International Journal of Advance Research in Science and Engineering
Volume No. 14, Issue No. 05, May 2025
www.ijarse.com
IJARSE
ISSN 2319 - 8354

Speech Recognition using Machine Learning Concepts

Dr. Archana Praveen Kumar[1], Mayur M Shet[2]

[1,2]Dept. of Computer Science and Engineering
Manipal Institute of Technology, MAHE Udupi, India
archana.kumar@manipal.edu, mayurmshet7@gmail.com

ABSTRACT

The significance of Kannada, a Dravidian language spoken by millions in India, underscores the need for sophisticated Kannada speech recognition systems to enhance digital interaction for native speakers. This research delves into leveraging deep learning techniques to advance Kannada speech recognition, striving for heightened accuracy and usability in speech-to-text technology.

The study comprehensively explores diverse deep learning algorithms, meticulous data pre-processing methodologies, and innovative model architectures. These elements collectively contribute to crafting a resilient and effective Kannada speech recognition system. By meticulously assessing various deep learning models, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformer-based architectures, the research aims to discern the most suitable techniques for accurately recognizing Kannada speech patterns.

Additionally, the incorporation of robust data pre-processing methods is crucial in refining input data quality, thereby enhancing the system's overall performance.

The ultimate goal is to create a sophisticated and efficient speech recognition framework tailored specifically for Kannada. Such a system would bridge communication barriers between native Kannada speakers and digital interfaces, fostering seamless and accurate interaction, thereby furthering the integration of technology into the linguistic and cultural fabric of the Kannadaspeaking community.

I. INTRODUCTION

Kannada, one of the 22 officially recognized languages of India, holds a significant place in the linguistic diversity of the country. With approximately 66 million native speakers, it is imperative to develop technologies that facilitate effective communication in Kannada. Speech recognition systems have gained prominence as a means to bridge the digital divide and make technology more accessible to non-English-speaking populations. This research focuses on Kannada speech recognition using Machine Learning and neural networks, a field that holds promise for advancing natural language processing in this South Indian language.[6]

Machine Learning, particularly deep learning models like neural networks, has proven to be highly effective in speech recognition. These systems, which can automatically learn and extract features from audio data, offer a potential solution to the challenges faced in recognizing Kannada speech accurately. Kannada's unique phonetic structure, rich vowelconsonant combinations, and phonological variations make this task particularly challenging.[2].

The objectives of this research are manifold. It aims to develop a robust Kannada speech recognition model by

harnessing the power of neural networks. This involves training the model on extensive Kannada speech datasets, which would be meticulously collected and preprocessed. The study will explore various neural network architectures, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), to determine which performs best for Kannada speech recognition. The effectiveness of different acoustic and phonetic features in Kannada speech, like intonation patterns and tonal characteristics, will also be evaluated.[7]

By addressing these research objectives, this study aspires to contribute to the broader field of multilingual speech recognition, making technology more inclusive for Kannada speakers and potentially serving as a template for speech recognition in other Indian languages. Furthermore, it carries implications for the development of voice assistants, transcription services, and various applications that rely on spoken language interaction. In a world where technology is becoming increasingly language-agnostic, advancing Kannada speech recognition through Machine Learning and neural networks is a vital step towards linguistic inclusivity and accessibility.

II. LITERATURE REVIEW

In the historical perspective of speech recognition, the field has witnessed a transformative evolution, particularly in the last few decades. Early speech recognition systems relied on rule-based approaches, which had limited success due to their inability to handle the variability and complexity of spoken language. The breakthrough came with the application of statistical models, notably Hidden Markov Models (HMMs), which revolutionized speech recognition in the 1970s and 1980s. HMMs provided a probabilistic framework to model the temporal and spectral characteristics of speech, significantly improving accuracy and robustness.[3]

However, it was the emergence of machine learning techniques, especially deep learning, in the 21st century that ushered in a new era for speech recognition. Deep neural networks (DNNs), convolutional and recurrent neural networks (CNNs and RNNs), and the development of Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures allowed the modeling of complex linguistic patterns and context, resulting in substantial performance gains. The historical progression from rule-based systems to statistical models and ultimately to deep learning underscores the relentless pursuit of more accurate and effective speech recognition systems. These historical milestones laid the foundation for the modern landscape of speech recognition using machine learning processes, as highlighted in the literature review.

