International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 07 | July - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

A Comprehensive Approach for Classification of Stuttering in Children

G. Tripthi[1], Ch. Monica[1]

Department of CSE, G. Narayanamma Institute of Technology and Science (for Women)

[1]*tripthi0203@gmail.com*, [1]*chmonica2002@gmail.com*

Abstract — Stuttering, a developmental speech disorder, poses significant challenges for children, impacting their fluency and interrupting the natural flow of speech. These disruptions have far-reaching consequences, particularly in communication, social interaction, and emotional well-being. Early intervention plays a crucial role in supporting optimal child development. This paper explores the potential of Machine Learning and Deep Learning techniques in the context of stuttering identification. By analyzing speech patterns and identifying specific characteristics associated with stuttering, these advanced computational methods offer valuable insights. Leveraging the power of these techniques, researchers and clinicians can enhance their understanding of stuttering, facilitate early detection, and develop effective interventions to support children affected by this speech disorder. This research paper aims to highlight the significance of applying Machine Learning and Deep Learning techniques as a promising approach for early intervention and improved outcomes in the field of stuttering identification and treatment in children.

Keywords: *Stuttering, Developmental Speech Disorder, Children, Telugu Language, Annotation, Stutter codes, Machine Learning, Deep Learning, SMOTE*

1. INTRODUCTION

Stuttering is a complicated condition that disrupts the flow and rhythm of speech, resulting in repetition, prolongation, and interruption of sounds, syllables, and words. The disorder can negatively impact a child's life, leading to frustration, anxiety, and social isolation due to difficulties in expressing themselves. Additionally, stuttering can impact academic performance and trigger emotional and behavioral issues such as low self-esteem, depression, and avoidance of social situations. Early detection and timely intervention are critical for improving long-term outcomes for children with stuttering. With prompt diagnosis and treatment, the severity and frequency of stuttering can be reduced, communication skills can be improved, and overall quality of life can be enhanced for individuals affected by the condition.

Machine learning and Deep Learning has emerged as a powerful tool in healthcare for its ability to analyze large datasets and identify patterns that are not easily detectable by human analysis. In recent years, it has been applied to various areas of healthcare, including the identification and diagnosis of speech and language disorders. This paper focuses on the application of Machine Learning algorithms and a few Deep Learning in stuttering identification among children. By leveraging the power of these technologies, we aim to develop an automated and accurate system that can identify stuttering in children at an early stage. This can lead to improved clinical outcomes, better treatment planning, and enhanced quality of life for affected children and their families.

The application of the algorithms in stuttering identification and diagnosis has the potential to improve outcomes and quality of life for children with stuttering, as early intervention has been shown to be associated with better treatment outcomes. Machine learning has the potential to enhance the diagnosis and treatment of stuttering by developing personalized treatment plans that meet the specific needs of each individual. This approach can lead to more efficient and effective treatment outcomes. Additionally, machine learning can help improve access to care and alleviate the burden on healthcare systems by reducing the time and resources needed for diagnosis and treatment. This study underscores the value of machine learning in enhancing the diagnosis and treatment of stuttering, which can positively impact the lives of those affected by the condition and their families.

The dataset used comprises speech samples from both children with and without stuttering, which are preprocessed and used to extract relevant features. We then employ supervised and unsupervised learning techniques to train several machine learning models by preprocessing the data. These models are trained to identify patterns and classify speech samples as either stuttered or non-stuttered. By using different models, we can compare their performances and select the most accurate and efficient one. The problem of imbalanced dataset in normal machine learning models is overcomed by using the SMOTE technique. The trained models can then be used to develop an automated system that can accurately identify stuttering in children, which can aid clinicians in

making more accurate diagnoses and improving treatment planning.

The rest of the paper discusses the related previous works in Section II, methodology followed in Section III, experiments that include feature extraction and modeling of data collected in Section IV, and presents the Results in Section V. Finally, the paper also discusses the conclusion and future scope of the presented work.

