

## Dialect Identification in Telugu Language Speech Utterance Using Modified Features with Deep Neural Network



Shivaprasad Satla<sup>1,2\*</sup>, Sadanandam Manchala<sup>1</sup>

<sup>1</sup> Department of CSE, Kakatiya University, Warangal 506009, India

<sup>2</sup> School of CS & AI, SR University, Warangal 506009, India

Corresponding Author Email: [s.shivaprasad@sru.edu.in](mailto:s.shivaprasad@sru.edu.in)

<https://doi.org/10.18280/ts.380623>

### ABSTRACT

**Received:** 19 April 2020

**Accepted:** 20 November 2021

#### Keywords:

*DNN, Telugu language, dialects, multilayer perceptron, HMM, GMM, MFCC*

Dialect Identification is the process of identifies the dialects of particular standard language. The Telugu Language is one of the historical and important languages. Like any other language Telugu also contains mainly three dialects Telangana, Costa Andhra and Rayalaseema. The research work in dialect identification is very less compare to Language identification because of dearth of database. In any dialects identification system, the database and feature engineering play vital roles because of most of the words are similar in pronunciation and also most of the researchers apply statistical approaches like Hidden Markov Model (HMM), Gaussian Mixture Model (GMM), etc. to work on speech processing applications. But in today's world, neural networks play a vital role in all application domains and produce good results. One of the types of the neural networks is Deep Neural Networks (DNN) and it is used to achieve the state of the art performance in several fields such as speech recognition, speaker identification. In this, the Deep Neural Network (DNN) based model Multilayer Perceptron is used to identify the regional dialects of the Telugu Language using enhanced Mel Frequency Cepstral Coefficients (MFCC) features. To do this, created a database of the Telugu dialects with the duration of 5h and 45m collected from different speakers in different environments. The results produced by DNN model compared with HMM and GMM model and it is observed that the DNN model provides good performance.

## 1. INTRODUCTION

The Telugu Language is one of the historical languages of India. It is a popular language which consists of Sanskrit elegance, Tamil sweetness and Kannada essence. Being one of the early languages in India, the Telugu language has been recognized "Ancient Language" in 2008 by the government of India. The majority Telugu speaking people belong to Telangana, Andhra Pradesh followed by Venkatalakshamma and Munirathamma [1]. According to the Census 2018, the Language stood in 15<sup>th</sup> position in the world with a population of around 85 million speaking it, 3<sup>rd</sup> language in highest number of native speakers in India, with 6.93% and also it is the most popular spoken language of the Dravidian Language family. It contains total 52 core letters are present in the Telugu language. The vowels are known as "Achchulu" and consonants are known as "Hallulu". In Telugu, vowels add short /o/ and /e/ along with /o:/ and /e:/ of Indo-Aryan languages [2].

The Automatic Speech Recognition (ASR) is a branch of Artificial Intelligence (AI) and pattern recognition where humans interact with device's interface through their voices as they do with other human beings [3]. It contains lot of applications like Alexa, Siri, Google assistant etc. [4, 5]. These applications make the digital world to the next level. The ASR performance is considerably less in dialect identification compared to Language Identification because of lack of standard databases and also similar pronunciations of words in

different dialects of a language. Dialect is nothing but way of uttering a standard language spoken in particular geographical region.

In ASR, the feature engineering i.e., feature extraction and feature selection is also very important to identify the dialects or a Language. These are the important characteristics of speech signals. There are different types features like Spectral features, Prosodic features, phonotactic, lexical etc. The spectral features like MFCC, SDC etc. [6] are extracted, if the signal represents in spectral envelope. These features are not providing good results, if the different words are pronounced in similar way. In this situation, Prosodic features provide the better results. The prosodic features are unique with respect to dialects of a Language like Stress, Pitch, Energy, intonation etc. [7]. Developing an accurate method for the detection of dialect will helps us to develop some sort of e-services like e-health, e-market, e-education, telephonic services etc. which helps for older and homebound people also.

It can even provide a better education service by identifying the dialect because if teaching happens in same dialect, then it is very easy to understand [8, 9]. Dialect Identification improving the performance of ASR, which is available in every electronic gadget. The usage of hand held devices is increased rapidly in a wide range of different applications and grown in popularity in which speech is a one input. This motivated us to design an automatic Telugu dialect identification system using shortest utterance of speech. Due to the popularity and usage of speech based systems, the

research area of dialect identification has attracted more researchers of speech processing to design a system for identifying regional dialects in Telugu. The Telugu language consists, three dialects namely Telangana, Costa Andhra, and Rayalaseema [10] but the main drawback of do the research in Dialect identification is dearth of database. For identify the Telugu dialects, created the database by collecting speeches from different speakers.

