stage is exposed to a Discrete Fourier Transform (DFT) so as to change over every speech for each frame in time domain to frequency domain analysis.

The DFT of the pre-prepared discourse outline is characterized as:

$$X_w(k) = \sum_{n=0}^{N-1} x_w(n)\, e^{-\frac{j2\pi kn}{N}}, \quad k = 0,1,2 \ldots\ldots N-1 \qquad (1)$$

Where, x(n) pre-processed speech frame and is the length of the DFT. The Fast Fourier Transform (FFT) is utilized as it is a quick calculation for executing the Discrete Fourier Transform (DFT) to implement in domain changes observed in frequency.

C) Mel Frequency Warping

Once the signal is changed over into the frequency domain through the Fast Fourier Transform (FFT), the subsequent stage in the process is to change and modify the frequency based power spectrum range as indicated by the mel scale in order to change over it into the mel range. To achieve this, the power spectrum of the speech signal is weighted by a progression of triangular filters calculated based on the frequencies whose bandwidth capacities and focal frequencies coordinate those of the sound-related basic band channels. The filters designed which convert the signal spectrum into a portrayal similar to the conduct of the human ear, for example the mel scale. The human ear is significantly more delicate to bring down frequencies (<1kHz) than the higher frequencies. This is found in the mel scale and is additionally why the filters are mel scaled so as to demonstrate this conduct (a larger number of filters in the lower frequency domain than in the higher ones). The designed filters are called mel-scale filters and as collective nature said to be mel-scale filter bank. (4) Is utilized to change over the frequency range to the mel range where m speaks to mels and f depicts the frequency.

4.2 Gender Recognition

The listener does not only decode the linguistic message from the speech signal, but at the same time he or she also infers paralinguistic information such as the age, gender and other properties of the speaker. This type of information is termed as voice information. This gender detection has application in many fields like

- Sorting telephone calls by gender for gender sensitive surveys.

- Identifying the gender and removing the gender specific components gives higher compression rate and enhanced the bandwidth. In terms of acoustic properties, voice information has been described by several acoustic parameters. Two of these parameters are of perceptual relevance:

1. the fundamental frequency (f0) and
2. The spectral formant frequencies.

Male has both the frequencies lower than the female. But formant frequency is something related to the vowels and hence it is text dependent and as our project is text dependent therefore gender classification here is done by using pitch/fundamental frequency extracted from different methods. Pitch is a very important feature which can be obtained from different methods in both time domain as well as frequency domain and also by the combination of both time domain and frequency domain.

Time domain methods include all those methods in which we work directly on the speech samples. The speech waveform is directly analyzed and the methods like short time autocorrelation, modified autocorrelation through clipping technique, normalized cross correlation function, average magnitude difference function, square difference function etc.

Similarly, in frequency domain methods the frequency content of the signal is initially computed and then the information is extracted from the spectrum. These methods include harmonic product spectrum analysis, Harmonic to sub harmonic ratio etc. There are also some methods which do not come under either time domain or frequency domain like wavelets, LPC analysis etc.

5.PROPOSED MODEL FOR IMPROVISING AGSR

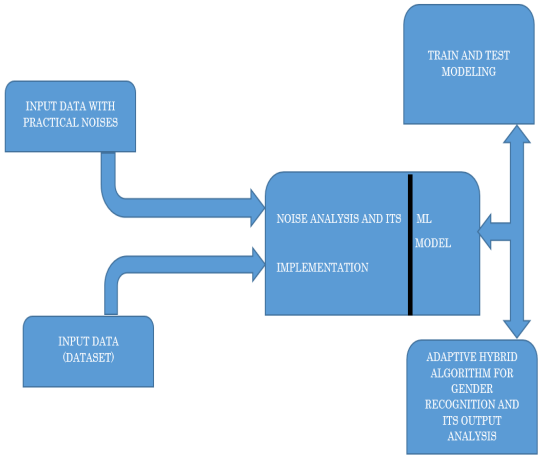

Figure 3: Representing the block diagram for proposed model

5.1 Block Diagram Description

The current design on figure-3 depicts with implementing of the different audio signals available based on two variations as mentioned (practical and Dataset) case which have to be processed with the current noise model and its implementation characteristics. Now these modelled data from the analysis mention in the next section of the design parametric criteria. The parametric criteria would suggest the modelling and implementation of the hybrid approach for noises and machine algorithms as combination of KNN and SVM algorithms.

For, noises we have considered iterative approach for LMS and NLMS which would provide better weights to ensure the noise model and its design analysis. The current model from the figure describes about input stream and its variations observed as these are self-noise built wave files and other are set of scenarios where the data modelling are separated with gender classify. Each model with efficient operation of the noise model with noise signal and its SNR values are less than 10dB are utilized for the classification of the gender recognition.

