International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-5, March 2020

Speech Classification for Kannada Language

Supriya B Rao, Sarika Hegde

Abstract: ***Speech classification is one of the challenging issues in speech processing. In this paper, we have done speech classification for the Kannada language. We have gathered a speech database from children aged 4-6 years. The dataset collected are pre-processed and speech feature extraction is done using Mel Frequency Cepstral Coefficients (MFCC) technique. After feature extraction Kannada alphabets are classified using six different Machine Learning (ML) classifiers. The classifier accuracies are compared with each other. Amongst the Deep Learning classifiers, Recursive Neural Network (RNN) gave the highest accuracy of around 93.6 %(for 300 epochs) and Random Forest (RF) gave the highest accuracy of around 88.9% which is a Machine Learning classifier.***

Keywords: Speech Classification, Kannada, Machine Learning

I. INTRODUCTION

Many systems have been developed which helps in analyzing, classifying and recognizing the speech signals. System developed comprises of both hardware as well as software. Such kind of systems is widely used in various fields such as agriculture, education, health sectors etc. Our objective is to develop a speech recognition system for Kannada Language. The Kannada speech recognition system should recognize the speech providing maximum accuracy. Kannada is a language spoken by the people of Karnataka State in South India. The Kannada alphabets include 13 vowel letters, 2 part-vowel, part -consonant letters and 25 structured consonants and 9 unstructured consonants [6].For our research work we have used Kannada vowel letters and part-vowel, part-consonant letters. Table 1 shows the list of vowels and also the Indian Language Transliterations (ITRANS) for the vowels.

Table- I: Kannada Vowels [7]

Vowels	ಅ	ಆ	ಇ	ಈ	ಉ	ಊ	ಋ	ಎ	ಏ	ಐ	ಒ	ಓ	ಔ
ITRANS	a	aa	i	I	u	U	Ru	e	E	i	o	O	ou

Table- II: Kannada part-vowel, part consonant [7]

Vowels	ಅಂ	ಅಃ
ITRANS	aM	aH

Revised Manuscript Received on March 30, 2020.
* Correspondence Author
Supriya B Rao*, Department of Computer Science and Engineering, NMAM Institute of Technology, Nitte, India. supriyabrao@gmail.com.
Sarika Hegde, Department of Computer Science and Engineering, NMAM Institute of Technology, Nitte, India.sarika.hegde@nitte.edu.in.

Signal processing, feature extraction and classification are the three main steps involved in developing a Kannada speech recognition system. Before feature extraction process the speech signals must undergo certain pre-processing steps like removal of background noise. Feature extraction is used for parametric representation of speech wave. It is the front end of signal processing. In this step the speech with linguistic content are kept and background noise will be discarded. Feature vector with multiple dimensions is obtained for each speech signal. For feature extraction various algorithms can be used like MFCC, Linear Prediction Coding (LPC) Perceptual Linear Prediction (PLP), etc. [1]. We have used MFCC algorithm for feature extraction. The next step is to classify the Kannada alphabets that we have taken. The classifiers are used to identify for which set of classes the alphabets are likely to fit. The dataset used in our work has total 15 classes .15 classes represents total of 15 Kannada alphabets staring from ಅ(a) to ಅಃ(aH). Machine learning classifiers like Support Vector Machine (SVM), Naïve Bayes (NB), and Random Forest (RF) and deep learning classifiers like Multilayer Perception (MLP), Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN) are used.