2. RELATED WORK

Shakeel A. Sheikh reviewed acoustic features, and statistical and deep learning-based stuttering/disfluency classification methods in paper [20]. They also presented several challenges and possible future directions. This paper provides an overview of the stuttering problem, discusses stuttering from acoustic and neurological perspectives, reviews historical works, reviews statistical and deep learning methods, and performs a benchmark analysis of state-of-the-art deep learning methods. **Abedal-Kareem Al-Banna**, [2] rigorously investigated the effective use of eight well-known machine learning classifiers, on two publicly available datasets (FluencyBank and SEP-28k[5]) to automatically detect stuttering disfluency using multiple objective metrics, i.e. prediction accuracy, recall, precision, F1-score, and AUC measures. The experimental results on the two datasets show that Random Forest classifier achieves the best performance, with an accuracy of 50.3% and 50.35%, a recall of 50% and 42%, a precision of 42% and 46%, and an F1 score of 42% and 34%, against the FluencyBank and SEP-28K datasets, respectively. The machine learning-based approaches may not be effective in accurate stuttering disfluency evaluation, due to diverse variations in speech rate, and differences in vocal tracts between children and adults. The use of deep learning approaches and Automatic Speech Recognition (ASR) with language models may improve outcomes, specifically for large scale and imbalanced datasets.

Abedal-Kareem Al-Banna, Eran Edirisinghe, Hui Fang, [1] proposed a new model for stuttering events detection that may help (SLP) to evaluate stuttering severity. The model is based on a log mel spectrogram and 2D atrous convolutional network designed to learn spectral and temporal features. The performance of the model is rigorously evaluated using two stuttering datasets UCLASS[15] and FluencyBank[3] using common speech metrics, i.e. F1-score, recall, and the area under the curve (AUC). The model outperforms state-of-the-art methods in prolongation with an F1 of 52% and 44.5% on the UCLASS and FluencyBank datasets, respectively. Also, 5% and 3% margins are gained on the UCLASS and FluencyBank datasets for fluent class.

The paper [11] is a systematic review of the literature on statistical and machine learning schemes for identifying symptoms of developmental stuttering from audio recordings. Twenty-seven papers met the quality standards that were set by **L. Barrett, J. Hu and P. Howell**. Comparison of results across studies was not possible because training and testing data, model architecture and feature inputs varied across studies. The limitations that were identified for comparison across studies included: no indication of application for the work, data were selected for training and testing models in ways that could lead to biases, studies used different datasets and attempted to locate different symptom types, feature inputs were reported in different ways and there was no standard way of reporting performance statistics. Recommendations were made about how these problems can be addressed in future work on this topic.

The existing systems for stuttering identification using machine learning algorithms have shown promising results in accurately detecting and classifying stuttered speech. However, further research is needed to address challenges such as data variability, and potential biases in training data. As machine learning techniques continue to evolve, it is likely that future stuttering identification systems will be even more effective in diagnosing and treating stuttering, ultimately improving the quality of life for those affected by this speech disorder. It should be noted that the majority of existing systems for stuttering identification using machine learning algorithms have been developed and evaluated solely on the English language, with limited research on regional languages. This highlights a crucial area for improvement, and the proposed system aims to address these gaps by incorporating a wider range of languages to increase the accessibility and effectiveness of stuttering identification and treatment for diverse populations.

3. METHODOLOGY

This section of the paper describes the dataset used in the study and the machine learning algorithms applied to the dataset in order to evaluate metrics that are used to assess the performance of the machine learning model. It also provides a summary of the key elements about how the machine learning model is built. The section highlights the detailed information on the dataset used, the machine learning algorithms that are employed, and the evaluation metrics used to measure the performance of these models. In summary, this section gives a quick overview of what to be expected in building the machine learning models.

***Table 1**: Stutter Code Description*

Stutter Code	Description
0	Clean - No Stuttering
1	Filled Pause/Interjection
2	Mono Syllable Repetition
3	Part word repetition
4	Multi Syllable Repetition
5	Unfilled Pause
6	Prolongation
7	Block Stuttering

The dataset TLD-ISC is used which consists of audio samples of 60 children that were collected by visiting the rural schools in Hyderabad. The data with 60 recordings was collected from both stuttering and non-stuttering children. The environment in which data collected was calm with no noise and had few noise absorbers around so as to avoid the problem of reverberation. These recordings were stored with a standard sampling rate of 44.1kHz and 16-bit resolution.