Our design proposes a novel scheme of the noise model and implementation of model for gender recognition for ensuring better accuracy. The design specification are improvised with the current noise models such LMS and NLMS, also for recognition we have provided with ensemble feature for KNN and SVM algorithm.

5.2 Hybrid Approach for Noise and machine learning design for Gender Recognition

a) Noise Model for Hybrid Algorithm Approach

Ensemble of LMS AND NLMS algorithm for design of noise model and its implementation would suggest different parametric factors affected on each audio signal utilized. To enhance particular design criteria on its performance, design procedure is being applied to the train and test model for the data samples utilized in the current proposed design.

b) Design Procedure for NOISE MODEL

i. Ensure the train and test data have noises less the 10dB.
ii. For greater SNR values each audio signals have to eliminate the different noises available in the current audio test and train data.
iii. The approach utilized on the optimal removal of the noises are the ensemble futuristic model for LMS and NLMS (or simply iterative model on LMS and NLMS)
iv. Iterative model and its implementation on the LMS and NLMS based on linear analysis.

5.3 LMS (Least Mean Squares) Algorithm

The LMS calculation is a surmised rendition of SDA, which substantially approximates R and p by supplanting the desire administrator by momentary worth. Thus,

$$R = E\,[x(n) * x^{t}(n)] \cong x(n) * x^{t}(n) \qquad (2)$$

$$p = E[d(n) * x(n) \approx d(n) * x(n)] \qquad (3)$$

Now we substitute equations (3) and (4) in stationary phase algorithm which provide a considerate equation as:

$$W_k(n+1) = W_k(n) + \mu[d(n) * x(n) - x(n)x^{T}(n) * W_k(n)] \qquad (4)$$

Common factors are meant to be separated and the equation has to rearrange as mention below:

$$W_k(n+1) = W_k(n) + \mu * x(n) * [d(n) - x^{T}(n) * W_k(n)] \qquad (5)$$

As we know that $y(n) = x^{T}(n) * W_k(n)$ and also we can equate $e(n) = d(n) - y(n)$, resulting the equation as:

$$W_k(n+1) - W_k(n) + \mu * x(n) * e(n) \qquad (6)$$

The above equation satisfies the adaptive model filter termed as least mean square algorithm (LMS).

Per the iterative and looping structure we can represent the LMS complexity as 2N+1 for multiplication per cycle and 2N for each cycle. The lower values of the parameter μ represents the better system performance and higher resulting noise bounded scenarios where $0 < \mu < 2/\lambda_{maximum}$, where λ maximum is largest eigen values of the correlation matrix.

5.4 NLMS (normalized least mean squares) algorithm

This algorithm provides an innovative approach to the existing adaptive filter algorithm (LMS) where an adaptive filter with coefficient or weight vector of order L, $W_k(n)=[w_1(n),w_2(n),\ldots,w_L(n)]^{T}$, is used to model an unknown system with impulse response $w^{*}=[w_1,w_2,\cdots,w_L]^{T}$. Here, $(\cdot)T$ denotes the transpose of a vector or a matrix. The unknown system and the adaptive filter are simultaneously excited by the same input ***x(n)***. The output of the unknown system ***d(n)*** is assumed to be corrupted by a measurement noise ***η(n)*** to form the desired signal ***d(n)*** for the adaptive filter. The NLMS algorithm under consideration assumes the following form:

$$W_k(n+1) = W_k(n) + \frac{\mu e(n)X(n)}{\varepsilon + \alpha X^{T}(n)X(n)} \qquad (7)$$

Where $x(n) - [x(n), x(n-1), \ldots x(n-L-1)]^{T}$ is the input vector at time instant n, μ is a positive step size constant to ensure convergence of the algorithm, and ε and α are positive constants.

5.5 Hybrid Weight Equation Analysis

From the above mention algorithms we have seen the equations suggesting the changes in weights for each model and its parametric criteria as modelled below:

$$W_k(n+1) = W_k(n) + \mu * x(n) * e(n) \text{ [From (6)}$$

$$W_k(n+1) - W_k(n) + \frac{\mu e(n)X(n)}{\varepsilon + \alpha X^{T}(n)X(n)} \text{ [From (7)}$$

From the weight values of the above equations we have utilized a linear iterative approach utilizing the weight which are implemented in LMS and NLMS.

Consider, a linear equation on single plane P(X, Y) as shown below:

$$f(x,y) = m1 * x + m2 * y \qquad (8)$$

Assuming f(x, y) would attain a constant or variable either x or y,

$$f(x,y) \approx cx \text{ or } cy \text{ or } c \qquad (9)$$

If $f(x,y) - c$ then,

From the equation (8) we have

$$y = \frac{m1}{m2} * x - \frac{1}{m2} * c \qquad (10)$$

Representing the one of the weight equations

Similarly $f(x,y) = cx \text{ or } cy$ would suggest the variations of x either in right side or left side depending on the values of c, m1 and m2. Finally, we propose a linear modeling for the current weights for each set of the algorithms as shown below:

$$W(n+1) = \lambda * f1 + \delta * f2 \qquad (11)$$

Here f1 and f2 are the weights from the equations 6 and 7. Also the values of $\lambda \text{ and } \delta$ are varied between $\lambda, \delta \subset (0,1)$.