From, the number of decades all researchers are applying the different machine learning techniques like Vector Quantization (VQ), SVM, K-means, HMM and GMM models [11, 12] to speech processing applications but DNN model make the Machine learning techniques to a step ahead. Now a days almost all the applications the neural networks are used. AI methods and applications are mostly solved by Neural Networks which is contains various hidden layers and output layer.

### 1.1 Telugu dialect database creation

A huge Telugu Corpus has been created by collecting the speech samples of different speakers using various recorders which are suitable for different environments and situations. In this case, speech samples were collected in online and offline and edited by using devices called PRAAT and Streaming audio recorder. The Chosen speakers ages are between 22 to 55 and different places are like working environments of office, schools, colleges and public place like parks, roadside, vendors etc.

The speakers who uttered speeches have been given freedom to speak on their own topic like own interests, habits, politics, self-description about family or home town etc.

The speech samples are recorded from different speakers including literates, illiterates and employees who are working in different occupations in different workplaces. The speech samples of different speaker were recorded in mono sound with the frequency of 44,100Hz. These speech samples are pre-processing using average filter to get rid of noise from the speech signal and also English words if any, in the speech the utterances are removed [13]. The speech wave records are 3-8 seconds for Telangana dialect, for Andhra dialect it is 3-7 seconds and for Rayalaseema it is 3-8 seconds. Dataset contains three datasets each having a place with the Telangana, Rayalaseema and Costa Andhra. Figure 1 represents the basic steps to create the Telugu dialects database from different speakers speech utterances.

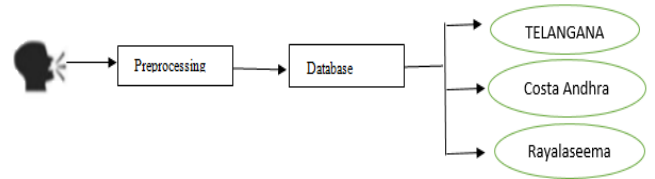


Figure 1. Block diagram of database creation

Total Five hours Forty-five minutes duration of speech Corpus is created from three dialects of Telangana, Costa Andhra and Rayalaseema speakers out of which 2 hour and 10 minutes was for Telangana dialect, 1h 57 minutes was for Andhra slang and 1 hour 38 minutes was for Rayalaseema slang. In Table 1 provides total Dataset created for identification of dialects.

Table 1. Complete dataset created

Dialect	Total time of speech data	Speakers for each dialect	Period of each sample	Age of speakers	Sampling Frequency
Telangana	2h 10 min	22	3-8	20-55	44,100Hz
Costa Andhra	1h 57 min	19	3-7	20-55	44,100Hz
Rayalaseema	1h 38 min	18	3-8	20-55	44,100Hz

## 2. FEATURE EXTRACTION

The feature extraction from input data is the first phase of any pattern recognition system. This step is part of training and testing phases of the system and these are similar. The feature vectors play a vital role in the accuracy and performance of the system. The selection of feature category and feature vectors in particular category influence the performance of system. In this proposed system, Mel Frequency Cepstral Coefficients (MFCC) and its derivative features are used to implement dialect identification system to identify Telugu dialects.

### 2.1 Mel Frequency Cepstral Coefficients (MFCC)

MFCCs are the most popular features which describe speech signal that is taken as an input from the speakers in ASR systems. It is useful for getting the spectral features from human voice. These features are derived based upon the frequency domain by using the Mel scale which is dependent upon human ear scale. Thus it will mimic the human ear. The main phases involved in MFCC feature extraction is shown in the Figure 2.

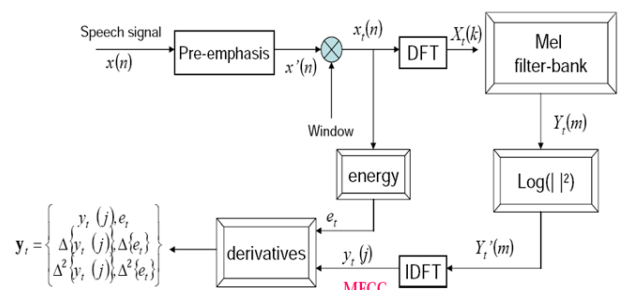


Figure 2. Basic structure of MFCC

The extraction of MFCC feature vectors from short duration of speech involves Pre-emphasis of speech signal, Framing and windowing of speech signal, Fast Fourier Transformation (FFT), Mel spectrum calculation and applying Discrete Cosine Transformation (DCT) [14].