The audio files from the dataset were then pre-processed before building the Machine Learning model. These audio samples were first divided into syllables using the Knowledge Nest for Speech (KNS) notation[9]. The start time and end time of each syllable were annotated.. The start time and the end time of the corresponding syllable were noted in a workbook along with their respective standard stuttered code as shown in Table 1. This data is later used in building the machine learning model to create an output label. This label had been assigned a value of either 0 or 1 based on the stutter code of each syllable. The input label was created by extracting the MFCC features of the audio sample which was a two dimensional array. The data is now completely pre-processed and ready to be fed to a Machine Learning model.

The Machine Learning algorithms used for identification of stuttering in children are Decision Trees (DT), Random Forest (RF). The models were trained with the above algorithms by importing the scikit-learn library. Each model was built using different split ratios into training and testing data – 70:30, 60:40 and 80:30. In each of these models different hyperparameters were experimented to find the best combination and their final results were also compared.

The Deep Learning models used – Convolutional Neural Network (CNN) and Artificial Neural Network (ANN) models. The CNN Sequential model and ANN Multi-Layer Perceptron was built under Deep Learning techniques. During the pre-processing stage, the audio samples were labeled, and these labeled samples served as the input to the models. The models are designed to output whether the speech in the given audio sample contains stuttering or not.

Finally, the performance of these Machine Learning was evaluated using different metrics. These metrics include – Accuracy that determines the percentage of correctly classified instances, Precision that gives percentage of instances classified as stuttering that are actually stuttering, Recall that notes the percentage of actual stuttering instances that are classified as stuttering and F1 score which is the harmonic mean of precision and recall. These metrics were used in comparing the four different algorithms used and also in determining the best algorithm for identification of stuttering in children.

4. EXPERIMENTS

The experiments conducted in this study aimed to evaluate the performance of the proposed stuttering identification system using machine learning algorithms. Through a series of experiments, we assessed the accuracy, precision, recall, and F1 score of the system in detecting and classifying stuttered speech.

A. Feature Extraction

The objective of feature extraction in stuttering identification is to simplify the speech signal by using a predefined number of signal components. This is because the entire information in the speech signal is often impractical to manage, and some of the information may not be relevant for the identification task. Feature extraction results in the improved performance of the model. Among the various techniques available for feature extraction from audio signals, MFCC is widely utilized due to its effectiveness and popularity. MFCC features[4] are particularly suited for datasets containing speech samples of children, as their voices tend to have lower pitch compared to adults. MFCC features are extracted with 39 coefficients for each audio frame. These 39 MFCC coefficients are extracted using the librosa library in python. Extraction of MFCC features for each audio frame resulted in a 2D array which can be fed directly to the Machine

learning models. The below Fig 1 shows a two line code in extracting MFCC features by importing the librosa library.

```
x, sr = librosa.load(audio_path)
mfccs = librosa.feature.mfcc(x, sr=sr,n_mfcc=39)
```

***Fig 1**: MFCC feature extraction*

B. Detection Models

Using the MFCC features extracted for each audio frame, different Machine Learning algorithms such as Decision Trees, Random Forest were trained for the identification of stuttered speech. Random Forest and Decision Tree algorithms were evaluated at different depths.

In addition to Machine learning algorithms, Deep learning models were also employed to train the stuttering identification system. Specifically, two types of deep learning models were explored - Sequential in Convolutional Neural Networks (CNN) and Multi-Layer Perceptron (MLP) in Artificial Neural Networks (ANN).

To assess the performance of these models, various test-train split ratios were also explored. An important point to consider while building machine learning models is that the dataset was imbalanced because of the fewer samples of stuttered speech and more samples of normal speech. To balance the dataset, a technique called SMOTE is used.

Synthetic Minority Over- Sampling Technique (SMOTE):

SMOTE (Synthetic Minority Over-sampling Technique)[14] is a popular data augmentation technique used in machine learning to address class imbalance problems. Class imbalance occurs when no. of instances in one class are lower than the number of instances in another class. This can lead to poor performance of machine learning models because the model tends to be biased towards the majority class. SMOTE works by generating synthetic samples by interpolation of the minority class such that the no.of samples in both majority and minority classes are equal using k nearest neighbors approach.

