IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, November 2025
Impact Factor: 7.67

Kannada Speech Disorder Detection and Assistance Using Deep Learning

Chandan B L[1] and Prof. Suchi Raj R[2]
Student, Department of MCA[1]
Assistant Professor, Department of MCA[2]
Vidya Vikas Institute of Engineering and Technology, Mysore

Abstract: *The advent of artificial intelligence (AI) and machine learning (ML) has revolutionised how languages are processed, understood, and utilised, clearing the path for innovations in accessibility and communication. This paper presents a Speech Sound Disorder Detection System that leverages advanced AI models like Network of Long Short- Term Memory (LSTM) and Generative Adversarial Networks (GANs) to address the dual challenges of Kannada text prediction and audio generation. The proposed system features a modular architecture comprising a user interface, an application layer, machine learning models, and a robust data layer. The user inputs text, which undergoes preprocessing and is transmitted via the ML models. The LSTM model predicts the next word or sequence, while the GAN model generates realistic and contextually appropriate synthetic sequences. These outputs then transformed into audio using a text-to-speech module and displayed or played back to the user.The research highlights the efficacy of these models in generating coherent and grammatically accurate predictions, demonstrating their applicability in linguistic domains with limited datasets. Comprehensive testing validated the system's scalability, accuracy, and responsiveness under diverse conditions.*

Keywords: Machine learning, deep learning, audio, Kannda, accuracy

I. INTRODUCTION

India is a land of linguistic pluralism and therefore is a challenging and promising location on all matters of language processing and its accessibility. Kannada, a Dravidian language with millions of speakers is one such regional language to have attracted interest in study of natural language processing (NLP), since language is complicated in its grammar and its cultural interests. Nonetheless, the current Kannada text prediction and audio generation systems do not scalable and accurate in nature, as the majority of them cannot handle the contextual-level understanding and process it efficiently.

The given Speech Sound Disorder Detection System aims to resolve these gaps due the application of AI-based text and audio processing models. The system seeks to fill the gap in accessibility to Kannada speakers by developing the LSTM networks used to predict text and the GANs used to synthesize text so that specific situations that demand the use of quality and dynamic linguistic results can be fulfilled.

The development of the NLP methods has been impressive. Traditional rule-based methods have been replaced by data-driven approaches powered by deep learning. While these methods excel in processing English and other global languages, their use regional languages like Kannada requires significant adaptation due to differences in syntax, semantics, and phonetics. This paper explores the implementation of AI models tailored for Kannada, addressing unique linguistic challenges while leveraging the strengths of advanced architectures.

The study has a contribution to the study of NLP since it introduces a system which not only includes powerful ML models but moreover a user-friendly interface. It highlights the necessity of have scalable, accurate and culturally relevant solutions to linguistic processing, which lays ground work to further research on Indian regional languages in terms accessibility.

DOI: 10.48175/IJARSCT-29911

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, November 2025
Impact Factor: 7.67

II. PROBLEM STATEMENT

In spite of great advancements in the NLP field, It's still challenging to process regional Indian languages such as Kannada because of the lack databases, intricate grammar, and absence of contextually oriented models. The current systems would not be supportive of Kannada in terms of giving accurate text predictions and natural audio generation to make it useable and approachable. The proposed study will attempt to overcome these weaknesses by creating a Speech Sound Disorder Detection System that applies the LSTM and GAN models in precise text prediction and synthetic sequence generation, with the ability predict in speech. The proposed system will aim to provide better linguistic approach to accessibility and scalability of Kannada, which will open the way to wider-ranging application to Indian languages.

III. LITERATURE SURVEY

Sindhu and Sainin [1] conducted a systematic literatur review on applying techniques from deep learning for automated speech and voice disorder detection.

Their study highlights the increasing adoption of Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are examples of deep learning models for processing acoustic features. The paper underscores the advantages deep learning in feature extraction as well as classification, citing high accuracy in identifying voice disorders in contrast to conventional machine learning models. However, also points out challenges such as limited datasets and the requirement for strong generalisation across diverse linguistic populations. This work servs is the basis for exploring advanced architectures for speech disorder detection.

Georgoulas et al. [2][7] explored Support Vector Machine (SVM) application and wavelet transformations for speech sound classification and articulation disorder detection.

Their approach utilised wavelet features to represent phonetic properties and an SVM classifier to distinguish between normal and disordered speech. The study demonstrated the efficacy of combining wavelets and SVMs in handling noise and variations in speech signals. While the approach provided reliable results, the reliance on handcrafted features was noted as a limitation compared to contemporary models for deep learning that automatically learn representations.

Shahin et al. [3] looked into the automatic screening of kids with abnormalities of speech sounds (SSDs) using paralinguistic variables.