5.6 Noise Model And Its Analysis

Considering the design of the Noises for the model as per the criteria of the filer proposed we have regarded as a mathematical model for the noises that have to be synchronize for each noise data considered.

We represent different noise data as a prime requirement for the design that satisfies the criteria of the Signal to Noise ratio that would range from 0, 5, 10 and 15 dB establishing the filter structure has to be utilize model chosen. As we know that filter structure used for design model should provide the model

equation to ensure the correct response of the design SNR is achieved based on the selection of the noises considered.

In our design we emphasize on the structural modeling and the computational equation generated based on the SNR which determines correct values achieved from variations on **μ, ϵ, δ** and **λ** for each algorithm utilized in the design. For all such values we need to estimate destined equation that would be generated from the formula:

$$\text{SNR (dB)} = 10\log(S/N) \quad (12)$$

$$S = N * 10^{0.1 \times SNR} \text{ or } N = S * 10^{-0.1 \times SNR} \quad (13)$$

So basically the equation can represented as:

$$p = q * 10^{k} \quad (14)$$

Here ***p*** and ***q*** are variables and ***k*** is constant factor which depends on the values of ***p*** and ***q***. In the current design criteria we know that a SNR range of 0-15 dB has to apply to value of the ***'k'*** where each values of the ***k*** is dependent on the noise which we have chosen. Now, the noises are: Babble, Factory, Destroy Engine, Car, Fire Engine and Train Noises. For such each noise model we generate the equations modeled as:

Initialize the Noise structure in SNR range. To implement this structure we propose the linear and quadratic equations based on the values of the δ, ϵ, μ and λ. Now as per the design we have utilized μ=0.01, $\epsilon=10^{-6}$, δ = 0.1 and $\lambda=10^{-2}$. Our design on Noises is comprised on the estimation of different noise scenario, on considering that Noise observed and noise calculated have their own set of noise generated equations.

The provisions of the parameters of the noise would vary based on the SNR values. The SNR is ranging from 0 15dB resulting different criteria of equations to be utilized for the correct SNR to on each noise was chosen. Now, to implement the equations we need to understand the direct dependent factor (SNR with 10 powers) based on the equation 1 (in noise model calculations). To apply an equation model, we need to realize its final equation for each SNR chosen and based on the frequency response of the each noise chosen for its coefficients of the input considered.

For SNR = 0 the obtained equation is:

$$s = Const - t \quad (15)$$

Here ***s*** and ***t*** are the output SNR and input SNR. Const is key parameter calculated as:

$$RMS\ of\ input\ [noise\ or\ data] - \sqrt{\sum_{i=1}^{N} s_i^2} \quad (16)$$

$$A_n = \frac{noise\ in\ RMS}{\delta} \quad (17)$$

$$A_{in} = input\ in\ RMS \quad (18)$$

For SNR =5:

$$P_n = \frac{2 * \delta * abs(noisx + A_n)}{2 * \delta} \quad (19)$$

For SNR = 10:

Equation is

$$P_n - \frac{\mu * noisex * \delta - rms\left(mean\left(noise_{in,p}\right)\right) - (A_{in} * A_n) - \mu * noisex * \delta * \mu^{noisex}}{\mu * noisex * \delta} \quad (20)$$

Here, noisex is SNR Equation for SNR=5.

and

For
SNR=15:

$$P_n = \frac{\mu * noisex * \delta - rms\left(mean\left(noise_{in,p}\right)\right) - (A_{in} * A_n) - \mu * noisex * \delta * \mu^{noisex}}{\mu * noisex * \delta} \quad (21)$$

Noise-x is SNR Equation for SNR=10.

Finally, the designed values for each SNR calculations have proven as 0, 5, 10 and 15 dB approximated for each noise chosen. The values for each SNR noises will abruptly change if the input is modified apart from the existing chosen one.

5.7 Machine Learning Algorithm Model

To improvise the design criteria and its learning model to provide the classification of the particular objects and its real time scenario for each set of the design model and its calculation based on the parametric criteria as mention in section III. Analysing these criteria and its relational aspects to the current problem of classification and recognition for each set of male voices and female voices. This modelling would results in the analysis of the different concepts utilized ensuring the mathematical formulation for each set of the algorithm applied. The current proposed model would suggest on each set of the weights observed on the formulation of the SVM and KNN. These weights are estimated with its class names and its functionality of optimizations applied. MATLAB modelling for KNN and SVM would represents on the usage of ensemble model of each optimisation parameters on each algorithm utilized.