SMOTE has several advantages over other techniques for addressing class imbalance. It is easy to implement, computationally efficient and can improve the performance of machine learning models. By using this technique, the issue of imbalanced datasets was effectively addressed, resulting in better identification performance. The details and outcomes of these steps are elaborated upon in the subsequent sections.

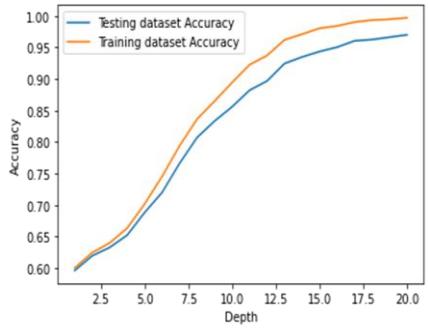
5. RESULTS

This section presents the outcomes of the experiments conducted to evaluate the performance of the proposed stuttering identification system using machine learning algorithms, including the accuracy, precision, recall, and F1 score of the system for detecting and classifying stuttered speech. A comparative analysis is made between Decision Tree and Random Forest models.

I. Decision Tree and Random Forest Algorithms

In the comparison between Decision Tree and Random Forest algorithms, various depths and criteria were evaluated. It was observed that when using entropy as the criterion, both Decision Tree and Random Forest algorithms achieved the highest accuracy compared to using the Gini criterion. This indicates that the entropy criterion was more effective in accurately predicting outcomes in both cases. The final result is considered by taking the max depth as the value which showed highest accuracy in the graphs below, fig 2, and fig. 3 for the split ratio 70:30 in Decision Tree and Random Forest respectively.

***Fig 2:** Performance of Random Forest model with respect to accuracy and max_depth*



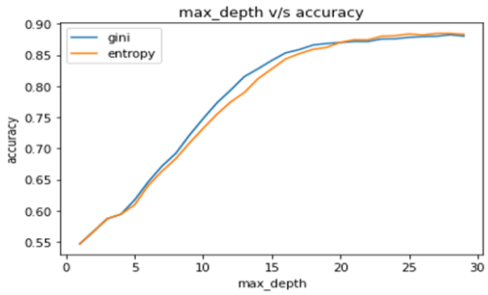


***Fig 3:** Performance of Decision Tree model with respect to Accuracy and max_depth*

II. Results for Different Models in terms of Split Ratios

The below Table 2 shows the comparative results of Decision Trees and Random Forest with different split ratios. The Decision Trees and Random Forest algorithms were implemented using gini index with respect to different split ratios. The evaluation metrics accuracy, precision, recall and f1-score were compared in the same context.

Table 2: *Comparative analysis of the Machine Learning models with different split ratios*

Split Ratio/Models	*Parameters*	*Decision Tree*	*Random Forest*
60-40 Split Ratio	**Accuracy**	96.36	93.36
	Precision	0.54	0.92
	Recall	0.02	0.04
	F1-Score	0.01	0.07
70-30 Split Ratio	**Accuracy**	93.24	93.53
	Precision	0.12	0.93
	Recall	0.003	0.02
	F1-Score	0.048	0.039
80-20 Split Ratio	**Accuracy**	92.91	93.15
	Precision	0.56	1
	Recall	0.03	0.04
	F1-Score	0.056	0.077

III. Results Comparison with respect to SMOTE

From the above cases an observation can be made that always the standard split ratio 70:30 produced the better results. The SMOTE technique was used in implementing all the above four algorithms only using the split ratio of 70:30. The results have shown better results in SVM, Random Forest and KNN but the performance was degraded in Decision Trees which is presented in the Table 3.

Table 3: *Overall comparison analysis of Machine Learning models*

Model	**Accuracy before SMOTE**	**Accuracy after SMOTE**
Random Forest	93.63	97.01
Decision Trees	93.36	88.32

The below Fig 4 is a graphical representation of accuracies for the two models before and applying the SMOTE. The performance of the Random Forest model was improved when compared to the Decision Trees algorithm. The results demonstrate that the proposed system, with the implementation of SMOTE, achieved high performance in identifying stuttered speech.