II. BACKGROUND RESEARCH

Diogo[2] The robust model makes use of SVM and double cross validation. The main advantage of this system is that it provides the children with the feedback based on their performance. The classification models are used to decide if the exercise is correct or not. The classification models like sustained vowel, increasing pitch, decreasing pitch are learned using the SVM. The feature extraction is done using OPENSMILE tool. The learned model provides high accuracy. The false negative rate is very low. Aishwarya[3] proposed a computer/mobile based technology which can be used in detecting, evaluating and also providing feedback to aphasia patients. Machine Learning model is used for phoneme recognition and also for decoding the words. Feature extraction is carried out to obtain phonemes of desired frame length. Sequence of phonemes is obtained by using HMM with trained language model. Further CNN classifier is used for classification of desired words. Wang [4] in this work the classification of normal and disordered speech is carried out. MFCC feature extraction with classifiers such as GM-SVM is used. Experimental results show that GMM-SVM classifiers classify pathological and normal speech more accurately than GMM classifier. In the environmental sound classification [8] ECS-10 dataset is used which has total of 10 classes. Environmental sounds like Dogs barking, Sea waves, Kids playing, Fire cracking etc are used. MFCC feature extraction is done. Total 6 classifiers are used and accuracies are compared. The accuracies obtained are as follows: SVM 81.7%, RF 80%, NB 69.7%, CNN

71.25% (100 epochs), MLP 63.125% (100 epochs), and RNN 66% (100 epochs).English fluency level [9] is classified using 5 different machine learning classifiers. There are three fluency levels low, medium and high. MFCC feature extraction is done. Mel coefficients are tuned. Accuracies are checked for different values of Mel coefficients. Total 5 classifiers are used out of which SVM achieved the highest accuracy (94.39%) and rest of the classifiers achieved accuracies above 89%.Urban sound classification [5] is done using Urban Sound dataset that has 10 classes which contains air conditioner, children playing, drilling etc sounds. After MFCC feature extraction classification accuracies are obtained for various classifiers like NB, J48 Decision Tree, SVM and RF. Among 4 classifiers Random Forest achieved highest accuracy around 94.17%.

III. SYSTEM DESIGN

Speech signals are processed and MFCC features are extracted. Using various classification algorithms the vowels will be classified into respective classes. The figure 1 shows the Kannada speech classification system.

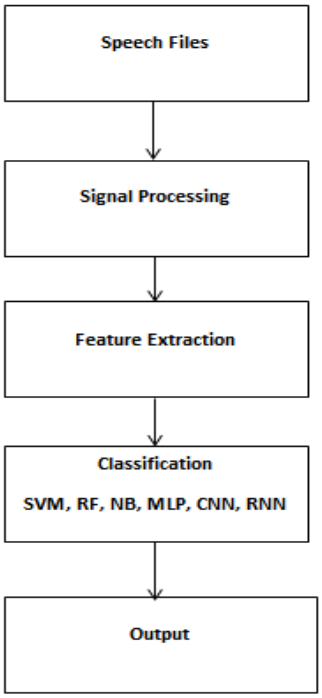

Fig.1. Kannada speech classification system

IV. RESULTS & DISCUSSIONS

A. Database

450 speech samples that are 1-3 seconds long were collected from 15 children of 4-6 years age. Each speech sample is a Pronunciation of Kannada alphabet. There are totally 15 classes for 15 Kannada alphabets. Out of 450 samples 294 samples are used for training and 126 samples are used for testing

B. MFCC Speech Features

MFCC feature extraction is used for parametric representation of speech signals. The first step in MFCC feature extraction is applying Pre-Emphasis filter. It is used for amplification of high frequencies in the signal. It is also helpful in avoiding the numerical problems that arise in the FFT process. It helps in improving the signal to noise ratio. Framing is a technique where the signals are divided into short frames. It is done to avoid the signal that keeps on changing over time the framing technique is applied. The signal remains constant over a short period of time. FFT is applied over these short frames. Adjacent frames are concatenated & frequency contours of signal are obtained that has good approximation. After dividing the signals into frames, Hamming window is applied to each frame. On each frame N-point FFT is performed. The frequency spectrum is calculated. Further the power spectrum is calculated. Filter bank is computed by applying the triangular filters. The coefficients calculated in the previous step can be highly correlated to each other. Due to this problem might occur in ML algorithm. To reduce the correlation between filter banks Discrete Cosine Transform (DCT) is applied [10].The Figure 2 shows the steps involved in MFCC feature extraction.