This model for SVM and KNN are being utilized on the FITCKNN and FITCSVM functions utilized on the current test and train data model. These test and train data would represent the different wav data observed from the data base of the current design model which would emphasize the design parametric would be based on the current machine learning algorithms as mentioned below:

5.7.1 SVM modelling

The classification model and its implementation would suffice the technique as classifier that impart a hyper-plane or a functionality based equation $g(x) = w^T * x + b$ utilizing the separation of the classes using optimal margin. The margin separated with the hyperplane would represent a linear SVM characteristics. To analyse mathematically we would consider a set of point's $\in x_i$ and its class models w1 and w2 which are linearly separable, the distance between the hyperplane and set of the class points as $\frac{|g(x)|}{\|w\|}$. This classifier aims at to locate w, b equals to the value of g(x) as 1 for nearest values or data values and $\in$ class w1 for w2 class its values assigned to -1. Hence from the g(x) equation we could represent the two functionalities for each class's w1 and w2 which are belongs to x in $g(x) = w^T * x + b$. The optimization function is given by $J(w) = \frac{1}{2}\|w\|^2$, with constraint as $y_i(w_i^T * x + b) \geq 1, i = 1, 2, 3, 4, 5, 6 \ldots N$. hence using Lagrangian function for the above equation we have $L(w, b, \lambda) = 0.5 * w^T w - \sum_{i=1}^{N} y_i(w_i^T * x + b)$.

5.7.2 KNN model

The KNN model is a supervised Machine learning algorithm depicts classification and regression models on different real time problems observed. The properties of the learning algorithm which doesn't require special training set and is parametric features but only represent based on the distance calculation on each set of train and test model.

i. Initiate the train and test model
ii. K nearest data points on each data sets
iii. Calculate the distance on Train and Test model with Euclidean, Manhattan or Hamming distance.
iv. Sort and assign the classified data based on the distances measured.

5.7.3 Hybrid Classify Model With Cepstrum Analysis

Cepstrum analysis on the current design model have utilized the different MFCC coefficients and pitch calculation as pseudo code figure-4. The calculation of the MFCC and pitch analysis for the current design model are represented with two distinct scenarios of train and test models ensuring the correct comparisons observed. This model proposes a clean 10dB variation of the SNR from the audio samples for each test and train model ensuring the different features tabulated as

- Mean
- Short Time Energy
- Spectral Centroid
- Pitch
- MFCC

```
    [trainFemaleSounds(t1), Fs] = audioread(myFile1(t1,1).name);
    femaleFeature(t1,2) = pitch(trainFemaleSounds(t1),Fs); %pitch_period feature for female :
    femaleFeature(t1,1) = -1; %female target = -1
[femaleMFCC, deltaFemale, deltaDeltaFemale] = mfcc(trainFemaleSounds(t1),Fs);
    meanFemale = mean(femaleMFCC,1);
    temp1 = mean(meanFemale);
    femaleFeature(t1,3) = temp1; %14 MFCC coefficient for female
    shortTimeEnergyFemale = ShortTimeEnergy(trainFemaleSounds(t1), ns, ns-overlap);
  femaleFeature(t1,4) = mean(shortTimeEnergyFemale); %short time energy feature for female
  spectralCentroidFemale = SpectralCentroid(trainFemaleSounds(t1), ns, ns-overlap, Fs);
  femaleFeature(t1,5) = mean(spectralCentroidFemale); %spectral centroid feature for female
```

The above code represents the pseudo code for MFCC and pitch calculation on the each audio samples taken.

Figure 4: Representing the pseudo code for the MLGR

6.RESULTS AND DISCUSSIONS

6.1 SVM Classification And Accuracy Results

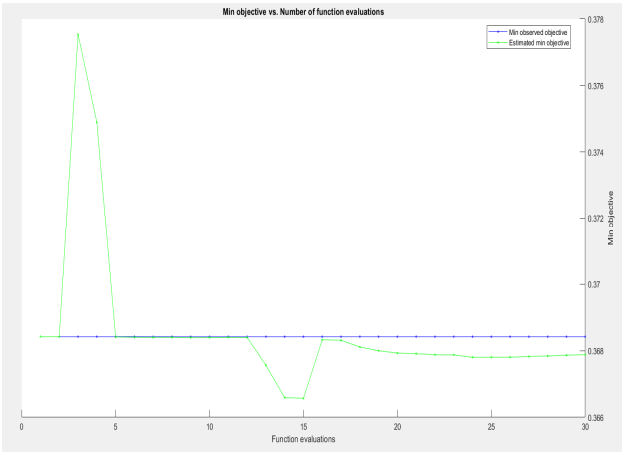

Figure 5: Representing the SVM objective graph on MALE and FEMALE characteristics for loss estimation