Pre-Emphasis is used to maintain the higher frequencies in speech signal. When the speech signal is recorded by micro phone, it deviates with original spectrum of vocal tract by -6 dB/octave. To reduce the glottal effects of vocal tract, pre-

emphasis is used after pre emphasis applied the Framing and Windowing. As speech signal is quasi stationary signal in nature, short period of speech signal is considered to extract the features of important characteristics of human speech signal. Therefore, the analysis of speech signal always is performed on short segments in which assumed that the characteristics of speech signal are stationary.

In this, the size of the window is 20ms and advanced every 10ms. The 20ms analysis window was used to get good spectral features and 10ms was used to track the temporal characteristics of speech utterance. The speech signals initially in time domain format. To convert the signal from time domain to the frequency domain, applied the FFT. After applying the FFT, the resultant signal was passed as input to the band-pass filters for compute the MEL Spectrum. The band-pass filters also known as Mel-filter bank. The energy levels in subsequent bands are correlated because of the smoothness of the vocal tract. The DCT is applied to convert the Mel frequency coefficients to cepstral coefficients. Because of the smoothness of vocal track, the adjacent bands in energy levels are correlated. The DCT is applied to get the uncorrelated coefficients by transform the Mel frequency coefficients.

### 2.2 Delta MFCC ( $\Delta$ MFCC)

Derivation of MFCC feature vectors gives Delta MFCC. These Delta MFCC features help to represent the related Delta features to the change in cepstral features with respect to time. They also represent the change between the frames. They give the temporal information in the speech signal for each frame. As MFCC feature vector size is 13, Delta MFCC size also 13.

### 2.3 Delta Delta MFCC ( $\Delta\Delta$ MFCC)

Derivation of Delta MFCC feature vectors gives Delta Delta MFCC feature vector ( $\Delta\Delta$  MFCC). They represent the change in the delta features between the frames. They introduce even longer temporal context. They let us know if there is a peak or valley on the look over part of trajectory. As Delta MFCC feature vector size is 13, Delta Delta MFCC size also 13.

The delta MFCC is defined as:

$$\Delta_k = f_k - f_{k-1} \tag{1}$$

where,  $f_k$  represents a feature and  $k$  represents time instance and the delta-delta MFCC features calculated as:

$$\Delta\Delta_k = \Delta_k - \Delta_{k-1} \tag{2}$$

These features cause increasing the performance and efficiency in extraction of data from speech utterances.

## 3. PROPOSED METHODOLOGY

Like any speech processing model, the general model of dialect identification system consists of two phases: training phase and testing phase as specified in Figure 3.

As in the state-of-the-art systems, there are popular methods to extract the feature vectors from the raw speech signal by using framing and windowed speech like Mel Frequency Cepstral Coefficients (MFCC), Prosodic, TEO, SDC features, etc. [15, 16]. These feature extract methods must be similar in training and testing phases.

In the first phase of general model, the suitable feature vectors are extracted from speech utterances of human beings. These extracted features are given as input to training model to create the reference model. In testing phase, the feature vectors of test speech utterances are given to the input for trained models of dialects and evaluated to get likelihood score. The model with maximum likelihood score gives the dialect of unknown utterance of speech. The speech Corpus of each dialect is divided into two categories of samples which are used for training and testing randomly.