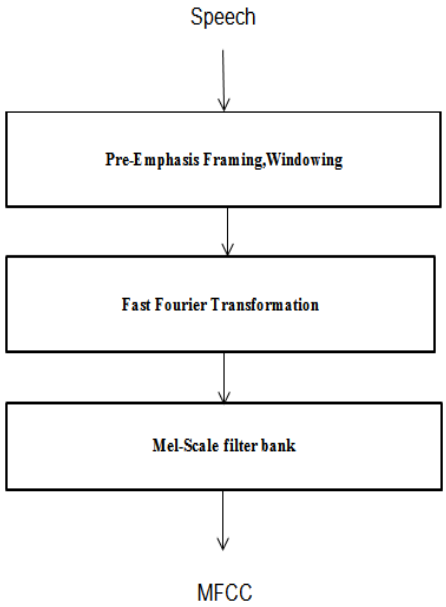

Fig.2. Block diagram of MFCC processor

C. Classifiers

Various classifiers namely SVM, MLP, CNN, RNN, NB, RF are implemented. Our main goal is calculate the test accuracies**.**

D. SVM

SVM is the most popular supervised machine learning algorithm. SVM makes use of kernel trick to handle nonlinear input. Different classes can be separated by SVM by constructing hyper plane in multidimensional space. Created optimal plane minimizes the error. Support Vectors are used in constructing the classifier [11]. We have used Radial basis function kernel for mapping input space in infinite dimensions space.

E. RF

Random samples are collected from the dataset and decision tree is constructed from each sample. Prediction result is obtained from each decision tree. Vote is given for each predicted result.

The prediction result with most votes is considered for final prediction [11].The confusion matrix for RF classifier is shown in the figure [3].The confusion matrix helps us to understand the status of training data. Each row and each column represents the particular Kannada alphabet. By analysing the confusion matrix we can understand how many times the classification is done right [3].

	ಅ	ಆ	ಇ	ಈ	ಉ	ಊ	ಋ	ಎ	ಏ	ಐ	ಒ	ಓ	ಔ	ಅಂ	ಅಃ
ಅ	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ಆ	0	8	2	0	0	0	0	0	0	0	0	0	0	0	0
ಇ	0	0	13	0	0	0	0	0	0	0	0	0	0	0	0
ಈ	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0
ಉ	0	0	0	1	6	0	0	0	0	0	0	0	0	0	0
ಊ	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0
ಋ	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0
ಎ	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0
ಏ	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0
ಐ	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0
ಒ	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0
ಓ	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0
ಔ	0	0	0	0	0	3	0	0	0	0	0	0	4	0	0
ಅಂ	0	0	0	0	0	0	0	0	0	0	0	0	0	5	6
ಅಃ	0	0	0	0	0	0	0	0	0	0	0	0	0	1	9

Fig. 3.Confusion matrix for RF Classifier

F. NB

NB is a statistical classification technique which uses Bayes Theorem. The first step of NB classifier is to understand the problem & identify the important features & labels. The NB classifier calculates the prior probability for class labels. Similar probability with each probability for each class is found. The values are put into Bays formula & posterior probability is calculated. The input provided will belong to the class with higher probability.

G. MLP, CNN, RNN

We have used the deep learning models like MLP, CNN & RNN. The MLP model uses two hidden layer which has 512*512 neurons. The output layer has 15 neurons. Neurons present in each hidden layer uses relu function. The output layer makes use of softmax function which converts the output to class probability. The softmax function is used for multiclass logistic regression. In CNN model there are totally four hidden layers. The first two hidden layers has 64 convolution filters. The next two hidden layer has 128 convolution filters. The output layer has 15 neurons in total corresponding to 15 classes. Hidden layer makes use of relu activation function. Output layer makes use of softmax function. RNN has two hidden layers which has 256*32 neurons. Output layer has 15 neurons. RNN is a LSTM (Long short-term memory).

