SPECTROGRAM-DRIVEN CLASSIFICATION OF DYSARTHRIC SPEECH FOR ENHANCED RECOGNITION

Made by

Mr. Madhav N (Reg. No.125015157, B.Tech Information Technology)

Ms. Mathangi N (Reg. No. 125156159, B.Tech Computer Science and Engineering

(with specialisation Artificial Intelligence and Data Science)

Mr. Syed Arshad S A (Reg. No. 125158068, B.Tech Computer Science and Engineering

(with specialisation IoT & Automation)

List of Figures

List of Tables

Abbreviations

ML	-	Machine Learning
SVM	-	Support Vector machines
ASR	-	Automatic Speech Recognition
LSTM	-	Long Short – Term Memory Networks
CNN	-	Convolutional Neural Networks
S-CNN	-	2-dimensional Spatial Convolutional Neural Networks
DL	-	Deep learning
SV	-	Speech Vision
MFCC	-	Mel-Frequency Cepstral Coefficients

Abstract

Dysarthria is a neurological motor speech disorder characterized by an individual's loss of control of their motor subsystems. Symptoms of dysarthria can vary significantly depending on the underlying cause and severity that may result in speech produced be moderately slurred or severely unintelligible as the disease progresses and is often accompanied by neurological conditions which makes them physically debilitated, making interfacing with digital devices via peripherals challenging or impossible. Automatic Speech Recognition (ASR) technologies can make a significant difference for them. However, ASR technologies have performed poorly in recognizing dysarthric speech, especially for severe dysarthria, due to challenges such as alternation and inaccuracy of dysarthric phonemes, the scarcity of dysarthric speech data, and the phoneme labelling imprecision. These alternations mask the discriminative acoustic attributes that ASR systems rely upon to recognize phonemes. During this investigation we realized that some correlations exist in voicegrams extracted from different dysarthric utterances of the same word. A voicegram is the visualization of spectrum frequencies and their dynamics over time, presented as a heat map. Our approach uses the voice spectrograms to better detect the speech of the individual, in which various spectral features are extracted and dimensionality reduction is performed using Principal Component Analysis (PCA). The reduced features are then fed into Machine learning classifiers.

Keywords:

Dysarthria, Neurological-Speech Disorders, Speech-to-Text, Spectrogram, Feature Extraction, Dimensionality Reduction, Machine Learning