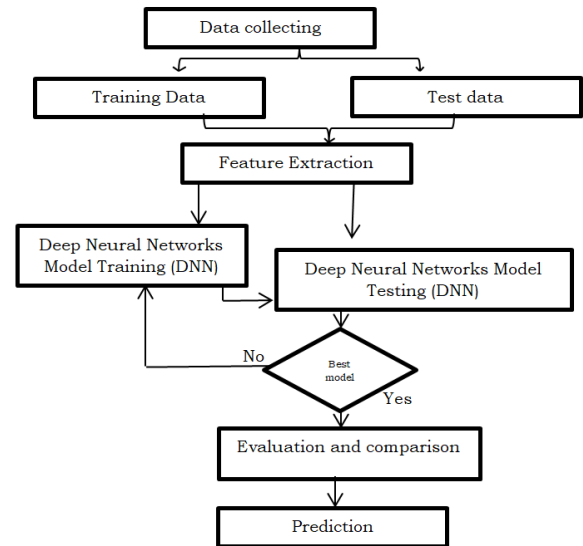


Figure 3. General model for dialect identification

### 3.1 Training deep neural networks

In training phase of DNN, dialect specific DNN is designed i.e., one DNN for one specific dialect using Stochastic gradient descent (SGD) and activation functions for Training speech of three dialects of Telugu language. The model designed for each dialect as shown in the Figure 4. In this work, MFCC feature vectors are extracted from Telugu speech utterances of different dialects of huge speech training samples from frame and utterance level. These features are given as input to DNN model for learning. By applying the backpropagation method, update the weights and learning rate in order to get expected value using error rate. The error rate or gradient at time step  $k$  is given in Eq. (3):

$$\frac{\partial E_k}{\partial W} = \frac{\partial E_k}{\partial \theta^k} \frac{\partial \theta^k}{\partial d_k} \left( \prod_{t=2}^k \frac{\partial d_t}{\partial d_{t-1}} \right) \frac{\partial d_1}{\partial W} \tag{3}$$

where,  $k$  is high, the equation tends to fade because the derivation of the Softmax activation function is less than 1. The Stochastic gradient approach allows for a high  $k$  value to be achieved with less iteration. The stochastic gradient descent (SGD) method is used to update the weights of the networks in order to reduce the error rate. It will repeat the process until the model is stabilized. In this, the activation function in output layer Softmax is used and it is used to normalize the output value given by model to 1.

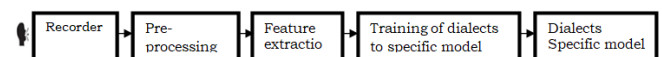


Figure 4. Basic diagram of training phase

### 3.2 Testing phase of DNN

The testing phase of DNN based dialect identification system gives the dialect of unknown utterance of Telugu speech utterance of unknown speech. It involves 39-dimensional MFCC+ΔMFCC+ΔΔMFCC feature extraction from unknown utterance from frame level and utterance level and evaluation against each DNN of three dialects of Telugu language. Like Machine learning model, the DNN based dialect identification system gives the dialect of unknown utterance of speech signal based on the maximum likelihood as shown in Figure 5. The output of training phase is represented by  $\beta$  (collectively represents the updated weight  $W_i$  and learning rate  $\theta_i$ ) is used to identify the dialects of unknown speech samples).

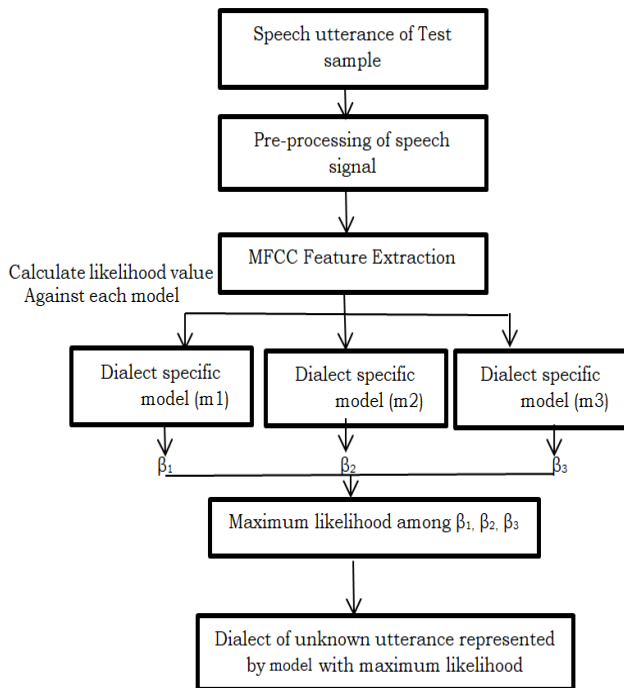


Figure 5. Dialect identification testing phase

### 3.3 Proposed method

The DNN can be considered as a branch of Machine learning where high level functions are retrieved from the input information and that can be incorporating several layers of nodes in the system. Through this mode creative and abstract component can be extracted from the input data.

The DNN is more popular because it is capable of extracting even complex features from large amount of data even from hours of speech data both Linear as well as non-linear because of their deeper architecture [17] as shown in Figure 6.