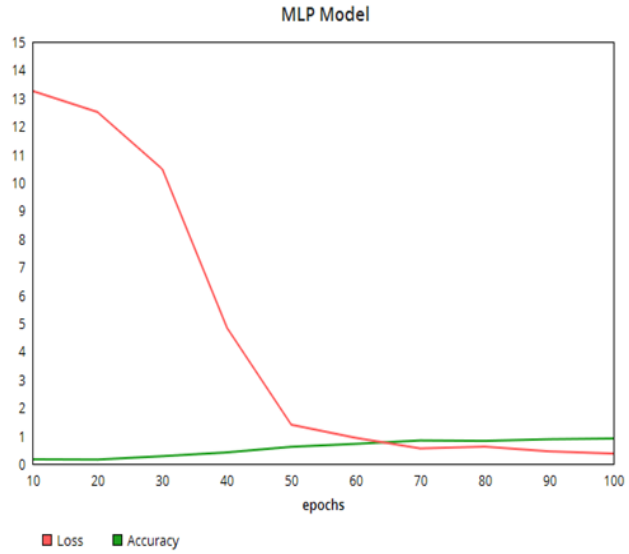

Fig.3. loss and accuracy of MLP model

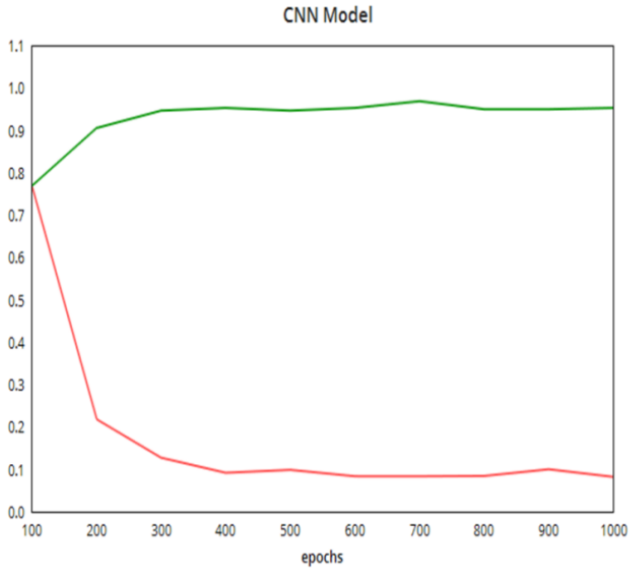

Fig.4. loss and accuracy of CNN model

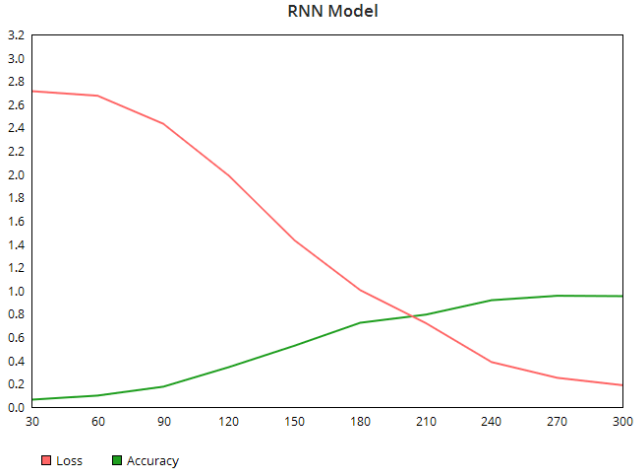

Fig.5. Loss and accuracy of RNN model

The loss vs accuracy graphs are plotted for CNN, RNN and MLP models as shown in the figure [3, 4, and 5]. In the above graphs the loss curve keeps on decreasing and stabilizes around some point. Accuracy curve keeps on increasing and stabilizes at some point. This indicates that the model is a good fit. For CNN we have plotted accuracy and loss curves for 1000 epochs, for MLP 100 epochs and RNN 300 epochs. The final accuracies are as follows: RNN 93.6% (300 epochs), MLP 89.6% (100 epochs), RF 88.9%, SVM 87.41%, CNN 85.18% (1000 epochs) and NB 56%.The below figure [7] shows the accuracy comparison between various classifiers.

