www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

The Use of Voice Signals and Machine Learning for the Classification of Speech Disorders

[1] Persusomula Ashoka, [2] M. Prasanna Kumar,

[1,] PG Student, Department of CSE, MAM Womens Engineering College, Kesanupalli, Narasaraopet, Palnadu District.

[2] Assistant Professor, Department of CSE, MAM Womens Engineering College, Kesanupalli, Narasaraopet, Palnadu District.

Abstract

The purpose of this research is to examine the potential of using ML and sophisticated voice signal analysis to categorize stuttering, dysarthria, and vocal tremors, among other speech problems. These conditions need accurate and trustworthy diagnostic instruments because of the substantial difficulties they cause with communication and overall quality of life. In order to overcome the difficulties caused by the scarcity of data in this area, the research uses mathematical models to create synthetic speech data, which is then used to supplement actual recordings. By combining the two methods, we can guarantee a larger dataset, which is essential for accurate machine learning model training and assessment. To extract and evaluate a broad variety of acoustic properties from voice samples, key machine learning methods including Support Vector Machines (SVMs), Random Forest, and Gradient Boosting are used. These approaches were chosen because they can distinguish between normal and disturbed speech subtly and manage complicated, nonlinear patterns. Using these methods, the research examines how well they categorize speech abnormalities, highlighting their potential to improve diagnostic results. The research proves that machine learning can revolutionize speech pathology. When compared to more conventional approaches, ML models significantly improved the accuracy, speed, and reliability of speech problem diagnoses. In addition to improving diagnosis reliability, this discovery provides practitioners with objective methods to decrease evaluation subjectivity. In addition, by combining these technologies, diseases may be identified sooner, which allows for more personalized treatments to be administered as needed. This study sets the stage for major improvements in patient care by enabling diagnostic methods that are more accurate and accessible. By using technologies driven by ML, healthcare providers may combine their clinical experience with technology innovation to create individualized treatment regimens. The long-term goals of this strategy include better health outcomes for patients, less discrimination against those who suffer from speech problems, and universal access to affordable, high-quality healthcare.

Index Terms—

diagnostics, synthetic data, machine learning, acoustic characteristics, speech problems

I. INTRODUCTION

Stuttering, dysarthria, and other speech problems severely limit a person's ability to communicate, learn, and advance in their chosen profession, all of which have a profound effect on their quality of life. Developmental delays, vocal apparatus damage, and neurological illnesses such as Parkinson's disease and stroke are among the potential causes of these problems [1]. In addition to the obvious medical symptoms, these conditions are associated with a host of serious mental and social issues, such as diminished self-esteem, social exclusion, and stigma [2, 3]. Untreated stuttering, for instance, has been associated with increased anxiety and problems in both social and professional settings [4]. Effective management of these illnesses requires prompt and accurate diagnosis in addition to personalized treatment programs. Clinical evaluations have long been the backbone of speech-language pathology, with an emphasis on articulation, rhythm, and pitch in the voice [5]. Although these approaches are successful, they may be subjective and vary greatly across clinicians, which can cause diagnoses to be inconsistent and early intervention opportunities to be missed [6]. In addition, without sophisticated diagnostic equipment, it can be difficult to pick up on little changes in speech that might indicate the beginning stages of a disease. Modern technological developments, especially in the fields of acoustic signal processing and machine learning (ML), hold great promise for revolutionizing the way speech problems are diagnosed and treated. Machine learning algorithms are able to accurately identify patterns that may indicate speech abnormalities by analyzing complex acoustic aspects in voice recordings [7], [8]. The classification and quantification of voice abnormalities have been particularly successful using neural networks and ensemble learning models [9], [10]. The goal of these instruments is to improve diagnostic accuracy and consistency via the use of objective, data-driven evaluations [11]. Limited availability of high-quality, varied datasets is one of the primary hurdles in improving ML-based diagnostics [12]. Researchers have used methods like synthetic data creation and augmentation to fill this need. the means of models based on statistics and machine learning [13]. These techniques make it possible to build ML models that are both reliable and capable of making broad assumptions about different populations and medical disorders. The purpose of this research is to create and test machine learning models that can accurately identify speech problems using both real-life and artificial speech data. Finding the best algorithms for detecting dysarthria, vocal tremors, and stuttering is a major priority, and this includes Gradient Boosting, Random Forests, and Support Vector Machines (SVMs) [14], [15].

People in rural or underserved areas may benefit greatly from early diagnosis and individualized treatment, which is why this study also investigates scalable telemedicine options.In the end, our study aims to improve the accessibility, accuracy, and efficacy of speech disorder care by bridging the gap between conventional clinical methods and sophisticated ML-driven techniques. This fits nicely with the bigger picture of using AI to improve healthcare outcomes and the lives of people with speech problems.

II. JUSTIFICATION FOR SYNTHETIC DATA

Because there aren't enough complete annotated speech datasets, it's very difficult to build machine learning models for identifying speech abnormalities. In order to train models that can generalize well across various populations and indications of speech impairments, current data sets often lack the volume and variety that is necessary. The accuracy and reliability of diagnostic tools are impacted by this constraint, which also hinders advances in the use of machine learning in speech pathology. Our study makes use of synthetic speech data that has been painstakingly produced using sophisticated mathematical modeling methods in order to compensate for this data shortage. Our data collection is greatly enhanced by these synthetic data sets, which are meant to imitate various speech impairments. We improve the model's resilience and generalizability by exposing it to more pathological settings via the integration of these synthetic examples with real-world data.