The neurons present in the first layer (not hidden) will take the input data and its output will become as input to the neurons present in successive layer and so on thus it gives the final output as represented in Figure 3 and 6. The produced final output will be like yes or no which in turn represented as probability. Every layer may contain one or more neurons and uses a function on these neurons called as “activation function”. This function normalize the output produced by neuron and sends the signal to next layer as input to the further connected neurons as shown in Figure 7.

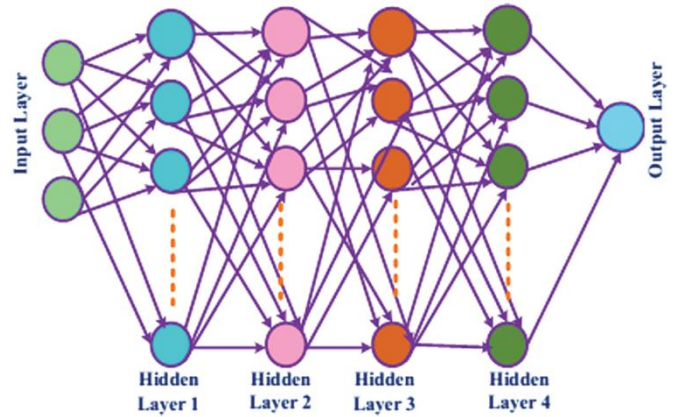


Figure 6. Basic DNN structure contain 2-hidden layer

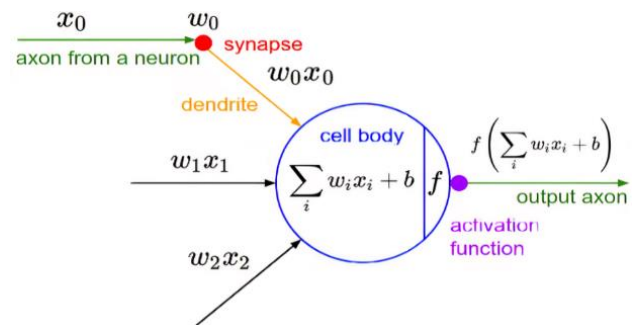


Figure 7. Neuron structure

The output of cell could be a weighted total of its inputs and bias terms, which is given to activation function  $F$ . The output of node in the  $n^{\text{th}}$  layer can be represented in Eq. (4):

$$y_{n,j} = F \left( \sum_i (y_{n-1,i} * w_{n,i,j}) + b_{n,j} \right) \quad (4)$$

The output is further passed only if the results value of incoming neurons is more than the threshold. A weight will be assigned to the connection that exists between the neurons of successive layers. This weight defines the importance of the input on the output for the next neuron as a result for the overall final output. The weights are initialized to random value and the weights will get modified by backpropagation, in each iteration in the training phase of DNN. The input layers use the ReLU (Redressed straight unit) activation function. Scientifically, it is characterized as  $y = \max(0, u)$ . The working of ReLU activation function is shown in Figure 8.

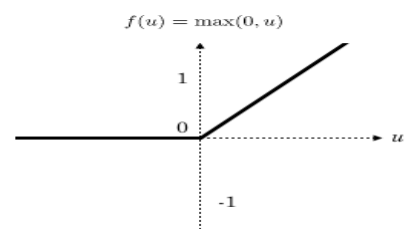


Figure 8. ReLU activation function

The activation function used in output layer is Softmax activation function and it is given in Eq. (4) and different parameters used in DNN model is shown in Table 2.



$$F(z_j) = \frac{e^{z_j}}{\sum_j(e^{z_j})} \quad (5)$$

where,  $z$  is input vector of  $j$  real numbers and  $e$  is constant.

**Table 2.** Parameters used in training DNN

Parameters	Value
Total hidden layers	2
Number of neurons in each layer	Hidden1=30, Hidden2=12
Input Size	39(MFCC+ $\Delta$ MFCC+ $\Delta\Delta$ MFCC)
Cross validation	True
Technology(to train)	Tensor flow

#### 4. RESULTS

In this, we used distinctive speech signals are used in training and testing phase. The speech signals that recorded in database creation are given to DNN at the time of training section and DNN is trained for each dialect separately. In proposed system, the speech signals that were used in the testing phase are Telangana, Costa Andhra and Rayalaseema dialects. The output of training phase of each dialect is represented by  $\beta$  (which contains the updated weights  $W_i$  and learning rate  $\theta_i$  of each dialect) which is given by DNN and it is given as input to the testing phase of DNN. Within the proposed system it generates 3  $\beta$  cherish each for every dialect. Dialects may be known, once a test for articulation gets a lot of affinity score for those each dialect. If a test sample gets a lot of affinity score for Telangana dialect, it demonstrates that it had been spoken in Telangana dialect and same just in case of Kosta Andhra and Rayalaseema additionally. In Figure 7 it shows the total parameters used by DNN model and calculating the accuracy by epochs=50.