Iter	Eval result	Objective	Objective runtime	BestSoFar (observed)	BestSoFar (estim.)	BoxConstraint	KernelScale
1	Best	0.36842	0.11406	0.36842	0.36842	0.074763	0.2784
2	Accept	0.36842	0.15193	0.36842	0.36842	164.08	57.809
3	Accept	0.52632	0.11984	0.36842	0.37754	85.706	0.18858
4	Accept	0.36842	0.12916	0.36842	0.37489	0.040674	515.84
5	Accept	0.36842	0.085673	0.36842	0.36841	0.043382	0.32076
6	Accept	0.36842	0.096913	0.36842	0.3684	3.1938	200.57
7	Accept	0.36842	0.14331	0.36842	0.3684	0.0687	31.026
8	Accept	0.36842	0.12557	0.36842	0.3684	362.17	926.86
9	Accept	0.36842	0.097105	0.36842	0.3684	0.0010463	56.324
10	Accept	0.47368	0.10631	0.36842	0.3684	0.0052658	0.0048177
11	Accept	0.36842	0.13038	0.36842	0.3684	0.082887	2.0702
12	Accept	0.36842	0.16128	0.36842	0.3684	0.0010092	975.1
13	Accept	0.36842	0.13461	0.36842	0.36756	54.076	289.38
14	Accept	0.36842	0.10244	0.36842	0.36659	999.31	218.45
15	Accept	0.36842	0.099627	0.36842	0.36657	0.0010945	1.2284
16	Accept	0.36842	0.10203	0.36842	0.36833	0.011871	0.8174
17	Accept	0.36842	0.12255	0.36842	0.36831	0.0010056	6.8163
18	Accept	0.36842	0.1245	0.36842	0.36911	3.3549	35.771
19	Accept	0.36842	0.16679	0.36842	0.368	0.0054535	230.31
20	Accept	0.36842	0.097538	0.36842	0.36793	0.29968	108.59

Iter	Eval result	Objective	Objective runtime	BestSoFar (observed)	BestSoFar (estim.)	BoxConstraint	KernelScale
21	Accept	0.36842	0.20919	0.36842	0.36791	0.84677	967.77
22	Accept	0.36842	0.11589	0.36842	0.36788	40.929	89.574
23	Accept	0.36842	0.10608	0.36842	0.36788	0.0088027	987.64
24	Accept	0.36842	0.1564	0.36842	0.3678	257.99	189.02
25	Accept	0.36842	0.09012	0.36842	0.3678	17.346	980.45
26	Accept	0.36842	0.096994	0.36842	0.36781	0.062615	0.63411
27	Accept	0.36842	0.10378	0.36842	0.36783	0.29926	11.2
28	Accept	0.36842	0.095426	0.36842	0.36784	0.0010252	293.59
29	Accept	0.36842	0.091058	0.36842	0.36786	995.43	964.2
30	Accept	0.36842	0.10071	0.36842	0.36788	0.015462	7.1244

```
Optimization completed.
MaxObjectiveEvaluations of 30 reached.
Total function evaluations: 30
Total elapsed time: 29.725 seconds.
Total objective function evaluation time: 3.5773

Best observed feasible point:
    BoxConstraint    KernelScale
    _____________    ___________

        0.074763         0.2784

Observed objective function value = 0.36842
Estimated objective function value = 0.36788
Function evaluation time = 0.11406

Best estimated feasible point (according to models):
    BoxConstraint    KernelScale
    _____________    ___________

        0.043382         0.32076

Estimated objective function value = 0.36788
Estimated function evaluation time = 0.11646
```

Figure 6: Representing the SVM functionalities for MALE and FEMALE optimization with best feature observed

The current model is applied with different algorithms such as SVM, KNN and HOC Ensemble with ADABOOST. Figure 5 and 6 depicts the performance scenario on the current GR classification and recognition resulting the overall loss value on each iteration value as 0.367. Hence overall loss is 30*0.367= 11 %. The accuracy is obtained with 100-loss = 89%.

6.2KNN Classification And Accuracy Results

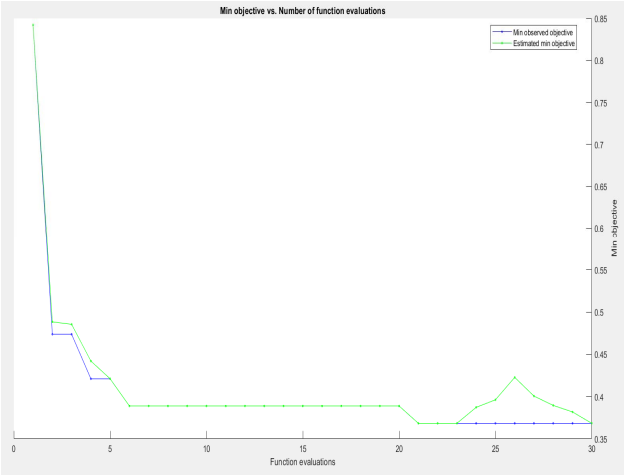

Figure 7: Representing the KNN objective graph on MALE and FEMALE characteristics for loss estimation

