IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.13349

STUDS, Speech Therapy Utility for Detection and Analysis of Stuttering

Hemanth RangeGowda S P[1], M Chinmaya Rao[2], Nishanth S Raj[3], Rakshitha Jain[4]

Mr. Amruth Ashok Gadag[5], Mr. Sunil Kumar S[6], Dr. Rakesh C V[7], Dr. Shubhaganga D[8],

Dr. Santosh M[9]

Student, Dept. of AI&ML, Mangalore Institute of Technology & Engineering, Karnataka, India[1-4]

Professor, Dept. of AI&ML, Mangalore Institute of Technology & Engineering, Karnataka, India[5,6]

Professor, Dept. of Speech and Hearing, MCHP, Manipal, Karnataka, India[7,8]

Professor, AIISH, Mysore, Karnataka, India[9]

Abstract: Abstract: Stuttering, a complex speech disorder, presents significant challenges in both diagnosis and treatment. In this study, we propose a novel approach for predicting stuttering severity in Kannada speech, aimed at enhancing therapeutic interventions for individuals affected by stuttering. Leveraging a dataset comprising video recordings of therapy sessions, our methodology involves the extraction of acoustic features from 3-second audio segments, including mean pitch, intensity, speech rate, and MFCCs. These features, along with annotations for disfluency types such as prolongation, repetition, and block, form the basis of a comprehensive dataset. Through the application of a CNN-LSTM hybrid neural network, we demonstrate promising results in predicting stuttering severity, with implications for personalized therapy strategies. Our research underscores the potential of machine learning techniques in improving the diagnosis and treatment of stuttering, paving the way for more effective interventions and improved outcomes for individuals with this speech disorder.

Keywords: MFCCs, CNN-LSTM, Kannada speech, stuttering.

I. INTRODUCTION

Stuttering, a challenging speech disorder characterized by disruptions in fluency and rhythm, presents significant obstacles for individuals and clinicians alike. Assessing stuttering severity remains a complex task, often reliant on subjective assessments. This paper proposes a novel approach for predicting stuttering severity in Kannada speech, leveraging machine learning and acoustic analysis techniques. By providing clinicians with an objective tool, our research aims to enhance treatment planning and outcomes.

Our methodology utilizes a dataset of therapy session recordings, capturing nuances of stuttering behavior through annotation and feature extraction. A hybrid CNN-LSTM architecture integrates advanced machine learning algorithms to enhance predictive capabilities. Implications of this research extend to clinical practice, offering tangible benefits for treatment planning and patient care. Subsequent sections detail our methodology, present experimental findings, and discuss broader implications for the field of speech pathology. Through our endeavors, we contribute to advancing stuttering research and developing innovative solutions to improve the lives of affected individuals.

II. LITERATURE SURVEY

The study on automatic recognition of stuttering in speech [1] compared event-based and interval-based segmentation methods, revealing that event-based segmentation yielded superior performance in recognition systems. It emphasized the impact of linguistic features, highlighted the need for thresholds in clinical applications, and discussed challenges in achieving fully automated stuttering severity assessment. The study recommended event-based procedures for improved machine learning model capabilities, despite limitations like manual labeling requirements and class imbalance issues.

The Stutter Diagnosis and Therapy System based on deep learning [2] focused on automating tasks in speech-language pathology. It included a Stutter Assessment for analyzing stuttering and Therapy Suggestions for recommending therapies based on patient progress.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024
DOI: 10.17148/IJARCCE.2024.13349

Using MFCC for feature extraction, GRCNN for diagnosis, and SVM for therapy recommendations, the system achieved high validation accuracies for identifying prolongation and repetition types of stuttering. Challenges included dataset structure and background noise impact, with plans for broader stuttering dysfluencies identification and therapy integration.

Advancements in stuttering detection [3] involved addressing class imbalance, data scarcity, and multi-contextual architectures. Introducing StutterNet with ResNet+BiLSTM and ConvLSTM models, the study proposed strategies like weighted cross entropy and data augmentation to enhance detection performance. Challenges included domain-specific data augmentation and the need for explainability analysis. The study highlighted promising results, suggesting further research to develop clinically usable stutter detection systems.

