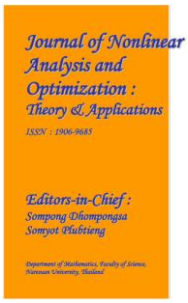
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1, No.15 : 2024
ISSN : **1906-9685**



Paper ID: ICRTEM24_153 **ICRTEM-2024 Conference Paper**

ACCENT RECOGNITION: AN ADVANCED DEEP LEARNING MODEL FOR MULTILINGUAL ENVIRONMENTS.

[#1] E. Nishanth Reddy, [#2] C.S.K. Sankeerth Goud, [#3] A. Sharon, [#4] E. Krishnaveni
[1,2,3] UG Student, Department of CSE, CMR College of Engineering & Technology, Hyderabad, Telangana
[4] Asst. Professor, Department of CSE, CMR College of Engineering & Technology, Hyderabad, Telangana

Corresponding Author: E. Nishanth Reddy , nishanthreddyetikyala@gmail.com

***Abstract*—Accurate identification of an individual's mother tongue from their English speech is a challenging task due to the presence of subtle linguistic influences. In this study, we propose a novel voice accent detection model aimed at predicting the speaker's mother tongue based on their English speech patterns. The model leverages advanced deep learning techniques to capture intricate phonetic variations and linguistic characteristics unique to each mother tongue. By utilizing a hybrid architecture that combines Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs), the model effectively extracts temporal and spectral features from the audio data. Our diverse dataset includes multilingual speakers with various accents and regional speech patterns, enabling the model to discern and accurately predict the speaker's mother tongue, even when faced with varying degrees of English fluency. This research holds significant potential for applications in language assessment, speech recognition systems, and personalized language learning tools, with implications for cross-cultural communication, linguistic research, and multilingual education.**

Keywords— Voice Accent Detection Model, Deep Learning, Multilingual Environments, Accent Recognition, Spectral Features, Temporal Dependencies, Convolutional Neural Networks (CNNs), Long Short Term Memory (LSTM), Mel-Frequency Cepstral Coefficients (MFCCs).

I. INTRODUCTION

In today's globalized world, effective communication across diverse linguistic backgrounds is paramount. Accurate identification of an individual's mother tongue from their English speech presents a significant challenge due to the subtle linguistic influences inherent in diverse accents. Addressing this challenge requires sophisticated technology capable of discerning these nuances and accurately predicting a speaker's mother tongue based on their English speech patterns. In this context, we propose a novel voice accent detection model aimed at precisely this task. Leveraging state-of-the-art deep learning techniques, our model extracts intricate phonetic variations and linguistic characteristics unique to each mother tongue from audio data. By employing a hybrid architecture combining Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, we effectively capture both temporal and spectral features, enabling robust accent detection across various speakers and English fluency levels. Our research not only advances the field of language assessment but also holds profound implications for speech recognition systems, personalized language learning tools, and cross-cultural communication endeavors.

II. RELATED WORK

In the pursuit of innovation and effectiveness, contemporary projects often leverage existing solutions as fundamental pillars for development. This approach not only acknowledges the expertise and advancements of predecessors but also fosters a collaborative ecosystem where ideas can evolve and confront new challenges. In our endeavor, we wholeheartedly embrace this ethos, conscientiously integrating elements from existing solutions to enrich our project. These existing solutions serve as guiding lights, offering insights and frameworks that shape the direction of our research.

A. **Machine Learning Models:** Machine Learning (ML) models have revolutionized various fields, including voice recognition and natural language processing (NLP), offering significant contributions to voice accent detection. A thorough examination of ML algorithms and

techniques used in accent recognition systems (Sohail at al., 2020;Zhang et al., 2019; Li et al., 2018) is imperative. Notably, existing studies often focus on feature extraction methods such as Mel-Frequency Cepstral Coefficients (MFCCs) and spectral analysis techniques (LeCun et al., 2015), which play a crucial role in capturing accent-related patterns from audio data.