Iter	Eval result	Objective	Objective runtime	BestSoFar (observed)	BestSoFar (estim.)	NumNeighbors	Distance
1	Best	0.84211	1.594	0.84211	0.84211	1	jaccard
2	Best	0.47368	0.40328	0.47368	0.48833	5	minkowski
3	Accept	0.68421	0.09998	0.47368	0.48545	1	cityblock
4	Best	0.42105	0.12299	0.42105	0.44203	4	chebychev
5	Accept	0.42105	0.10248	0.42105	0.42106	4	chebychev
6	Best	0.38889	0.19963	0.38889	0.38891	4	correlation
7	Accept	0.44444	0.15237	0.38889	0.38891	4	cosine
8	Accept	0.42105	0.11644	0.38889	0.3889	4	cityblock
9	Accept	0.42105	0.07552	0.38889	0.3889	4	euclidean
10	Accept	0.63158	0.091156	0.38889	0.3889	4	hamming
11	Accept	0.47368	0.40068	0.38889	0.3889	5	mahalanobis
12	Accept	0.47368	0.59998	0.38889	0.3889	5	seuclidean
13	Accept	0.61111	0.24769	0.38889	0.3889	5	spearman
14	Accept	0.63158	0.10117	0.38889	0.3889	5	jaccard
15	Accept	0.52632	0.12636	0.38889	0.3889	2	euclidean
16	Accept	0.63158	0.092296	0.38889	0.3889	2	hamming
17	Accept	0.61111	0.10629	0.38889	0.38891	2	cosine
18	Accept	0.44444	0.15199	0.38889	0.3889	9	cosine
19	Accept	0.44444	0.12697	0.38889	0.3889	9	correlation
20	Accept	0.5	0.11632	0.38889	0.38891	2	correlation
21	Best	0.36842	0.083754	0.36842	0.36843	8	cityblock
22	Accept	0.47368	0.10139	0.36842	0.36843	9	euclidean
23	Accept	0.36842	0.078809	0.36842	0.36843	10	chebychev
24	Accept	0.47368	0.10267	0.36842	0.38729	7	chebychev
25	Accept	0.42105	0.16221	0.36842	0.39621	10	cityblock
26	Accept	0.47368	0.21553	0.36842	0.42272	7	cityblock
27	Accept	0.36842	0.079376	0.36842	0.4006	10	chebychev
28	Accept	0.36842	0.11078	0.36842	0.38956	10	chebychev
29	Accept	0.36842	0.07122	0.36842	0.38187	10	chebychev
30	Accept	0.68421	0.12567	0.36842	0.36842	1	chebychev

```
Optimization completed.
MaxObjectiveEvaluations of 30 reached.
Total function evaluations: 30
Total elapsed time: 44.4908 seconds.
Total objective function evaluation time: 6.159

Best observed feasible point:
    NumNeighbors    Distance
    ____________    _________

        8           cityblock

Observed objective function value = 0.36842
Estimated objective function value = 0.36842
Function evaluation time = 0.083754

Best estimated feasible point (according to models):
    NumNeighbors    Distance
    ____________    _________

        10          chebychev

Estimated objective function value = 0.36842
Estimated function evaluation time = 0.088007
```

Figure 8: Representing the KNN functionalities for MALE and FEMALE optimized distance with best feature observed

Similarly we have observed in the figures 7 and 8 with KNN model for different distance algorithms, which have attained the loss % as 11 equallance to SVM. Hence accuracy will be 89%.

6.3Hybrid Algorithm Designed (HOCA)

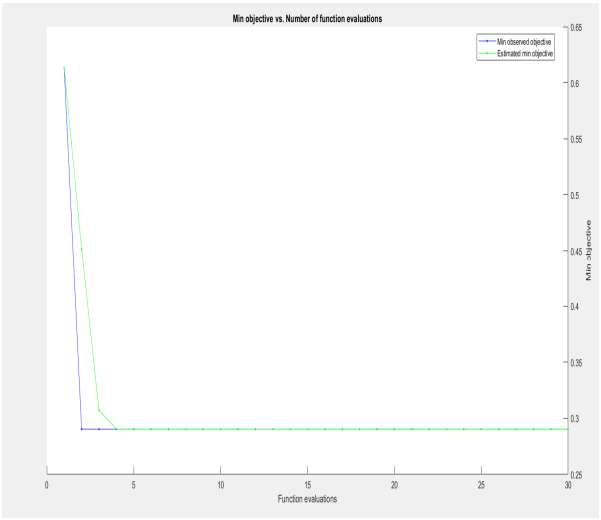

Figure 9: Representing the proposed model (HOCA) objective graph on MALE and FEMALE characteristics for loss estimation