SNAPSHOTS

TRAINING AND VALIDATION LOSS, ACCURACY OF MODEL

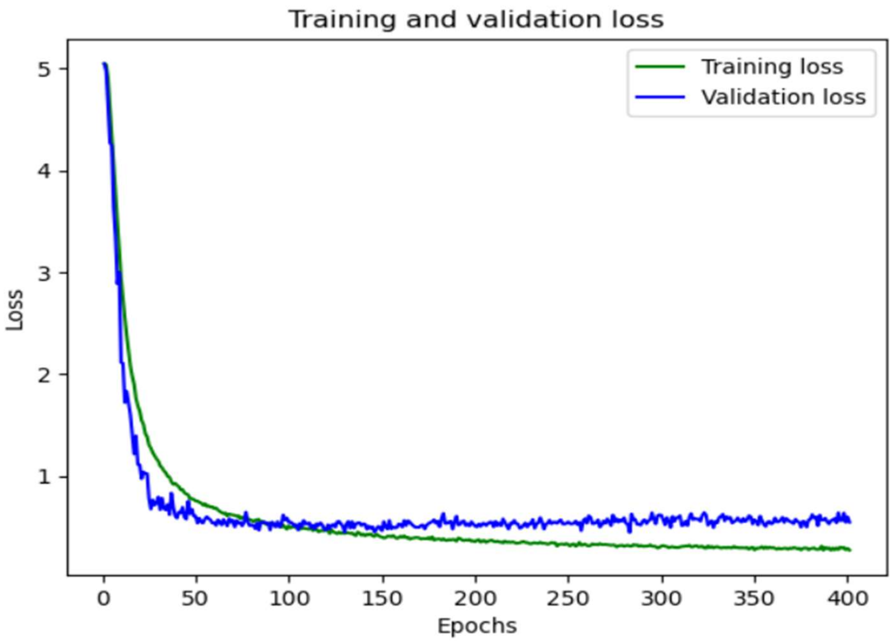


Fig 4.1 Training and Validation Loss