**Table 3.** Complete dataset for both training and testing

Dialect	Total time of speech data	Speakers for each dialect	No. of Test samples
Telangana	2h 10 min	72	175
Costa Andhra	1h 57 min	69	140
Rayalaseema	1h 38 min	58	137

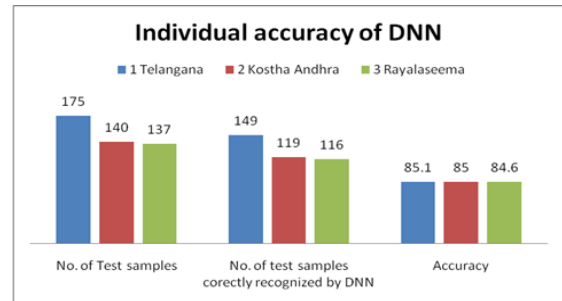
Table 3 above provides the complete details of database used for the dialect identification. In identify the dialects applied the process to 50 epochs. The accuracy of the DNN model for each iteration is called learning the model.

The total number of test samples and recognized samples are shown in Table 4.

**Table 4.** Individual accuracy of DNN

Dialects	No. of Test samples	No. of test samples correctly recognized by DNN	Accuracy
Telangana	175	149	85.1
Costa Andhra	140	119	85.0
Rayalaseema	137	116	84.6

From the above results, out of 175 samples 149 Telamngana samples are identified by DNN. The accuracy of model with respect to Telangana is 85.1. For Andhra out of 140 test samples model identified the 119 and produce accuracy is 85%. For Rayalaseema, 116 test samples are correctly classified by DNN model out of 137 samples. The accuracy given by DNN model is 84.6. Overall, the DNN Model produce good performance with respect to Telangana dialects. The comparison of accuracies produced by DNN is shown in Figure 9.



**Figure 9.** Diagrammatic representation of individual accuracy of DNN

#### 4.1 Comparison of results with different models

In these different models like GMM and HMM performance is compared with proposed DNN based model.

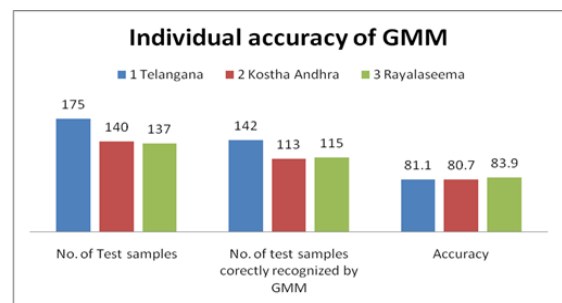
##### 4.1.1 Gaussian mixture model (GMM)

GMM model is also applied to identify the dialects of Telugu language. The accuracy produced by GMM model is given in Table 5. The GMM based dialect identification is designed with the input sequence of 39-dimensional feature vectors (MFCC+ $\Delta$ MFCC+ $\Delta\Delta$ MFCC) with 32 Gaussian mixtures and achieved the good performance.

From the results, GMM model produced the accuracy 81.1%, 80.7% and 83.9% for Telangana, Costa Andhra and Rayalaseema dialects respectively. The corresponding graph is shown in Figure 10.

**Table 5.** Individual accuracy of GMM

Dialects	No. of Test samples	No. of test samples correctly recognized by GMM	Accuracy
Telangana	175	142	81.1
Costa Andhra	140	113	80.7
Rayalaseema	137	115	83.9



**Figure 10.** Performance of GMM model in dialect identification

Overall, an average performance of GMM model is 81.8%. It performed well in identify the Rayalaseema dialects.