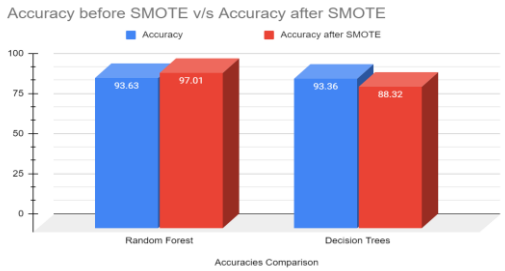


Fig 4: *Comparison of Accuracies for the Decision Tree and Random Forest models with respect to SMOTE technique*

IV. Comparison of Deep Learning Models

The below Table 3 represents the comparative results of Sequential models and Multi Layer Perceptron models. Two Sequential models were implemented differing the number of hidden layers. One Sequential model is built with 1 hidden layer and the other is implemented with 11 hidden layers. These models have shown less performance when compared to the Machine Learning algorithms, the reason is to be discussed in the next section. The Fig 5, is a graphical comparison of Deep Learning model and its accuracies.

Table 4: *Comparison of Deep Learning Models*

Model	Accuracy
Multi Layer Perceptron	92.93
Sequential Model (Single Layer)	90.82
Sequential Model (11 Layers)	77.29

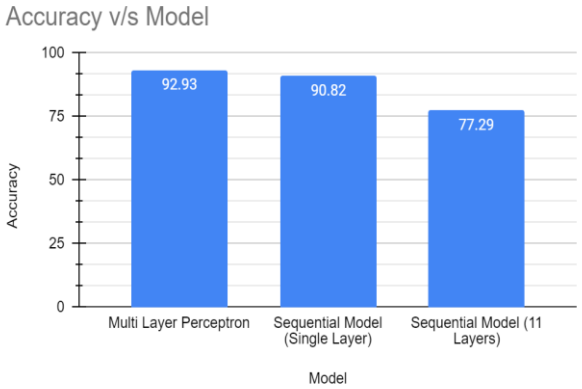


Fig 5*: Comparison of Deep Learning Model accuracies*

The results demonstrate that the proposed system, utilizing both Machine Learning and Deep Learning algorithms, achieved high accuracy in detecting and classifying stuttered speech. By leveraging SMOTE technique, the system was able to effectively address imbalanced datasets.

6. DISCUSSIONS AND CONCLUSIONS

This paper has presented and attempted to classify a dataset for identification of stuttering in children for Telugu language. The creation of a dataset for stuttering identification in children is an important step in advancing the field of speech therapy and improving the diagnosis and treatment of stuttering in young patients. A well-designed and annotated dataset can provide valuable insights into the characteristics of stuttering in children and help researchers and clinicians develop new and more effective methods for addressing this speech disorder.

By automating the process of identifying stuttering using Machine Learning and Deep Learning algorithms, clinicians can save time and resources legally and ethically. Additionally, this work opens up the possibility of developing mobile applications that can help parents and care givers monitor their children's speech and identify stuttering in real-time, enabling them to seek prompt intervention.

In this paper, the potential of Machine Learning and also Deep Learning algorithms for detecting stuttering in children is investigated . The study used a dataset consisting of audio recordings of 60 samples, with a mix of stuttered and non-stuttered speech samples. The ML and DL algorithms used were Decision Tree, Random Forest, CNN and ANN-MLP. The problem of imbalanced dataset was also addressed to improve the overall accuracy using SMOTE technique. The results show that the Random Forest algorithm achieved the highest accuracy of 97.01% in identifying stuttering. These results significantly suggest that Machine Learning algorithms have the potential to serve as a reliable and efficient tool for the early identification of stuttering in children.

To summarize, the study highlighted the effectiveness of Machine Learning algorithms in accurately identifying stuttering in children through audio recordings. The impressive accuracy achieved by the Random Forest algorithm indicates that Machine Learning can be a dependable and efficient means of early detection and intervention. These findings are particularly relevant for clinicians, parents, and caregivers, as automating stuttering identification can streamline the process and enable timely intervention.

Furthermore, this work can be extended to investigate the potential of Machine Learning algorithms in identifying other speech disorders and developing mobile applications for speech monitoring and intervention. Overall, Machine Learning and Deep Learning algorithms have the potential to revolutionize the field of speech pathology and improve outcomes for children with speech disorders.