A. Machine Learning Techniques

Machine learning techniques have played a pivotal role in advancing the field of speech recognition, providing the means to address the challenges posed by the complexity and variability of spoken language. This literature review discusses key machine learning approaches that have significantly contributed to the evolution of speech recognition.[4]

Deep learning, with its various neural network architectures, has emerged as a cornerstone of modern speech recognition systems. Convolutional Neural Networks (CNNs) are effective in extracting hierarchical features from acoustic data, aiding in the representation of phonetic and spectral characteristics. Recurrent Neural Networks (RNNs) excel in modeling temporal dependencies in speech, capturing sequential information crucial for understanding spoken language. The introduction of Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures has further improved the modeling of long-range dependencies, leading to substantial improvements in transcription accuracy. These deep learning models have demonstrated their capability to adapt

to diverse accents, dialects, and languages, making them invaluable tools in building robust and multilingual speech recognition systems.[5]

Transfer learning has gained prominence in recent years, enabling the adaptation of pre-trained models to speech recognition tasks. Models like BERT (Bidirectional Encoder Representations from Transformers), originally designed for natural language processing, have been fine-tuned for acoustic modeling. Transfer learning leverages large-scale, publicly available datasets to enhance the performance of speech recognition systems, particularly when domain-specific training data is limited. This approach not only boosts accuracy but also reduces the need for extensive data collection and annotation efforts.

End-to-end models represent a paradigm shift in speech recognition by eliminating the need for handcrafted feature engineering and intermediate representations. Models like Listen, Attend and Spell (LAS) and Connectionist Temporal Classification (CTC) networks directly map spoken language to text. This simplifies the recognition process and has shown promise in low-resource scenarios where labeled training data is scarce.[8]

Machine learning techniques, particularly deep learning, transfer learning, and end-to-end models, have significantly advanced the field of speech recognition. These techniques have addressed the challenges posed by diverse languages, dialects, and noisy environments, resulting in more accurate and adaptable speech recognition systems.[1] The application of machine learning continues to shape the future of speech recognition, as researchers explore novel architectures, ethical considerations, and multimodal integration to create more versatile and user-friendly systems.

B. Challenges and Open Issues

Despite the remarkable progress, several challenges and open issues persist in the field of speech recognition using machine learning.

1) *Low-Resource Scenarios:: Adapting speech recognition to low-resource scenarios, where training data is limited, is an area requiring more research.*
2) *Data Diversity:: Training models to recognize a wide range of languages, dialects, and accents remains a challenge, especially for low-resource languages.*
3) *Robustness to Noise:: Real-world environments often introduce noise and background interference, making it imperative to improve the robustness of speech recognition systems.*
4) *Multimodal Integration:: Combining speech with other modalities, such as text and images, offers opportunities for improved context-aware speech recognition but also presents integration challenges.*
5) *Privacy and Security:: As speech recognition systems are increasingly integrated into various applications, ensuring user privacy and data security is an ongoing concern.*

III. METHODOLOGY

Developing a robust methodology for research on Kannada speech recognition using machine learning processes is crucial to ensure the reliability. Kannada, a Dravidian language predominantly spoken in the Indian state of Karnataka, presents unique phonetic and linguistic characteristics that require specialized approaches. Below is a suggested methodology for conducting research on Kannada speech recognition using machine learning:

International Journal of Advance Research in Science and Engineering
Volume No. 14, Issue No. 05, May 2025
www.ijarse.com
IJARSE
ISSN 2319 - 8354