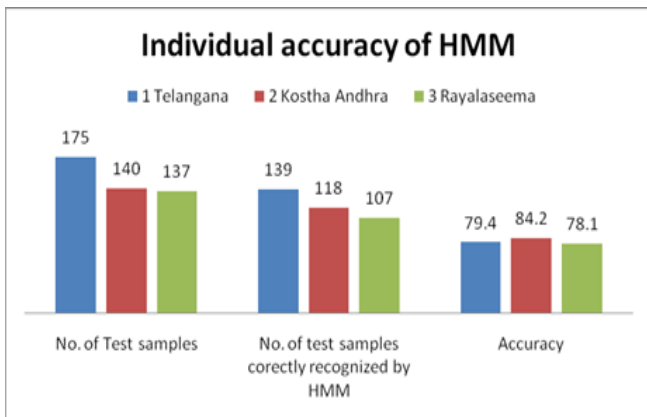
#### 4.1.2 Hidden Markov Model (HMM)

The 39-dimensional feature vectors (MFCC+ΔMFCC+ΔΔMFCC) are extracted from the training speech of each dialect of the Telugu language and used these feature vectors as the input sequence. The HMM-based dialect identification system is designed with three states for the input sequence of 39-dimensional feature vectors (MFCC+ΔMFCC+ΔΔMFCC) and achieved the performance is 79.4%, 84.2% and 78.1% for Telangana, Costa Andhra and Rayalaseema respectively as shown in Table 6.

**Table 6.** Individual accuracy of HMM

Dialects	No. of Test samples	No. of test samples correctly recognized by HMM	Accuracy
Telangana	175	139	79.4
Costa Andhra	140	118	84.2
Rayalaseema	137	107	78.1

The detailed results produced by HMM model is shown in Figure 11.



**Figure 11.** Diagrammatic representation of individual accuracy of HMM

**Table 7.** Total No. of test samples recognized by individual model

Dialects	No. Of Test Samples	No. of Test Samples Correctly Recognized by			Accuracy		
		GMM	HMM	DNN	GMM	HMM	DNN
Telangana	175	142	139	149	81.1	79.4	85.1
Costa Andhra	140	113	118	119	80.7	84.2	85
Rayalaseema	137	115	107	116	83.9	78.1	84.6
TOTAL	452	370	364	385	81.8	80.7	84.5

**Table 8.** Overall performance of differed models

Dialects	No. of Test samples	Complete test samples	Accuracy of GMM	Accuracy of HMM	Accuracy of DNN
Telangana	175				
Costa Andhra	140	452	82	80.7	84.5
Rayalaseema	137				

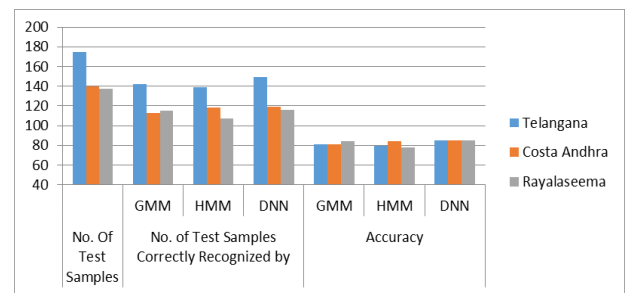
From the above table, HMM model produced good results in Costa Andhra dialects identification and it produced 79.4%, 84.2% and 78.1% for Telangana, Costa Andhra and Rayalaseema respectively. The overall accuracies produced by different models as shown in Table 7.

#### 4.2 Overall accuracy

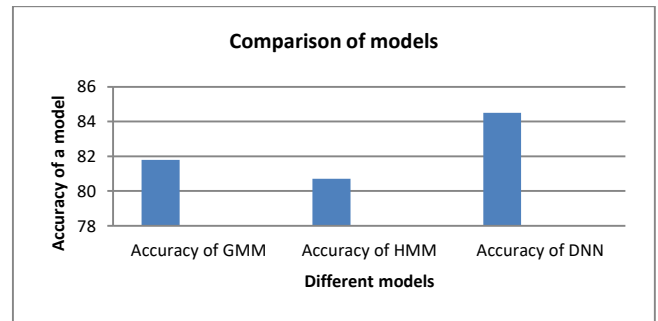
The complete performance of different models are shown in Table 7. The corresponding graph is shown in Figure 12.

From the Table 7 results, it is observed that DNN model correctly classifies the 385 test samples out of 452 and produced the 84.5% accuracy compare to remaining models.

From the Table 8 results, DNN model produced good results with 84.5% compare to remaining GMM and HMM model. The total samples used for testing purpose is 452. The corresponding graph as shown in Figure 13.