Iter	Eval result	Objective	Objective runtime	BestSoFar (observed)	BestSoFar (estim.)	Method	NumLearningCycles	LearnRate	MinLeafSize
1	Best	0.29032	10.922	0.29032	0.29032	AdaBoostM2	148	0.89416	9
2	Accept	0.48387	2.0241	0.29032	0.30235	RUSBoost	19	0.001296	1
3	Accept	0.29032	5.4102	0.29032	0.29033	AdaBoostM2	74	0.0018357	8
4	Accept	0.51613	26.147	0.29032	0.29031	AdaBoostM2	427	0.001509	1
5	Accept	0.6129	22.171	0.29032	0.30765	AdaBoostM2	349	0.043335	15
6	Accept	0.45161	5.1464	0.29032	0.44086	AdaBoostM2	76	0.0018378	6
7	Accept	0.48387	25.211	0.29032	0.447	AdaBoostM2	396	0.0035495	5
8	Accept	0.29032	4.6896	0.29032	0.29033	AdaBoostM2	74	0.0017991	10
9	Accept	0.29032	16.19	0.29032	0.29029	AdaBoostM2	257	0.0010666	9
10	Accept	0.35484	1.0731	0.29032	0.29032	Bag	10	-	8
11	Accept	0.41935	4.9473	0.29032	0.29031	AdaBoostM2	81	0.99297	7
12	Accept	0.29032	0.66565	0.29032	0.29032	Bag	10	-	10
13	Accept	0.29032	2.5621	0.29032	0.29031	AdaBoostM2	39	0.0029222	9
14	Accept	0.29032	27.188	0.29032	0.29031	Bag	499	-	11
15	Accept	0.6129	0.73867	0.29032	0.29032	Bag	11	-	13
16	Accept	0.32258	0.78364	0.29032	0.29032	AdaBoostM2	11	0.25302	11
17	Accept	0.29032	26.607	0.29032	0.29032	Bag	484	-	9
18	Accept	0.51613	1.4893	0.29032	0.29032	Bag	23	-	5
19	Accept	0.41935	1.4474	0.29032	0.29032	RUSBoost	19	0.0014128	9
20	Accept	0.6129	8.5343	0.29032	0.29032	RUSBoost	131	0.94879	12

Iter	Eval result	Objective	Objective runtime	BestSoFar (observed)	BestSoFar (estim.)	Method	NumLearningCycles	LearnRate	MinLeafSize
21	Accept	0.45161	1.8779	0.29032	0.29032	RUSBoost	26	0.0010062	7
22	Accept	0.51613	1.2909	0.29032	0.29032	RUSBoost	16	0.46242	4
23	Accept	0.51613	24.287	0.29032	0.29032	Bag	446	-	2
24	Accept	0.45161	33.364	0.29032	0.29032	RUSBoost	496	0.010469	2
25	Accept	0.51613	0.82096	0.29032	0.29032	AdaBoostM2	10	0.56406	2
26	Accept	0.48387	0.6785	0.29032	0.29032	Bag	10	-	1
27	Accept	0.51613	33.235	0.29032	0.29032	AdaBoostM2	498	0.0013414	3
28	Accept	0.45161	1.3779	0.29032	0.29032	Bag	26	-	3
29	Accept	0.48387	36.196	0.29032	0.29032	RUSBoost	489	0.092358	3
30	Accept	0.32258	29.138	0.29032	0.29032	Bag	468	-	7

```
Optimization completed.
MaxObjectiveEvaluations of 30 reached.
Total function evaluations: 30
Total elapsed time: 317.0714 seconds.
Total objective function evaluation time: 255.2415

Best observed feasible point:
      Method      NumLearningCycles    LearnRate    MinLeafSize
    ___________    _________________    _________    ___________

    AdaBoostM2           29             0.0063498         9

Observed objective function value = 0.29032
Estimated objective function value = 0.29034
Function evaluation time = 1.5676

Best estimated feasible point (according to models):
      Method      NumLearningCycles    LearnRate    MinLeafSize
    ___________    _________________    _________    ___________

    AdaBoostM2           18             0.0097221         9

Estimated objective function value = 0.29034
Estimated function evaluation time = 1.2023
```

Figure 9: Representing the optimization algorithms and its parametric features for HOC model

Here, the algorithms are two different ML models with boosting algorithms utilized on the current data set mention in the directory path. For each set of the data appended on each test and train locations would only result in greater iterations and larger simulation time. Hence from the above figures 7 and 8 we have seen the optimal loss is 0.29 with 30 iterations resulting 8.7%, and with an accuracy of 91.3%.

Table 1: Representing the comparison results of the existing and proposed model for accuracy and losses.