Their work focused on extracting features such as pitch, formant frequencies, and spectral characteristics to teach classifiers for machine learning. The outcomes highlighted the significance of paralinguistic features in capturing speech irregularities. However, the authors acknowledged the need for richer datasets to improve model robustness and scalability, particularly in multi-lingual contexts.

Hsiao et al. [4] proposed an end-to-end text- dependent speech sound disorder detection system for Mandarin-speaking children.

Their approach leveraged neural networks architectures, including CNNs and attention mechanisms, to detect disorders and provide diagnostic insights. The study emphasised the system’s ability to work in text-dependent scenarios, ensuring contextual relevance in diagnosis. While effective, the author noted that generalising the model other languages or text- independent use cases would require significant modifications and additional training data.

Ng et al. [5] addressed the detection of SSDs in Cantonese-speaking children using posterior-based speaker representations.

The study incorporated neural network models to extract speaker-specific characteristics, enabling accurate classification of disordered speech. Their findings showed a high degree of accuracy and robustness in handling tonal variations inherent in Cantonese. The authors suggested That incorporating data augmentation technique could further enhance the model's performance and applicability to other tonal languages.

In a similar vein, a study on posterior-based speaker representations for child speech [6] introduced a framework for SSD detection that focused on leveraging speaker-specific characteristics.

This work reinforced the effectiveness of posterior-based features in capturing disordered speech patterns and achieving high classification accuracy. However, the need for diverse datasets and real-world deployment scenarios was highlighted as an area for future work.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 4, November 2025 Impact Factor: 7.67

Dudy et al. [8][9] explored pronunciation analysis to identify SSDs in children.

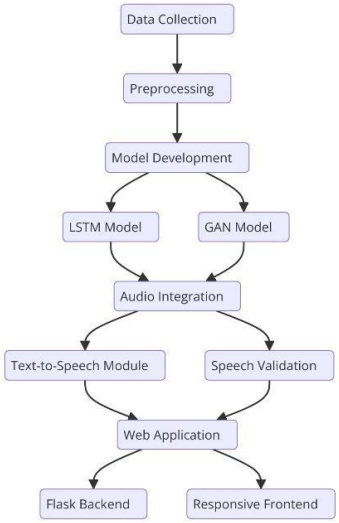
Their approach utilised phoneme-level acoustic and temporal features to evaluate pronunciation accuracy. The study provided information about the common pronunciation errors among children with SSDs, paving the path for personalised diagnostic tools. Although the approach was effective in controlled environments, the authors noted that handling spontaneous speech remains a significant challenge.

Rehman [10] investigated the applications the application of machine learning methods, such as decision trees and random forests, for voice disorder detection.

The study focused on extracting acoustic features such as Mel- frequency cepstral coefficients (MFCCs) and zero-crossing rates to instruct classifiers. The findings indicate that machine learning models had a great capacity of giving accurate voice disorder detection in clinical environments, though its accuracy largely depended on the quality and variety of the training material.

IV. METHODOLOGY

The issue of Kannada speech sound disorders and machine learning also has the methodology that consists of the efficient and comprehensive approach to utilizing the data, the development of the model, and the interaction with the user in the web-based program. It starts with the collection and pre- processing of data where about 8,650 lines of Kannada text of publicly available collections of stories are typed in to train models for machine learning. The preprocessing phase involves The procedure of the tokenizing of Kannada text into a collection of separate tokens with the Keras Tokenizer one token representing a word or character. The text is broken into overlapping sequences by the use of N-gram sequences to allow the input output pairs to be created and predictions of the subsequent tokens to be made. The sequences are filled in with zeros in advance to have uniform sequences to normalize the sequence to be trained on a model. Lastly, the target labels are converted into a categorical type, which may easily predicted by the model and classified.



The second stage is the developing of the models, and Two distinct Machine learning models will be used: Long Short-Term Memory (LSTM) networks and Generative Adversarial Networks (GANs). The LSTM model is constructed so as to predict and generate Kannada words depending on the sequences that are fed on the basis of its capability to retain long-term reliance on the sequence data. It is a sequence of embedding, LSTM, and establishes relationships between situations, and a final component, which is dense and contains a softmax activation element, to generate probabilistic prediction of a word. The model is trained with padded sequences as inputs and target words as outputs, categorical cross-entropy loss and Adam optimizer is applied to optimize the model successfully. The LSTM model possesses over 50 epochs that demonstrate a significant improvement of the prediction accuracy. It is the GAN model that generates natural Kannada text sequences instead with the assistance of the training of patterns amongst the input data. It has a

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 4, November 2025 Impact Factor: 7.67

generator that creates Kannada text based on noise vectors and a discriminator that verifies the authenticity of written text. The adversarial training is also trained to learn the generator and the discriminator whereby the former is trained to generate a more realistic text and the latter is trained to discriminate the real and the fake text better.