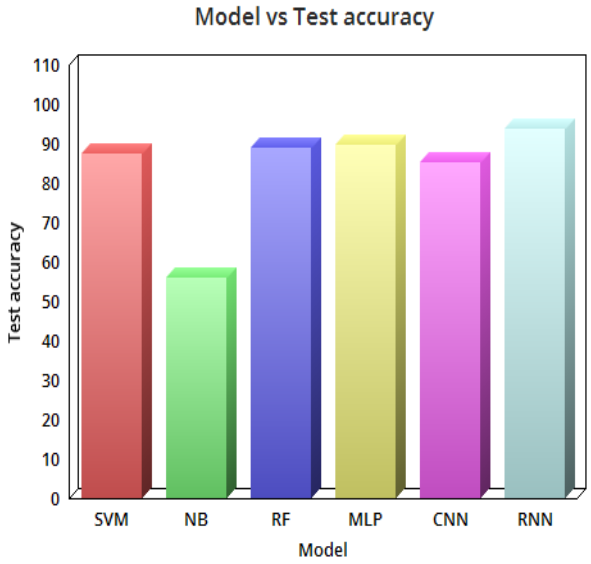

Fig.6. Model vs Test accuracy

V. CONCLUSION

In this paper MFCC features are extracted and Kannada alphabets are classified. We have used six different ML models to classify Kannada words. As a first step, we have determined the appropriate number of Mel cepstral coefficients for our data set. RNN which is a deep learning classifier provided highest accuracy about 93.6%.Amongst the ML classifiers RF gave us the highest accuracy around 88.9%.In the case of the neural networks architecture we can still explore adding specific hidden layers and modifying the number of neurons per layer. Further work includes classification of consonants and Kannada words and also we can aim at improving the accuracies of the classifiers.

ACKNOWLEDGMENT

We would like to acknowledge and express our gratitude to the Swasthi Shree Nemisagara Varneeji School(SNV), Karkala, who helped us in recording speech samples from children of pre-primary class. We have used these speech samples for our research work.

REFERENCES

1. Gupta, Divya, Poonam Bansal, and Kavita Choudhary. "The state of the art of feature extraction techniques in speech recognition." In *Speech and language processing for human-machine communications*, pp. 195-207. Springer, Singapore, 2018.
2. Diogo, Mariana, et al. "Robust scoring of voice exercises in computer-based speech therapy systems." Signal Processing Conference (EUSIPCO), 2016 24th European. IEEE, 2016.
3. Aishwarya, Jaya, et al. "Kannada Speech Recognition System for Aphasic people." 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE, 2018.
4. Wang, X., Zhang, J., & Yan, Y. (2009, October). Automatic Detection of Pathological Voices Using GMM-SVM Method. In 2009 2nd International Conference on Biomedical Engineering and Informatics (pp. 1-4). IEEE.
5. Afshankaleem, I. Shanti Prabha,"Ënhancement of Urban Sound Classification Using Various Feature extraction Techniques",IJRTE, Volume-7,Febrauary 2019.
6. https://en.wikipedia.org/wiki/Kannada
7. https://shodhganga.inflibnet.ac.in/bitstream/10603/104462/12/12_chapter%202.pdf
8. https://github.com/agrija9/Environmental-Sound-Classification-ESC-using-neural-networks-and-other-classifiers.
9. https://github.com/agrija9/Avalinguo-Dataset-Speaker-Fluency-Level-Clasification-Paper-/tree/master/code.
10. https://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html.
11. https://www.datacamp.com/community/tutorials/svm-classification-scikit-learn-python.

AUTHORS PROFILE

Supriya B Rao, BE in Computer Science and Engineering. Pursuing M. Tech in Computer Science and Engineering at NMAMIT Nitte, Karkala. Area of interest is Machine Learning, Speech Recognition. Research work carried out on Automatic Speech Therapy System for Kannada Language.

Sarika Hegde, Working as Associate Professor in the Department of Computer Science and Engineering at NMAM Institute of Technology, Nitte, India. Her research interests include Speech Recognition, Voice/Speech Disorder Detection, Statistics and Machine Learning, Data Visualization.