SN0	Algorithms	Accuracy For 30 Iterations (%)	Losses For 30 Iterations (%)
1	SVM	89%	11 %
2	KNN with chebysev distance	89%	11 %
3	Ensemble with ADABOOST	90.4%	9.6%
4	Proposed Approach (HOCA)	91.3%	8.7%

The above Table-1 depicts the design model comparisons on each algorithms as mentioned and its accuracy and losses observed via applying to the Ensemble learning network.

7.CONCLUSIONS

Speech and its analysis have become an important aspects of the current real time modelling of the problems associated with audio processing. Ensuring the data set and its functional features for the audio data would improvise the analysis on the output observed with each set of the voice data as “male” and “female”. Our design provides better approach for the implementation of various data sets on voices such as male and females depending upon the data set collected. The occurrence of the data and its relational features such as would provide a classification model parameters on each test or train voice sample as:

- Mean
- Short Time Energy
- Spectral Centroid
- Pitch
- MFCC

From, the samples considered as at least 10 male and 10 female voice data, from the kaggle voice data set and real time recorded voice samples which imparts different voice pitches for each male and female criteria. Modelling with these samples would require training model and test model for to enact the different test samples considered. Previous approaches from MFCC learning and vector blast quantization have been proven to recognize the voice based on the different

samples and its distances located at each section. Applying the different boosting algorithms have shown the incremental change of the accuracy about 3% for only 10 samples of each male and female audio data.

SCOPE

To implement on the design model and its real time noise analysis with Projection algorithm. Ensuring the design algorithm with the best noise model would suggest best results and imparts different data sets with larger voice samples. Enhanced data model with better pitch signal capabilities would improvise the design parametric aspects. To implement CNN or DEEP NN models via NVIDIA GPU based on the larger samples would increase the performance drastically and ensuring the better recognition rate.

ACKNOWLEDGEMENT

We thankfully acknowledge management of KL University to provide each source and required facilities for completion of this work.

REFERENCES

1. Jisung Wang, Sangki Kim, Yeha Lee, **SpeechAugmentation Using Wavenet in Speech Recognition,** *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), **Electronic ISBN:** 978-1-4799-8131-1*
2. Samik Sadhu, Ruizhi Li, Hynek Hermansky, **M-vectors: Sub-band Based Energy Modulation Features for Multi-stream Automatic Speech Recognition,** *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), **Electronic ISBN:** 978-1-4799-8131-1*
https://doi.org/10.1109/ICASSP.2019.8682710
3. Andrew Rosenberg, Yu Zhang, Bhuvana Ramabhadran ,Ye Jia, Pedro Moreno, Yonghui Wu, Zelin Wu, **Speech Recognition with Augmented Synthesized Speech**, *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), **Electronic ISBN:** 978-1-7281-0306-8.*
4. Hainan Xu, Shuoyang Ding, Shinji Watanabe, **Improving End-to-end Speech Recognition with Pronunciation-assisted Sub-word Modelling,** *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), **Electronic ISBN:** 978-1-4799-8131-1*
5. Burnett, M., & Kulesza, T, **End-User Development in Internet of Things: We the People**. (2015) *In International Reports on Socio-Informatics (IRSI)*, Vol.12, Iss.2, pp. 81-86.
6. Childers, D. G., Wu, K., Bae, K. S., & Hicks, D. M. (1988, April). **Automatic recognition of gender by voice**. *In Acoustics, Speech, and Signal Processing, 1988. IEEE-ICASSP-88., 1988 International Conference on* (pp. 603-606).
7. Sigmund, M. **Gender distinction using short segments of speech signal**. (2008). *International Journal of Computer Science and Network Security*, pp-8(10), 159-162.
8. Tolba, H, **A high-performance text-independent speaker identification of Arabic speakers using a CHMM-based approach**. (2011) *Alexandria Engineering Journal*, pp-50(1), 43-47.
https://doi.org/10.1016/j.aej.2011.01.007
9. Chisaki, Y., Nakashima, H., Shiroshita, S., Usagawa, T., & Ebata, M. **A pitch detection method based on continuous wavelet transform for harmonic signal**. (2003). *Acoustical science and technology*, pp-24(1), 7-16.
10. Sigmund, M. **Gender distinction using short segments of speech signal**. (2008). *International Journal of Computer Science and Network Security*, pp-8(10), 59-162.
11. Bissell, C. C. **Nyquist Rate Sampling**. (1990). *International Journal of Electrical Engineering Education*, pp-27(1), 77-79.
12. Ali, M. S., Islam, M. S., & Hossain, M. A. **Gender recognition system using speech signal**. (2012). *International Journal of Computer Science, Engineering and Information Technology (IJCSEIT)*, pp-2(1), 1-9.
https://doi.org/10.5121/ijcseit.2012.2101
13. Hasan Erokyar, **Age and Gender Recognition for Speech Applications based on Support Vector Machines,** MS Thesis, Dept. Elect. Eng., South Florida Univ.2014.