A systematic review of machine learning approaches for developmental stuttering detection [4] emphasized transparent reporting for improved comparability across studies. It discussed challenges in data reporting and model architecture comparisons, advocating for clear documentation of dataset shapes and features. The review called for meta-analytic techniques to bridge knowledge gaps and enhance understanding of stutter classification accuracy. It provided insights into current ML models for stuttering detection and suggested areas for future research improvements.

StutterNet, a deep learning-based stuttering detection system utilizing TDNN [5], excelled in capturing temporal and contextual disfluency aspects in speech. Outperforming existing methods, StutterNet addressed various stuttering types as a multi-class classification problem, significantly reducing parameters through TDNN parameter sharing. Trained with PyTorch in Python, StutterNet demonstrated significant improvements over ResNet+BiLSTM in accuracy and MCC. Future work includes real-world scenario evaluations and exploration of advanced TDNN variants for continued enhancement.

The SEP-28k dataset comprises over 28,000 annotated speech clips from public podcasts featuring people who stutter [6], aiming to enhance dysfluency detection in speech. It addresses clinical assessment needs and improves speech recognition technology accessibility. Benchmarking against Fluency Bank showed significant performance gains with increased training data, discussing challenges in dysfluency detection, dataset curation, Fluency Bank comparisons, annotation difficulties, study methods, model analysis, and evaluation metrics. Emphasizing the dataset's importance, it encourages further exploration and extends the focus to dysfluencies in individuals with other speech disorders.

The identification of primary and collateral tracks in stuttered speech [7] introduces a novel evaluation framework for disfluency detection, blending clinical and NLP perspectives. Using a new dataset from semi-directed interviews, it contrasts text-based and acoustic-prosodic features, suggesting new audio features. Highlighted challenges include limited pathological annotated datasets and a lack of clear evaluation protocols, stressing collaborative interests across research domains. The document presents schemes for identifying primary and collateral tracks, emphasizing precision, recall, F1-score, error rate metrics, and introducing new Audio Span Features for improved detection and identification.

Detecting stuttering events in orthographic transcripts of children's speech [8], utilizing machine learning approaches for automated diagnosis. Comparing HELM and CRF models, CRF outperformed by 2.2% in baseline experiments, with data augmentation enhancing performance, especially for rare events. By expanding human-transcribed stuttering speech data, the study addressed training data limitations and high dimensionality challenges. Annotating stuttering types in transcripts, the study incorporated UCLASS data, added human transcriptions, and explored data normalization, feature extraction, and data augmentation effects for classifier performance improvement.

Automatic Speech Recognition [9] provides an in-depth exploration of ASR fundamentals, focusing on LVCSR and HMM's role in speech recognition. It covers HMM architecture, feature extraction, acoustic models, Viterbi decoding, training procedures, application areas, evaluation metrics, historical context, and Gaussian models' impact. It discusses ASR's evolution, application domains, and challenges, emphasizing word error rate evaluation and ARPA-funded programs' influence on ASR development.

III. NOVELTY OF THE PROPOSED WORK

The proposed work presents novel contributions in the field of stuttering severity prediction specifically for Kannada speakers. The integrated methodology combines audio preprocessing, acoustic feature extraction, and manual annotation, offering an analysis of stuttering behavior. The utilization of chunk segmentation and manual annotation allows the analysis of disfluency types within individual audio segments, enabling the examination of stuttering patterns.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.13349

The implementation of a CNN-LSTM hybrid neural network further enhances prediction by capturing both temporal dependencies and speech features representing an innovative model architecture. The clinical impact of providing therapists with objective measures of stuttering severity based on audio recordings facilitates assessment and personalized treatment planning, ultimately improving the speech for Kannada-speaking patients.

