

## Abstract

The development of voice recognition systems tailored to vernacular dialects holds transformative potential for enhancing accessibility and inclusivity in technology. This thesis focuses on creating a voice recognition model specifically designed for vernacular Gujarati dialects, addressing the unique linguistic and phonetic challenges inherent in regional variations of the language.

The key part of this research was to gather a diverse and representative spoken Gujarati corpora sourced via varied public repositories, which includes radio broadcast, interview, folk song, community recording and public availability speech corpora. This dataset includes a variety of dialectal variation in phonology, syntax and usage to guarantee robustness and inclusivity to the development of the models.

A dialect-specific recognition system using advanced techniques in voice recognition system, including deep learning architectures the proposed framework and model was developed. The model is further enriched with dialectal linguistic features integrated to its architecture, phoneme based pretraining to increase recognition accuracy, and transfer learning to adapt general speech recognition systems to dialect specific nuances.

The model was evaluated and found to achieve substantial improvement in phoneme recognition accuracy over baseline systems. The results show that modeling context-aware, high quality, diverse datasets are crucial to vernacular speech recognition. The system developed is there to provide practical applications for voice enabled user interface, digital accessibility and protection of linguistic diversity more specific examples of such languages which are least represented.

This work contributes to the emerging area of regional language processing with an end-to-end framework that can be used for future work on low-resource languages and dialects and to build inclusive, ubiquitous and accessible technology solutions in multilingual communities.