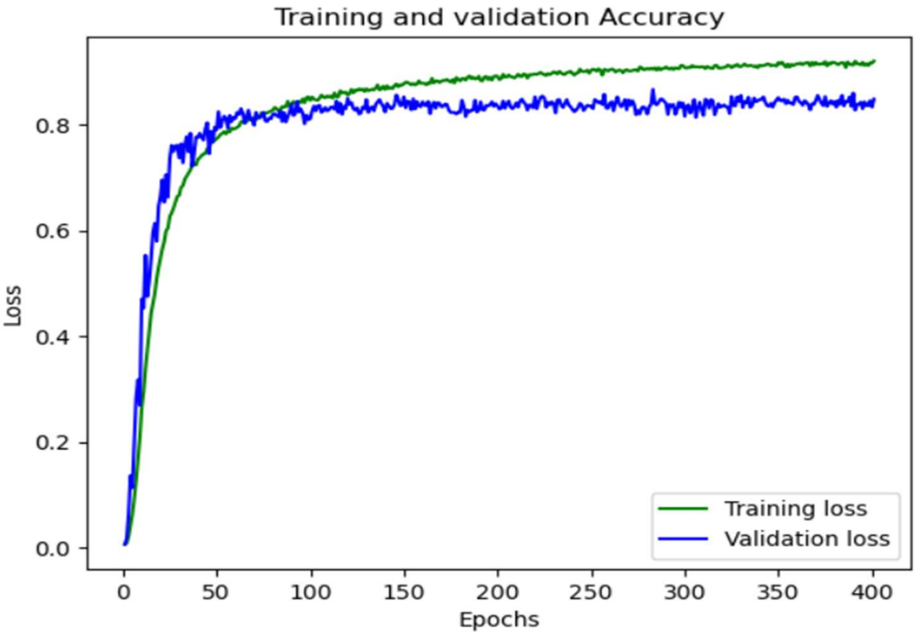


Fig 4.2 Training and Validation Accuracy

FOR 1000 SAMPLES

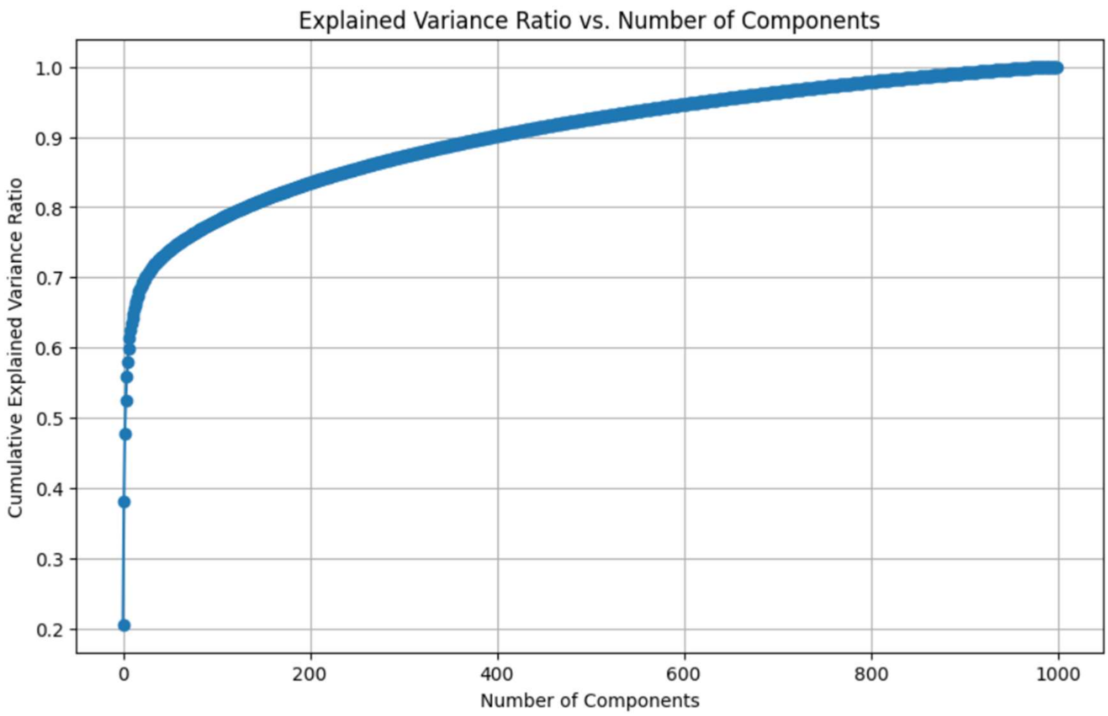


Fig 4.3 Variance Ratio vs Number of Components