REFERENCES

[1] Abedal-Kareem Al-Banna, Eran Edirisinghe, Hui Fang, “Stuttering Detection Using Atrous Convolutional Neural Networks”, DOI: 10.1109/ICICS55353.2022.9811183
[2] Abedal-Kareem Al-Banna, “Stuttering Disfluency Detection Using Machine Learning Approaches”, 2022, Vol.21, No. 02, 2250020
[3] Amirhossein Hajavi, Tedd Kourkounakis and Ali Etemad , “FluentNet: End-to-End Detection of Speech Disfluency with Deep Learning” , *ArXiv* abs/2009.11394 (2020)
[4] Anil Kumar Vuppala, “Towards a Database for Detection of Multiple Speech Disfluencies in Indian English”,2021, National Conference on Communications (NCC), Kanpur, India, 2021, pp. 1-6, doi: 10.1109/NCC52529.2021.9530043.
[5] Colin Lea, “Sep – 28K : A Dataset for Stuttering Event detection from podcasts with people who stutter”, 2021, 349583752

[6] E. Shriberg, “To’errrr’is human: ecology and acoustics of speech disfluencies,” Journal of the International Phonetic Association, pp. 153–169, 2001
[7] G. Lemaˆıtre, F. Nogueira, and C. K. Aridas, “Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning,” The Journal of Machine Learning Research, vol. 18, no. 1, pp. 559–563, 2017.
[8] J. Proenc¸a, D. Celorico, S. Candeias, C. Lopes, and F. Perdigao, “Chil- ˜ dren’s reading aloud performance: A database and automatic detection of disfluencies,” in Sixteenth Annual Conference of the International Speech Communication Association, 2015.
[9] Kalyani Nara, K. V. N Sunitha, “Syllable Analysis to Build a Dictation System in Telugu language”, 2010, CoRR. abs/1001.2263.
[10] L. Barrett, J. Hu and P. Howell, "Systematic Review of Machine Learning Approaches for Detecting Developmental Stuttering," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 1160-1172, 2022, doi: 10.1109/TASLP.2022.3155295.
[11] M. Kaushik, M. Trinkle, and A. Hashemi-Sakhtsari, “Automatic detection and removal of disfluencies from spontaneous speech,” in Proceedings of the Australasian International Conference on Speech Science and Technology (SST), vol. 70, 2010.
[12] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “Smote: synthetic minority over-sampling technique,” Journal of artificial intelligence research, vol. 16, pp. 321–357, 2002
[13] Naomi Eichorn, “Assessment of Stuttering Disorders in Children and Adults ® Chapter in A Guide to Clinical Assessment and Professional Report Writing in Speech-Language Pathology”, 2012
[14] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, W. Philip Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique”, (2002) 321–357, 2002
[15] Peter Howell, Stephen Davis, and Jon Bartrip “The UCLASS archive of stuttered speech”,2010
[16] S. Garg, U. Mehrotra, G. Krishna and A. K. Vuppala, "Towards a Database For Detection of Multiple Speech Disfluencies in Indian English," *2021 National Conference on Communications (NCC)*, Kanpur, India, 2021, pp. 1-6, doi: 10.1109/NCC52529.2021.9530043.
[17] S. Oue, R. Marxer, and F. Rudzicz, “Automatic dysfluency detection in dysarthric speech using deep belief networks,” in Proceedings of SLPAT 2015: 6th Workshop on Speech and Language Processing for Assistive Technologies, 2015, pp. 60–64.
[18] S. R. Maskey, Y. Gao, and B. Zhou, “Disfluency detection for a speechto-speech translation system using phrase-level machine translation with weighted finite state transducers,” Dec. 28 2010, uS Patent 7,860,719.
[19] Sebastian P. Bayerl, “KSoF: The Kassel State of Fluency Dataset – Therapy Centered Dataset of Stuttering”, 2022 Proceedings of the Thirteenth Language Resources and Evaluation Conference
[20] Shakeel A. Sheikh, “Machine Learning for Stuttering Identification: Review, Challenges and Future Directions”, 2022, ISSN 0925-2312, S0925231222012772