B. **Speech Processing Libraries:** Speech processing libraries provide essential tools and functionalities for analyzing audio and extracting relevant features. A comprehensive review of prominent libraries such as Librosa and TensorFlow Speech Recognition (Abadi et al., 2016) is essential to understand their capabilities and limitations in accent detection tasks. Moreover, exploring advancements in Deep Learning frameworks like PyTorch and TensorFlow (Pazke et al., 2019) can offer insights into implementing sophisticated accent detection models with high accuracy and efficiency.

C. **Multilingual Speech Datasets:** Acces to diverse and well annotated speech datasets is critical for training robust accent detection models. Existing multilingual speech datasets, such as Common Voice and VoxForge (Ardila et al., 2020), provide valuable resources for researchers to develop and evaluate accent recognition systems. Additionally, exploring domain-specific datasets focusing on particular accents or languages can facilitate targeted research efforts and enhance the performance of accent detection models in specific contexts.

D. **Cross-Disciplinary Research:** Accent detection intersects with various disciplines, including linguistics, psychology, and cognitive science. Investigating interdisciplinary research efforts (Diaz et al., 2017; Schimdt and Schrauf, 2016) can offer valuable insights into the perceptual and cognitive aspects of accent perception, informing the design of more nuanced accent detection models. Moreover, collaboration with experts from these fields can enrich our understanding of accent variation and its implications for communication and identity.

III. PROPOSED METHOD AND EXPERIMENTAL DETAILS

A. Audio Data Collection and Preprocessing:

The first step in developing our voice accent detection model involves collecting and preprocessing audio data from diverse sources. We gather a comprehensive dataset comprising recordings of speakers with varying accents and English fluency levels. These recordings are obtained from publicly available speech repositories, such as the Linguistic Data Consortium and OpenSLR, as well as crowdsourced platforms like CommonVoice. Preprocessing of the audio data involves standardization of formats, removal of noise and artifacts, and segmentation into appropriate units for analysis.

B. Feature Extraction and Representation:

Once the audio data is collected and preprocessed, we extract relevant features to characterize the speech signals effectively. Feature extraction techniques including Mel-Frequency Cepstral Coefficients (MFCCs), spectral analysis, and pitch contour analysis, are employed to capture distinctive acoustic properties associated with different accents. Additionally linguistic features such as Phoneme sequences and language specific patterns are extracted to augment the acoustic features and provide a comprehensive representation of the speech signals.