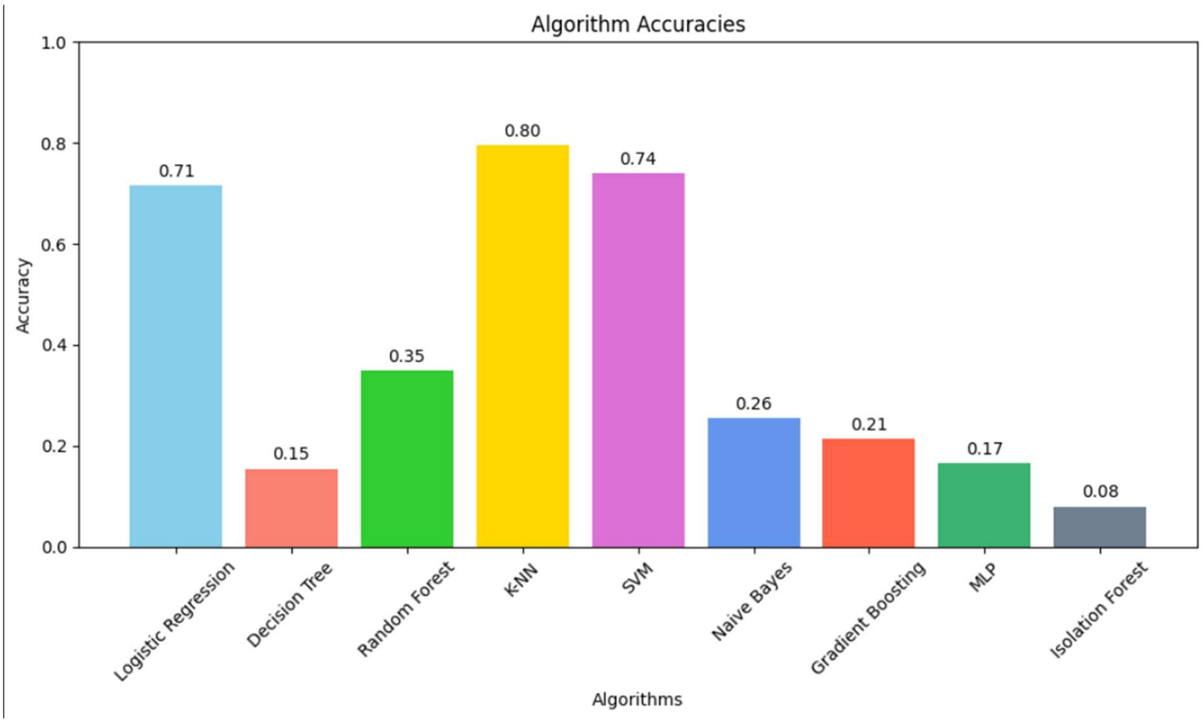


Fig 4.4 Algorithm Accuracies

<u>FOR 2000 SAMPLES</u>

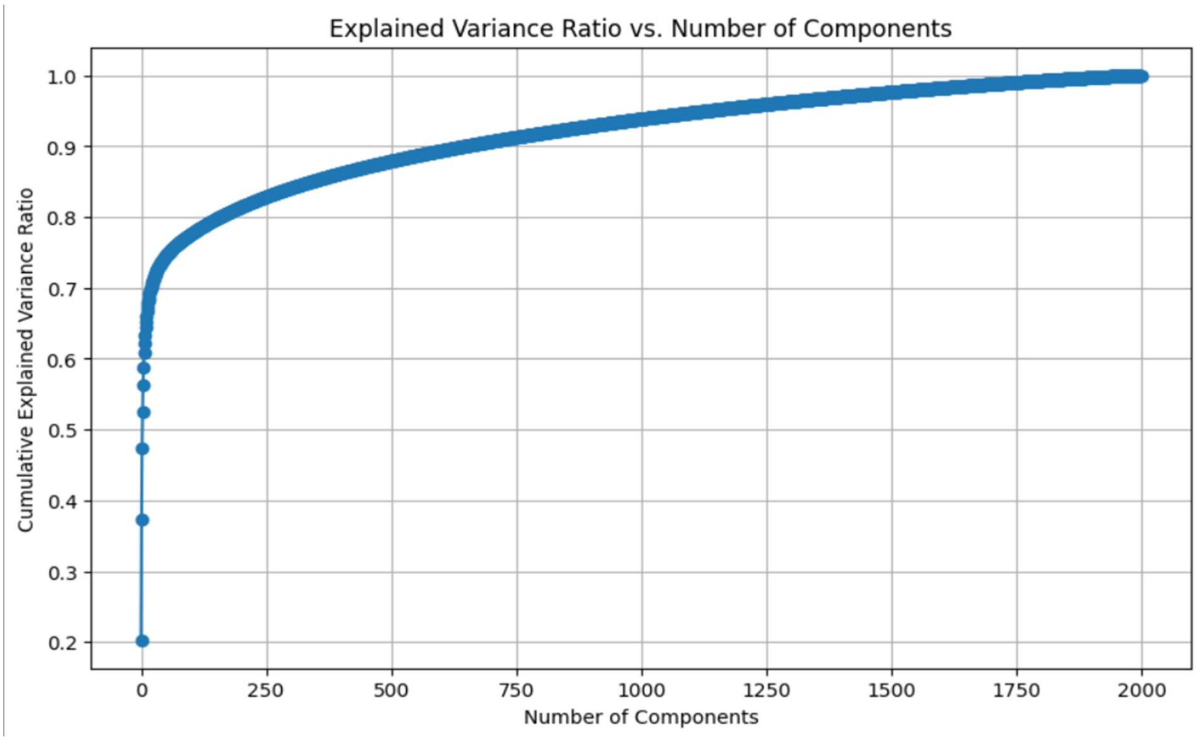