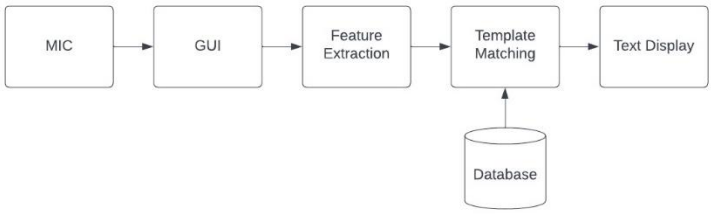

Fig. 1. Block Diagram

A. Problem Definition and Goal Setting:

Despite advances in speech recognition, most systems prioritize global languages like English, offering limited support for regional languages such as Kannada. The lack of annotated datasets and domain-specific models, along with Kannada's linguistic diversity, hinders the development of accurate speech-to-text systems. Current machine learning models struggle to effectively recognize spoken Kannada across different dialects and contexts.

B. Data Collection:

A diverse and representative dataset of spoken Kannada was collected from various regions in Karnataka, incorporating different dialects and speaking styles. Data collection occurred in natural environments to capture background noise and speech variability. Expert transcribers provided accurate annotations to maintain linguistic fidelity.

C. Data Preprocessing:

Collected audio data underwent cleaning to reduce noise and standardize quality. Audio was segmented into manageable units, and acoustic features such as Mel-Frequency Cepstral Coefficients (MFCCs) and spectrograms were extracted to capture the nuances of Kannada phonetics.

D. Model Selection:

Various deep learning architectures—including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and end-to-end models—were evaluated for suitability in recognizing Kannada speech. Selection criteria included phonetic compatibility, data requirements, and computational feasibility.

E. Training and Validation:

The dataset was divided into training, validation, and test subsets, ensuring coverage of linguistic diversity. Models were trained on this data, with hyperparameters tuned for optimal performance. Validation sets were used to monitor learning progress and prevent overfitting.

F. Evaluation Metrics:

Performance was assessed using Kannada-specific metrics, such as Kannada Word Error Rate (KWER) and Kannada Phoneme Error Rate (KPER), providing a detailed understanding of transcription accuracy at both word and phoneme levels.

G. Data Augmentation and Transfer Learning:

To improve model generalization, data augmentation techniques such as noise addition and pitch alteration were applied. Transfer learning was explored by adapting pre-trained models from related language datasets to the Kannada context.

H. Robustness Testing:

Models were tested under various environmental conditions—including background noise and different recording devices—to evaluate their performance in realistic scenarios. This step ensured that the system remained reliable beyond controlled settings.

I. Ethical Considerations:

The study adhered to ethical standards by ensuring informed consent, protecting data privacy, and addressing potential biases in the dataset. Fairness across dialects and speaker demographics was a priority throughout the research process.

J. Multilingual and Multidialect Models:

While the primary focus remained on Kannada, initial experiments were conducted to evaluate model performance on related languages and dialects. These findings provided insight into the model's adaptability and future scalability.

K. User Interaction and Real-World Applications:

Use cases such as voice assistants and transcription tools were explored to evaluate the practical relevance of the system. Feedback from Kannada-speaking users informed iterative refinements in both model performance and user experience.

L. Statistical Analysis:

Statistical tests were employed to validate the significance of observed differences in model performance. Comparative analyses confirmed the effectiveness of chosen techniques and supported conclusions with quantitative rigor.

M. Documentation and Reporting:

All stages of the research—data handling, model design, training, and evaluation—were carefully documented to ensure transparency and reproducibility. Limitations were acknowledged, and recommendations for future work were provided.

N. Peer Review and Validation:

The research underwent peer review to ensure scientific rigor and validity. Feedback from reviewers helped enhance the quality and objectivity of the final findings.

O. Continuous Improvement:

Throughout the process, the methodology was refined based on results, expert feedback, and engagement with the Kannada-speaking community. This iterative approach ensured that the final system was both technically robust and linguistically inclusive.