**Figure 12.** Performance of models



**Figure 13.** Comparison of different models in Telugu dialect identification

## 5. CONCLUSIONS

In this applied the MLP-DNN model to identify the three dialects of Telugu language i.e., Telangana, Costa Andhra and Rayalaseema and compare the results with GMM, HMM models. For DNN model, we use 2 hidden layers with different activation functions and each one contains 12 and 10 neurons respectively, and one output layer with softmax activation functions used. In this, total 39 MFCC features extracted from speech corpus and given to DNN for train and testing. We created 5h 45min dialect dataset, out of which 1h 10 min used for testing purpose. We applied 3 different data models GMM, HMM, DNN to dataset. GMM produces 81.8%, HMM is 80.7% and DNN produced 84.5% accuracy. Overall DNN produced good accuracy. In future, increase the dataset size and also epochs to increase the accuracy. We can also apply different models and feature extraction techniques to identify accurate dialect.

## REFERENCES

- [1] Venkatalakshamma, M., Munirathamma, N. (2014). Importance of Andhra Pradesh mother tongue-a study on Telugu language. *Indian Journal of Applied Research*, 4(12): 1-2.
- [2] Babu, B.P. (1976). A phonetic and phonological study of some characteristic features of Telugu English including reference to the source and target languages. University of London, School of Oriental and African Studies (United Kingdom).
- [3] Kumar, A., Mittal, V. (2019). Speech recognition: A complete perspective. *International Journal of Recent Technology and Engineering*, 7(6C).
- [4] Anusuya, M.A., Katti, S.K. (2010). Speech recognition by machine, A review. *International Journal of Computer Science and Information Security (IJCSIS)*, 6(3): 181-205. <https://arxiv.org/abs/1001.2267>.
- [5] Nassif, A.B., Shahin, I., Attili, I., Azzeh, M., Shaalan, K. (2019). Speech recognition using deep neural networks: A systematic review. *IEEE Access*, 7: 19143-19165. <https://doi.org/10.1109/ACCESS.2019.2896880>
- [6] Tzudir, M., Sarmah, P., Prasanna, S.M. (2018). Dialect identification using tonal and spectral features in two dialects of Ao. In *SLTU*, pp. 137-141. <https://doi.org/10.21437/SLTU.2018-29>
- [7] Chittaragi, N.B., Prakash, A., Koolagudi, S.G. (2018). Dialect identification using spectral and prosodic features on single and ensemble classifiers. *Arabian Journal for Science and Engineering*, 43(8): 4289-4302. <https://doi.org/10.1007/s13369-017-2941-0>
- [8] Sastry, J.V. (1987). A study of Telugu regional and social dialects: A prosodic analysis. University of London, School of Oriental and African Studies (United Kingdom).
- [9] Londhe, N.D., Kshirsagar, G.B. (2017). Speaker independent isolated words recognition system for Chhattisgarhi dialect. In *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, pp. 1-6. <https://doi.org/10.1109/ICIIECS.2017.8276169>
- [10] Shivaprasad, S., Sadanandam, M. (2020). Identification of regional dialects of Telugu language using text independent speech processing models. *International Journal of Speech Technology*, 25(1): 1-8. <https://doi.org/10.1007/s10772-020-09678-y>
- [11] Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A. R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N., Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6): 82-97. <https://doi.org/10.1109/MSP.2012.2205597>
- [12] Mengistu, A.D., Melesew, D. (2017). Text independent Amharic language dialect recognition: A hybrid approach of VQ and GMM. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 10(1): 215-222.
- [13] Ouali, M.A., Ghanai, M., Chafaa, K. (2020). TLBO optimization algorithm based-type2 fuzzy adaptive filter for ECG signals denoising. *Traitement du Signal*, 37(4): 541-553. <https://doi.org/10.18280/ts.370401>
- [14] Chitturi, R., Hansen, J.H. (2007). Multi-stream dialect classification using SVM-GMM hybrid classifiers. In *2007 IEEE Workshop on Automatic Speech Recognition & Understanding (ASRU)*, Kyoto, Japan, pp. 431-436. <https://doi.org/10.1109/ASRU.2007.4430151>
- [15] Mohamed, A.R., Sainath, T.N., Dahl, G., Ramabhadran, B., Hinton, G.E., Picheny, M.A. (2011). Deep belief networks using discriminative features for phone recognition. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, pp. 5060-5063. <https://doi.org/10.1109/ICASSP.2011.5947494>
- [16] Sadanandam, M., KamakshiPrasad, V. (2019). New features for language recognition from speech signal. *International Journal of Advanced Intelligence Paradigms*. <https://doi.org/10.1504/IJAIP.2019.10023086>
- [17] Bulla, P., Anantha, L., Peram, S. (2020). Deep neural networks with transfer learning model for brain tumors classification. *Traitement du Signal*, 37(4): 593-601. <https://doi.org/10.18280/ts.370407>