Fig 4.5 Variance Ratio vs Number of Components

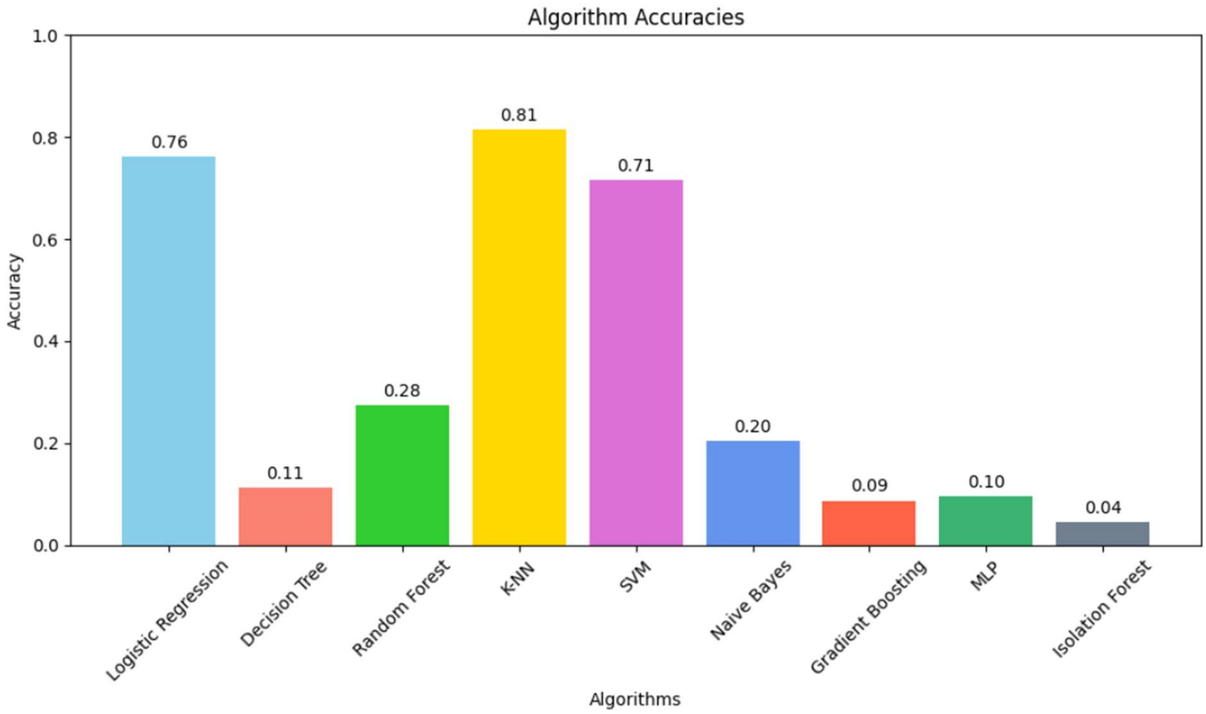
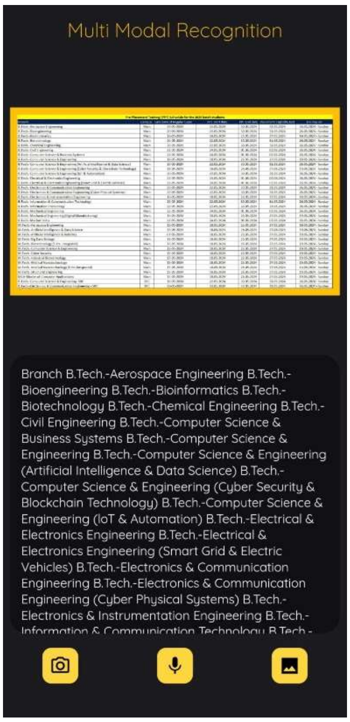
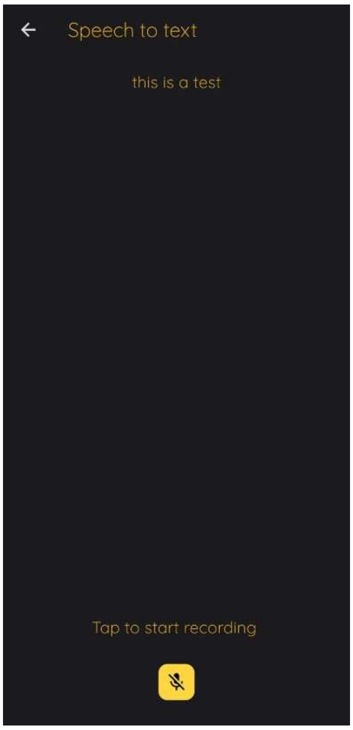