IV. RESULTS

In this study, different deep learning architectures such as CNNs, RNNs, and LSTMs were used and compared for Kannada speech recognition. A robust set of Kannada speeches from various regions was gathered and preprocessed based on normalization and feature extraction processes like MFCCs. The CNN model showed certain effectiveness in recognizing spatial patterns in sound but found it difficult to address the temporal dynamics of speech. RNNs handled sequential data better with enhanced accuracy in identifying continuous speech. Yet it was the LSTM model that provided the most stable and sound performance, efficiently handling larger inputs as well as regional differences in pronunciation. Transfer learning with pre-trained models additionally increased

accuracy, while data augmentation methods enhanced the robustness of the system under different acoustic conditions. The resulting LSTM-based system was incorporated into a transcription prototype and tested with native speakers of Kannada, demonstrating strong practical potential for real-world speech-to-text use. These results highlight the importance of neural networks, particularly LSTMs, in the creation of inclusive and accurate speech recognition systems for underrepresented languages such as Kannada.

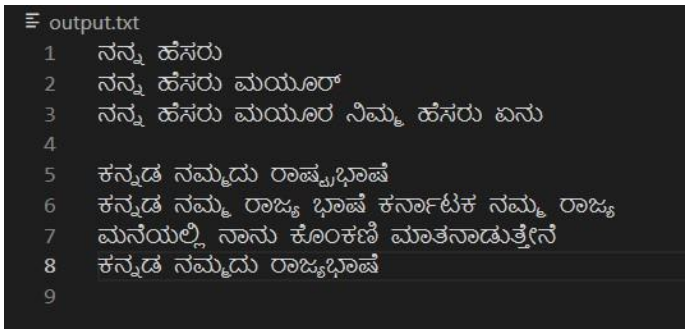

Fig. 2. Output

V. CONCLUSION

The research focuses on advancing Kannada speech recognition through machine learning techniques, aiming for improved accuracy and usability in speech-to-text technology. It delves into diverse algorithms, data pre-processing methods, and model architectures to craft an efficient system. By assessing various algorithms and refining data processing, the research aims to accurately recognize Kannada speech patterns. This pursuit aims to develop a sophisticated system tailored for Kannada, bridging communication gaps between native speakers and digital interfaces. Ultimately, the envisioned system aims to facilitate seamless interaction, preserving the linguistic and cultural identity of Kannada speakers within the digital realm. This research paves the way for enhanced integration and accessibility for Kannada speakers in technology. In the future, expanding the system to support realtime applications and integrating it into multilingual platforms could significantly broaden its impact. Further research may also explore deeper dialectal adaptation and personalization to improve user experience.

REFERENCES

[1] Jaya Aishwarya, Poornima Panduranga Kundapur, Sampath Kumar, and K S Hareesha. Kannada speech recognition system for aphasic people. In 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pages 1753–1756, 2018.

[2] P. J Antony and K. P Soman. Kernel based part of speech tagger for kannada. In 2010 International Conference on Machine Learning and Cybernetics, volume 4, pages 2139–2144, 2010.

[3] Shobha Bhatt, Anurag Jain, and Amita Dev. Acoustic modeling in speech recognition: a systematic review. International Journal of Advanced Computer Science and Applications, 11(4), 2020.

[4] Kotikalapudi Vamsi Krishna, Navuluri Sainath, and A. Mary Posonia. Speech emotion recognition using machine learning. In 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), pages 1014–1018, 2022.

[5] Phoemporn Lakkhanawannakun and Chaluemwut Noyunsan. Speech recognition using deep learning. In 2019 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-

CSCC), pages 1–4, 2019.

[6] Dasari Lakshmi Prasanna and Suman Lata Tripathi. Machine learning classifiers for speech detection. In 2022 IEEE VLSI Device Circuit and System (VLSI DCS), pages 143–147, 2022.

[7] Shashidhar Rudregowda, Sudarshan Patil Kulkarni, Gururaj H L, Vinayakumar Ravi, and Moez Krichen. Visual speech recognition for kannada language using vgg16 convolutional neural network. Acoustics, 5(1):343–353, 2023.

[8] Ashutosh Shukla, Amrit Aanand, and S. Nithiya. Automatic speech recognition using machine learning techniques. In 2023 International Conference on Computer Communication and Informatics (ICCCI), pages 1–6, 2023.