

Chapter 7

Conclusions

This research is about exploring, reviewing, and proposing various speech recognition methodologies for Gujarati language. In the chapter 1 of this thesis, we reviewed that human auditory perception imitation has given rise to this interdisciplinary field of speech recognition. The types and applications of speech recognition are addressed in chapter 1. This chapter also contains the objectives, importance, literature survey, related speech recognition in Indian regional languages, and our work. Then, in the chapter 2, various preliminaries and methodologies are discussed, which have been used in our models for speech recognition in Gujarati language. The detailing of the models created was described in chapters 3 and 4.

In chapter 3, the models are for isolated speech recognition. In all the models in this chapter, we have considered the mel-frequency cepstral coefficient for feature extraction from speech. The highest accuracy obtained was 90% for the model based on classification using a radial basis function network.

In chapter 4, two types of models are considered: isolated words and continuous sentences. In this chapter, we have considered mel-frequency discrete wavelet coefficients for feature extraction.

- For the recognition of isolated words, the model based on the hidden Markov model gave the highest recognition accuracy 100%.
- For the augmented dataset, multilayered perceptron gave the highest recognition accuracy at 85%.
- For the recognition of continuous sentences, a model based on multilayered perceptron trained with the gradient descent method gave the highest recognition accuracy of 84.62%.

We used ensemble learning models in our work, discussed in the chapter 5. Bagging ensemble models were used to combine odd numbers of models based on multilayered perceptrons and hidden Markov models. An ensemble learning model based on multilayered perceptron showed accuracy of 100% and 85.71% for two sentences, respectively, as compared to the hidden Markov model with 87.50% and 85.71% for the same. We also created a graphical user interface, discussed in the chapter 6. It is useful for recording a speech using a pre-defined vocabulary and then recognising it using our trained models.

The objective of this research work was to explore, review, and propose various speech recognition methodologies for the Gujarati language. This is nicely addressed by various models, as discussed in chapters 3 to 6. Our project is a kick-start for ASR in Gujarati language, however, it is limited to a small vocabulary with static training and is limited for urban dialects.

7.1 Future scope:

The insights gained from this research work motivate us to continue expanding the work.

- In the future, we plan to increase the size and diversity of the dataset. Specifically, we will collect recordings from a broader range of Gujarati speakers, including those from various dialects. This will ensure that the dataset is more representative and comprehensive. Using such a large dataset of speech recordings from Gujarati speakers across various dialects in the models discussed in our research will improve the generalization of the models, thereby enhancing their usability.
- Additionally, we intend to explore advanced deep learning techniques, such as recurrent neural networks (RNNs), to make the speech recognition model more dynamic and adaptable. RNNs are preferred over MLPs for sequential data, such as speech, because they are specifically designed to capture the temporal dynamics inherent in such data. By training on a larger dataset of speech recordings from Gujarati speakers across various dialects, RNNs can make the model more robust and better suited for real-world applications.
- A key objective will be to integrate a feature into the speech recognition system that allows it to incorporate new speech data from users during usage. This will require developing a mechanism to recognize new speech inputs and simultaneously add them as new instances to the dataset, enabling dynamic growth of the dataset.

Successfully addressing these future directions will pave the way for the development of open-source Gujarati speech recognition software, which holds the potential to benefit a wide range of users and applications.