Fig 4.6 Algorithm Accuracies

APPLICATION INTERFACE





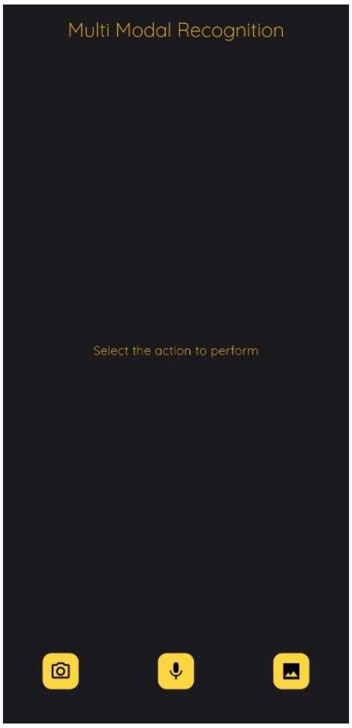


Fig 4.7 Application Interface

Table 4.1. Performance analysis

No. of samples/Classifier	1000 image samples	2000 image samples
SVM	74	71.5
Logistic Reg	71.5	76.25
KNN	79.5	81.5
Gradient Boosting	21.5	8.75
Decision Tree	15.5	11
Random Forest	35	28
Gaussian Naïve Bayes	25.5	20
Multi-Layer Perceptron	17	10
Isolation forest	8	4

CONCLUSION AND FUTURE PLANS

Three challenges were identified in developing dysarthric ASR systems and a system called Speech Vision, was proposed, that attempts to address them. The challenges were alternation and inaccuracy of phonemes in dysarthric speech, making conventional ASR systems less effective in recognizing dysarthric speech, unavailability of dysarthric speech samples, and inaccuracy and difficulties with labeling dysarthric phonemes. Different measures were taken to address the issue, which were converting word utterances into visual-feature representation and attempting to recognize shapes instead of phonemes, and generation of synthetic dysarthric speech using text-to-speech technologies, which upon observation seemed to show better results when compared to conventional methods. By using these gained insights, we have built a machine-learning model that can be used by dysarthric individuals to help them in their daily lives and make communication easy for them. Future plans include implementing our model into a mobile application that can be used by those individuals to make their daily lives easier and less dependent. Furthermore, the app can be integrated with custom machine-learning models, that can be fine-tuned for the individuals to better assist them in their life. It can include features like easy access to family contacts and other necessary services for quick and easy access. With this project, we would like to improve the quality of those individuals who suffer from these ailments.

REFERENCES

[1] M. Fernández-Díaz and A. Gallardo-Antolín, “An attention long short term memory based system for automatic classification of speech intelligibility,” Eng. Appl. Artif. Intell., vol. 96, Nov. 2020, Art. no. 103976, doi: 10.1016/j.engappai.2020.103976.

[2] C. Whillans, M. Lawrie, E. A. Cardell, C. Kelly, and R. Wenke, “A systematic review of group intervention for acquired dysarthria in adults,” Disability Rehabil., pp. 1–17, Dec. 2020, doi: 10.1080/09638288. 2020.1859629.

[3] N. P. Narendra and P. Alku, “Automatic assessment of intelligibility in speakers with dysarthria from coded telephone speech using glottal features,” Comput. Speech Lang., vol. 65, Jan. 2021, Art. no. 101117, doi: 10.1016/j.csl.2020.101117.

[4] Y. Zhao, M. Kuruvilla-Dugdale, and M. Song, “Voice conversion for persons with amyotrophic lateral sclerosis,” IEEE J. Biomed. Health Inform., vol. 24, no. 10, pp. 2942–2949, Oct. 2020, doi: 10. 1109/JBHI.2019.2961844.

[5] I. Calvo et al., “Evaluation of an automatic speech recognition platform for dysarthric speech,” Folia Phoniatrica et Logopaedica, pp. 1–10, Nov. 2020, doi: 10.1159/000511042

[6] S. R. Shahamiri, V. Lal and D. Shah, "Dysarthric Speech Transformer: A Sequence-to-Sequence Dysarthric Speech Recognition System," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 3407-3416, 2023, doi: 10.1109/TNSRE.2023.3307020.

[7] Tjandra, Andros, et al. "Towards Accurate Dysarthric Speech Recognition." *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH)*. 2017.

[8] Gómez-Alanís, Alma, et al. "Deep Learning for Dysarthric Speech Recognition: Insights and Directions." IEEE Transactions on Neural Systems and Rehabilitation Engineering 29.4 (2021): 1594-1604.

[9] Yuan, Jing, et al. "Towards Adaptive Dysarthric Speech Recognition with Deep Learning Techniques." Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH). 2019.

[10] Dukes, Krista, et al. "Enhancing Dysarthric Speech Recognition with Deep Transfer Learning." Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH). 2018.