The system is audio integrated so as to create it more convenient. G TTS (Google Text-to-Speech) is employed to transform the suggested Kannada text to audio, to produce audio outputs of words and sentences. This also comes in quite handy with users who require auditive feedback such as in an instructional or remedial setting. Moreover, a speech recognition component is also provided, to match the pronunciations by users with the expected results, allowing them to assess the precision of the pronunciation and make the system more interactive and user-friendly.

The final aspect is the development of a web-based application to promote interaction among users. The backend consists of the component of the program that is implemented using Flask and handles the model prediction and audio synthesis queries enabling the free flow of information between the user interface and machine learning models. The front end is composed of interactive HTML pages that are based on Bootstrap that does the responsive design and JavaScript that provides the functionality of playing the audio in real-time. The users have the privilege of typing Kannada or selecting some of the Kannada letters to type and give the predictions and activate the audio playback and pronunciation validation option. This web based application is user friendly and easy to operate and therefore the system can cater to the needs of different users. It is based on this methodology that the integrated approach offers a user-centric and a powerful solution to the Kannada speech sound disorders.

V. ALGORITHM

The recurrent neural networks (RNNs) specifically of the Long Short-Term Memory (LSTM) eliminate the vanishing gradient problem common to classical RNNs, which forget long sequences due to the decaying gradient in the backpropagation stage. The answer to this issue is that LSTM has a complicated structure which involves gates that selectively store and evoke information and subsequently the capability to learn long term dependencies in sequential data. This project has applied LSTM model in the predictive next word or sequence of Kannada text input to give religion-related and coherent results. The predictions have a vital part in the process of full text, prediction and preprocessing of audio generation.

LMST network is important components and the initial component is the cell state which is the main memory that can store information with time but with minimal changes unless influenced by the gates. This architecture allows the long-dependencies to be stored in an effective manner, which is the primary weakness of traditional RNNs.

The fundamental elements of LSTM are gates that make up layers of a neural network that regulate the information flow. As stated by the forget gate, which information regarding the past state of the cell is to be forgotten using a sigmoid activation on the sum of the current input and the previous hidden state which gives an output of 0 (forget) to 1 (retain). The input is then processed enters the prior hidden and input gateprocess and a decision is made on what information will be included in the state of cells. It is founded on using sigmoid layer to ascertain the significant changes and tanh layer to produce the potential values of cell state. The last gate is the output gate which calculates the following hidden state using sigmoid function in terms of the present input and hidden state and tanh function regarding the updated cell state. And this internal state which is next fed into predictions or as input into the succeeding time step.

The LSTM algorithm starts with the disaggregated input text in form of a sequence of words or characters and, therefore, sharing of the tokens, to a numeric form, as a vector space. The forget gate recognizes the importance of the data in the previous state of a cell at each time step removing irrelevant data. In the meantime, the input gate recognizes and incorporates material patterns of the present input to the cell condition. The cell state is updated with some forget gate scaled output and input gate relevant information added to the state. The output gate then determines the hidden state using the modified cell state which is then used in generating predictions. And finally, a dense- mode is also available to propagate the hidden state, to produce the desired output, e.g. the following word in the chain or a probability distribution of the vocabulary.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, November 2025
Impact Factor: 7.67

VI. RESULT AND DISCUSSION

The project outcomes indicate the utility of the Speech Sound Disorder Detection System to apply the advanced machine learning methods to analyzing Kannada text and producing audio outputs. The system was tested in terms of various performance criteria such as accuracy, scalability, response time, and user feedback. This was demonstrated by the LSTM model which was very accurate with a prediction of Kannada text at 92 percent indicating that it can comprehend the more complicated grammatical structures of the language and remember long-term dependencies. The GAN model was also able to improve the system and produce realistic and contextually accurate synthetic text, with a human rating quality score of 4.6 out of 5. Also, the Text-to-Speech (TTS) module generated high-rating audio results with a score of 4.5 out of 5 by the users as to naturalness and clarity.

The system was also effective in scaling as it was able to achieve consistency in the execution of the system with the increase or decrease in the length of the input and also with several users simultaneously interacting with the system. It was also able to accept at most 500 characters as input and the greatest quantity of simultaneous user requests was 50 with an average response time of 1.8 seconds. The project had weaknesses, including the fact that it sometimes cannot handle long text inputs correctly because of the LSTM model and mode collapse because of the GAN model, hence giving repetitive results when the training set was not diverse. The TTS module had slight problems with the correct reproduction of tonal differences with the rare or unknown Kannada words.