IV. METHODOLOGY

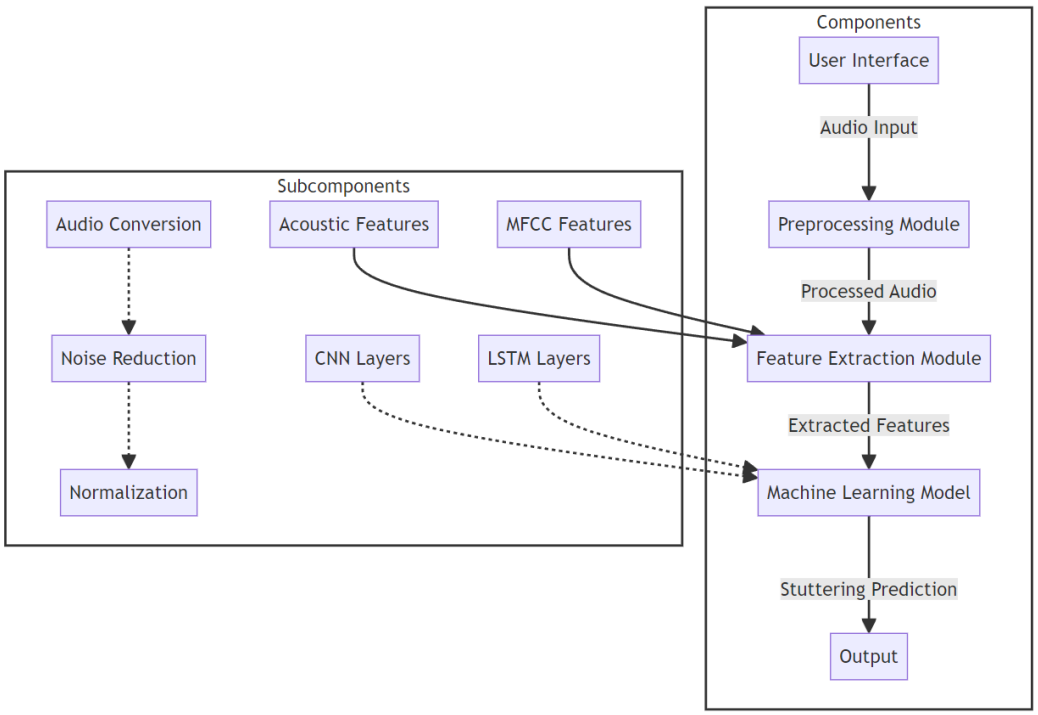

Fig 1.1 System architecture

The methodology revolves around the systematic approach employed to develop and evaluate a predictive model for stuttering severity prediction in Kannada-speaking patients. The process begins with dataset acquisition, wherein video recordings of therapeutic sessions are collected from a reliable source under ethical guidelines, ensuring patient confidentiality and consent. These recordings are then processed to extract audio files and segmented into fixed-duration chunks, facilitating uniform feature extraction and model training. Following this, preprocessing techniques are applied to enhance the quality of the audio data, including normalization and noise reduction. Acoustic features and Mel-Frequency Cepstral Coefficients (MFCCs) are extracted to capture various aspects of speech production and spectral characteristics. Manual annotation of the dataset is conducted to label stuttering severity and provide contextual information.

The model architecture comprises a novel CNN-LSTM hybrid neural network designed to capture spatial and temporal dependencies in the data. The CNN layers extract spatial features from temporal representations, while LSTM layers capture temporal dependencies in sequential data. The annotated dataset is then split into training, validation, and test sets using stratified sampling to ensure a balanced distribution of stuttering severity labels. The model is trained using the training data with the Adam optimizer and binary cross-entropy loss function.

Hyperparameters such as batch size, learning rate, and number of epochs are optimized through grid search and cross-validation techniques. Evaluation metrics including accuracy, precision, recall, F1 score, and AUC-ROC are utilized to assess model performance. Comparisons with baseline models highlight the superiority of the proposed CNN-LSTM hybrid model in predicting stuttering severity. Ethical considerations such as data privacy, consent, and bias mitigation strategies are strictly adhered to throughout the study. Efforts are made to mitigate potential biases in the dataset through careful sampling strategies and data preprocessing techniques. The methodology ensures a systematic and comprehensive approach to developing and evaluating the predictive model, as depicted in the accompanying diagram.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024
DOI: 10.17148/IJARCCE.2024.13349