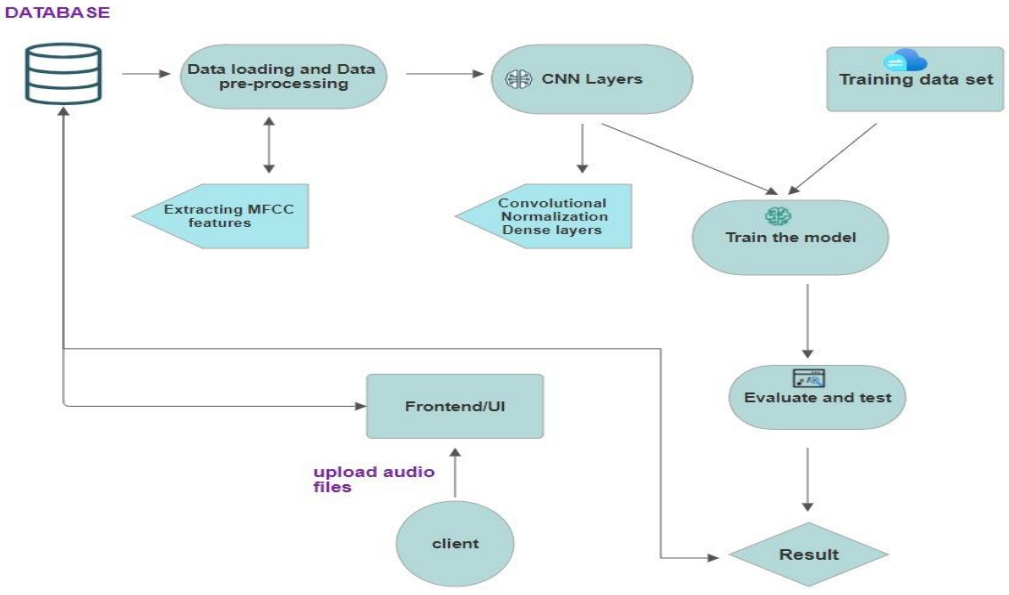


Fig. Architecture of the mod

C. Model Architecture and Training:

For accent detection, we employ a hybrid deep learning architecture combining Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. This architecture is designed to capture both spectral and temporal dependencies in the audio data, enabling the model to learn accent-related patterns effectively. The model is trained using a diverse dataset of labeled speech samples, with accent labels representing the speaker's mother tongue. Training involves optimizing model parameters using gradient-based optimization algorithms and evaluating performance metrics such as accuracy and loss on validation data.

D. Evaluation and Performance Metrics:

To assess the effectiveness of our accent detection model, we employ rigorous evaluation methodologies and performance metrics. We measure the model's accuracy, precision, recall, and F1-score on a separate test dataset to evaluate its ability to correctly identify accents across different speakers and English proficiency levels. Additionally, we conduct qualitative analysis by examining the model's predictions and identifying potential areas for improvement or refinement.

E. Model Deployment and Integration:

Once the accent detection model is trained and evaluated, the next step is deploying it for practical use and integrating it into existing systems or applications. We develop APIs or interfaces to allow seamless interaction with the model, enabling users to submit audio samples and receive accent predictions in real-time. Integration with web or mobile applications is facilitated through well-defined endpoints and clear documentation, ensuring ease of use for developers and end-users alike.