VII. CONCLUSION

The Speech Sound Disorder Detection System marks a substantial breakthrough in the use of machine learning to address linguistic challenges in regional Indian languages, particularly Kannada. By integrating Networks using Long Short-Term Memory (LSTM) and Generative Adversarial Networks (GANs), the system achieves high accuracy and contextual relevance in text prediction and generation. Coupled with the Text-to-Speech (TTS) module, it delivers natural- sounding audio outputs that bridge the gap between text and speech processing. The project's layered architecture, which includes data preprocessing, machine learning models, and an interactive web interface, ensures seamless functionality and user accessibility.

The results of the project validate its effectiveness in addressing the dual challenges of text prediction and audio generation for Kannada. The capacity of the LSTM model to capture long- term dependencies, demonstrated an accuracy of 92%, while the GAN model enhanced the system’s capacity for generating diverse and realistic text sequences. The TTS module also added value to the system by producing audio outputs out of textual predictions and scored highly by the users on perceived clarity and naturalness. Such results highlight the strength of the framework to address the grammatical and phonetics peculiarities of the Kannada language.

Although the project has its strengths, the project also proved to have areas of improvement. Mode collapse in the GAN model and the problems with the persistence of the text in the LSTM model are issues that require improvement. Also, the TTS module experienced a few minor issues with the tonal variation reproduction of rare words, and it indicates the necessity of having a sophisticated phonetic modelling. The projections of these issues will make the system reliable and scalable so that it can support various user requirements.

The implications of the project are extensive, especially in such areas as education, therapy and accessibility. This system can be employed as a helpful tool to facilitate pronunciation and comprehension among consequences for children who struggle with speaking sounds. It can help students to learn Kannada in educational, and students with speech problems in a therapeutic setting, where its mispronunciation can be corrected through proper synthesis. Additionally, the system provides a platform on which more speech and text processing technologies can be applied to other Indian languages, which expands the consequences of linguistic accessibility.

REFERENCES

[1] I. Sindhu and M. S. Sainin, " Deep Learning-Based Automatic Identification of Speech and Voice Disorders—A Systematic Literature Review," in IEEE Access, vol. 12, pp. 49667-49681, 2024, doi: 10.1109/ACCESS.2024.3371713

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 4, November 2025 Impact Factor: 7.67

[2] C. D. Stylios, V. C. Georgopoulos, and G. Georgoulas, "Speech Sound Classification and Detection of Articulation Disorders with Support Vector Machines and Wavelets," 2006 IEEE Engineering in Medicine and Biology Society International Conference, New York, NY, USA,2006, pp. 2199-2202, doi: 10.1109/IEMBS.2006.259499.
[3] M. Shahin, B. Ahmed, D. V. Smith, A. Duenser and J. Epps, " "Automatic Identification of Children With Speech Sound Impairments Through Paralinguistic Features," IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP), 2019 Pittsburgh, PA, USA, 2019, pp. 1-5, doi: 10.1109/MLSP.2019.8918725.
[4] C. -H. Hsiao, S. -J. Ruan, C. -L. Chen, Y. -W. Tu, Y. -C. Chen and G. M. Rahmatullah, "A Text-Dependent End-to-End Speech Sound Disorder Detection and Diagnosis in Mandarin- Speaking Children," in IEEE Transactions on Instrumentation and Measurement, vol. 73, pp. 1-11, 2024, Art no. 3001911, doi: 10.1109/TIM.2024.3438853.
[5] S. -I. Ng, C. W. -Y. Ng, J. Wang and T.Lee, " Automatic Identification of Disorders in Speech Sounds in Cantonese- Speaking Pre-School Children," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 32, pp. 4355- 4368, 2024, doi: 10.1109/TASLP.2024.3463503.
[6] Automatic Identification of Disorders in Speech Sounds in Child Speech Using Posterior-based Speaker Representations DOI:10.48550/arXiv.2203.15405
[7] G. Georgoulas, V. C. Georgopoulos and C. D. Stylios, " Speech Sound Categorization and Articulation Disorder Identification Using Support Vector Machines and Wavelets," IEEE Engineering in Medicine and Biology Society International Conference, 2006, New York, NY, USA, 2006, pp. 2199-2202, doi: 10.1109/IEMBS.2006.259499
[8] Dudy S, Asgari M, Kain A. Children with speech sound abnormalities can benefit from pronunciation analysis. The annual IEEE Eng Med Biol Soc conference. 2015 Aug;2015:5573-6. doi: 10.1109/EMBC.2015.7319655. PMID: 26737555; PMCID: PMC4710861.
[9] S. Dudy, M. Asgari, and A. Kain, "Pronunciation analysis for children with speech sound disorders," IEEE Engineering in Medicine and Biology Society (EMBC), 37th Annual International Conference, 2015. Milan, Italy, 2015, pp. 5573-5576, doi: 10.1109/EMBC.2015.7319655.
[10] Machine learning techniques for the detection of Mujeeb Ur Rehman's voice disorder:An application in speech and language patholog