V. RESULTS

The results of the models built are utilized for calculating the percentage of stuttering, which aids in assessing the severity of stuttering and guiding therapeutic interventions. The outcome of the model predictions is used to determine the presence and type of stuttering disfluencies in the audio data. The percentage of stuttering is calculated by aggregating the predictions over the entire dataset. This percentage is further broken down into different types of stuttering, including prolongation, repetition, and block, providing a comprehensive understanding of the stuttering profile. Graphical representations of the percentage of each stuttering type are provided to visually depict the distribution and severity of stuttering across the dataset. These graphical representations serve as valuable tools for therapists in evaluating the stuttering severity and tailoring treatment strategies accordingly, ensuring targeted interventions for individuals with varying degrees of stuttering severity. The calculation methodology involves dividing the total number of stuttered speech segments by the total number of speech segments in the dataset and multiplying by 100 to obtain the percentage of stuttering. This approach enables objective assessment and monitoring of stuttering severity, facilitating informed decision-making in clinical practice

VI. CONCLUSION

In conclusion, the development of a Kannada stuttering severity prediction tool represents a significant advancement in the field of speech therapy. By leveraging advanced machine learning techniques and language-specific approaches, the tool offers therapists a valuable resource for objective assessment and monitoring of stuttering severity in Kannada-speaking individuals.

Through integrating audio preprocessing, feature extraction, and innovative model architecture, our research demonstrates the feasibility and effectiveness of using computational methods to augment traditional stuttering assessment practices. The tool's potential to improve treatment planning and intervention strategies underscores its importance in enhancing the quality of care for individuals with stuttering disorders. Looking ahead, continued research and development efforts will focus on further refining the prediction model, expanding its linguistic scope, and exploring avenues for real-world implementation. By addressing these future challenges, we aim to advance the field of stuttering assessment and contribute to the improvement of therapeutic outcomes for individuals with communication disorders. This concluding statement emphasizes the significance of your research findings, acknowledges future directions for exploration and development, and underscores the broader impact of your work in the field of speech therapy

REFERENCES

[1] Sheikh, S.A., Shahidullah, M., Hirsch, F. and Ouni, S., 2023. Advancing stuttering detection via data augmentation, class-balanced loss and multi-contextual deep learning. IEEE Journal of Biomedical and Health Informatics.
[2] Reinolds, F., Neto, C. and Machado, J., 2022. Deep learning for activity recognition using audio and video. Electronics, 11(5), p.782.
[3] Sheikh, S.A., Sahidullah, M., Hirsch, F. and Ouni, S., 2021, August. Stutternet: Stuttering detection using time delay neural network. In 2021 29th European Signal Processing Conference (EUSIPCO) (pp. 426-430). IEEE.
[4] Prabhu, Y. and Seliya, N., 2022, December. A CNN-Based Automated Stuttering Identification System. In 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 1601-1605). IEEE.
[5] Beleza, S.R.A., Shimomoto, E.K., Souza, L.S. and Fukui, K., 2023. Slow feature subspace: A video representation based on slow feature analysis for action recognition. Machine Learning with Applications, 14, p.100493.
[6] Kourkounakis, T., Hajavi, A. and Etemad, A., 2021. Fluentnet: End-to-end detection of stuttered speech disfluencies with deep learning. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 29, pp.2986-2999.
[7] Sugamura, N. and Itakura, F., 1986. Speech analysis and synthesis methods developed at ECL in NTT—From LPC to LSP—. Speech communication, 5(2), pp.199-215.
[8] Barrett, L., Hu, J. and Howell, P., 2022. Systematic review of machine learning approaches for detecting developmental stuttering. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 30, pp.1160-1172.
[9] Choi, K., Luo, Y. and Hwang, J.N., 2001. Hidden Markov model inversion for audio-to-visual conversion in an MPEG-4 facial animation system. Journal of VLSI signal processing systems for signal, image and video technology, 29, pp.51-61.
[10] Bose, A. and Tripathy, B.K., 2020. Deep learning for audio signal classification. Deep learning research and applications, pp.105-136.