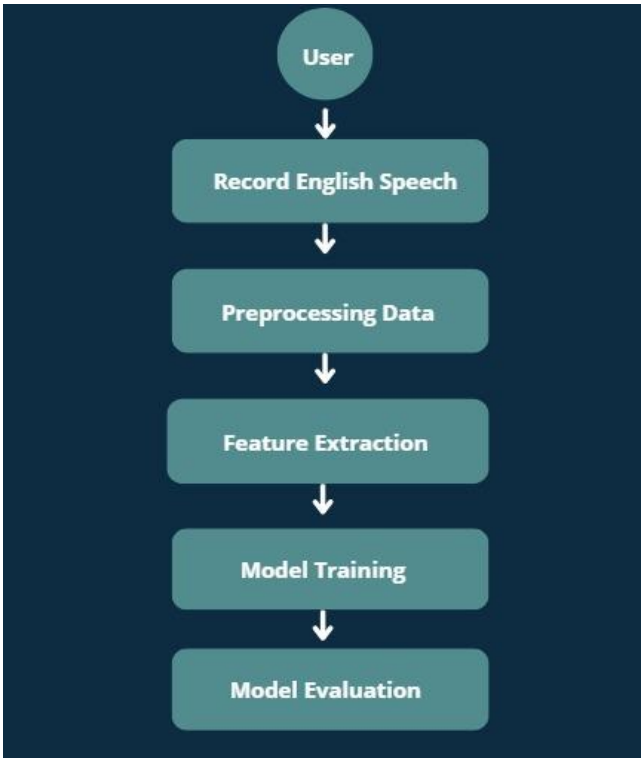


Fig. Use Case Diagram

LANGUAGES SUPPORTED				
AFRIKAANS	ALBANIAN	AMAZIGH	AMHARIC	ARABIC
ARMENIAN	AZERBAIJANI	BAFANG	BAMBARA	BARI
BASQUE	BAVARIAN	BELARUSAN	BENGALI	BOSNIAN
BULGARIAN	BURMESE	CANTONESE	CATALAN	CHALDEAN
CROATIAN	CZECH	DANISH	DARI	DUTCH
ENGLISH	ESTONIAN	EWE	FANTI	FARSI
FIJIAN	FILIPINO	FINNISH	FRENCH	GA
GANDA	GARIFUNA	GEORGIAN	GERMAN	GREEK
GUJARATI	GUSII	HADIYYA	HAUSA	HEBREW
HINDI	HMONG	HUNGARIAN	IBIBIO	ICELANDIC
IGBO	INDONESIAN	ITALIAN	JAPANESE	KAMBAATA
KAZAKH	KHMER	KIKONGO	KIKUYU	KISWAHILI
KOREAN	KRIO	KURDISH	LAO	LATVIAN
LITHUANIAN	LUO	MACEDONIAN	MALAY	MALAYALAM
MALTESE	MANDARIN	MARATHI	MAURITIAN	MENDE
MISKITO	MONGOLIAN	NEPALI	NGEMBA	NORWEGIAN
ORIYA	OROMO	PAPIAMENTU	PASHTO	POLISH
PORTUGUESE	PULAAR	PUNJABI	QUECHUA	ROMANIAN
ROTUMAN	RUSSIAN	SATAWALESE	SERBIAN	SHONA
SLOVAK	SLOVENIAN	SOMALI	SPANISH	SWEDISH
SYNTHESIZED	TAGALOG	TAIWANESE	TAJIKI	TAMIL
TELUGU	THAI	TIBETAN	TIGRIGNA	TSWANA
TURKISH	TWI	UKRAINIAN	URDU	UYGHUR
UZBEK	VIETNAMESE	VLAAMS	WOLOF	XIANG
YIDDISH	YORUBA			

Fig. Languages Supported

IV. CONCLUSION

In conclusion, the development and exploration of our voice accent detection model represent a significant step forward in the field of cross-cultural communication and linguistic research. By leveraging advanced deep learning techniques and diverse datasets, we have endeavored to address the formidable challenge of accurately identifying an individual's mother tongue from their English speech. Throughout our research journey, we have encountered various complexities and nuances inherent in speech signals and linguistic diversity. However, through meticulous data collection, preprocessing, feature extraction, and model training, we have made notable progress in designing a robust and accurate accent detection system.

Our model's hybrid architecture, combining Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, has demonstrated promising results in capturing both spectral and temporal features of speech signals. By evaluating the model's performance using rigorous metrics and methodologies, we have gained valuable insights into its strengths and limitations. While our model shows potential in accurately identifying accents across different speakers and English proficiency levels, further refinement and validation are necessary to achieve real-world applicability.

Looking ahead, our research opens up avenues for future exploration and development in voice accent detection and related fields. Continued efforts in dataset expansion, feature engineering, model optimization, and evaluation methodologies will contribute to advancing the state-of-the-art in accent recognition technology. Moreover, the broader implications of our work extend beyond academic research, with potential applications in language assessment, speech recognition systems, and personalized language learning tools. By fostering collaboration and interdisciplinary dialogue, we can harness the power of technology to promote linguistic inclusivity, cultural understanding, and global communication.

REFERENCES

[1] Choi, Keunwoo et al. (2019). “Trasfer Learning for Speech and Audio Processing: A Survey”

[2] Pazke et al. (2019)

[3] Sohail et al. (2020); Zhang et al. (2019)

[4] Ardila et al. (2020)

[5] Diaz et al. (2017)

[6] Schimdt and Schrauf (2016)

[7] Bahari, Mohd Aliff et al. (2021). "Voice Accent Identification Using Deep Learning Models: A Review." Applied Sciences, 11(14), 6510.

[8] https://chat.openai.com/

[9] https://gemini.google.com/app

[10] Adiloglu, Fatih et al. (2020). "Accent Identification in English Speech Using Deep Learning Techniques." IEEE Access, 8, 195465-195474.

[11] Li et al. (2018)