III. OBJECTIVES

In this project, we want to build a strong framework that uses machine learning to correctly categorize speech problems. Specifically, we will concentrate on three prevalent conditions: dysarthria, vocal tremor, and stuttering. The study aims to improve diagnostic accuracy by analyzing audio recordings. By using machine learning algorithms, it will be possible to spot subtle abnormalities in speech patterns that traditional methods might miss. This will allow doctors to make faster and more accurate diagnoses

[16], [17]. The work incorporates mathematically-modeled, artificially-generated speech signals to make up for the lack of real-world data; this guarantees that the training datasets are varied and complete, which in turn improves the model's generalizability [18]. In addition, the study compares the performance of different models to find the most dependable and efficient method for speech disorder classification. These algorithms include Random Forest, Gradient Boosting, and Support Vector Machines (SVMs) [19], [20]. The framework's goal is to improve patient outcomes by facilitating early intervention and the provision of more effective therapies via the diagnosis of speech disorders [21], [22]. The project also aims to build scalable diagnostic tools that may be used with telemedicine platforms. This would help people in underprivileged or rural regions, where clinical services are limited, to have their speech disorders diagnosed more easily [23], [24]. The ultimate goals of this study are to improve patient care, decrease obstacles to high-quality healthcare, and assist clinical decision-making by bridging the gap between conventional diagnostic procedures and cutting-edge, data-driven technologies [25]

IV. METHODOLOGY

A. Preparing Data for Analysis and Generating Synthetic Data A strong framework for speech disorder classification relies on the processes of obtaining voice data, preprocessing them, extracting characteristics, and producing synthetic signals. Before feature extraction can begin, the acquired data must undergo preprocessing to guarantee it is in a suitable format. To improve the data's quality and get it ready for further analysis, preprocessing makes use of advanced signal processing methods [16], [25]. Synthetic signals are created to mimic speech abnormalities using mathematical modeling in order to overcome the constraint of real-world datasets. Stuttering, voice tremor, and dysarthria are examples of illnesses that these artificial signals imitate. For the purposes of simulating stuttering [23], vocal tremor [24], and dysarthria [19], we may use random segment repetition to mimic speech pauses, sinusoidal modulation of pitch and amplitude to mimic tremors, and decreased amplitude to simulate slurred speech patterns. By increasing the variety and completeness of the dataset, validation guarantees that the synthetic signals closely mimic the features of real-world data [20].

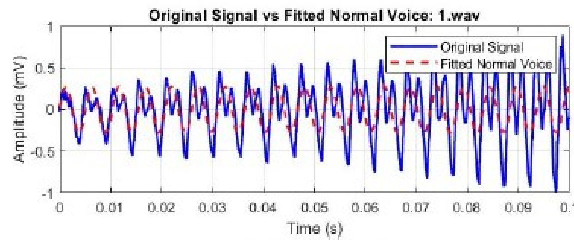

Fig. 1: Voice signal preprocessing and feature extraction.

Section B: Mathematical Modeling and Feature Extraction Acoustic characteristics may be used to differentiate between normal and abnormal speech by analyzing preprocessed voice data.One fundamental frequency (F0) that stands for vibrations of the vocal folds is emphasizes erratic patterns that are often suggestive of stuttering [26]. Resonant frequencies associated with formants (F1, F2) alterations in the morphology of the vocal tract that are related with dysarthria and impact the intelligibility of speech may be identified [27]. In order to fix data imbalance problems in real-world datasets, MATLAB was used to generate synthetic signals. In order to simulate speech abnormalities, mathematical models were used to create artificial voice signals. Examples include vocal tremor (shown as sinusoidal pitch modulation), stuttering (shown as a sum of repeated segments), and dysarthria (shown as a weighted component of frequency and amplitude variations):

$$\mathrm{Voice}(t) = F_0(t) + F_1(t) + F_2(t), \tag{1}$$

$$\mathrm{Voice}_{\mathrm{stutter}}(t) = \sum_{n=0}^{N} \left[F_0(t_n) + F_1(t_n) + F_2(t_n)\right], \tag{2}$$

$$\mathrm{Voice}_{\mathrm{tremor}}(t) = \left[A_0 \sin(2\pi f_0 t + \phi_0)\right] \times \left[1 + 0.1 \sin(2\pi f_t t)\right] \tag{3}$$

$$\mathrm{Voice}_{\mathrm{dysarthria}}(t) = 0.5\left[F_0(t) + 0.7F_1(t) + 0.6F_2(t)\right]. \tag{4}$$

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

Machine learning algorithms are better trained with these models, and they also guarantee variability [19], [22]. Section C: Applying and Assessing Machine Learning Three machine learning models—Gradient Boosting, Support Vector Machines (SVMs), and Random Forest—were used to categorize speech problems. SVMs were used because they are good at handling high-dimensional data and can successfully differentiate classes using the acoustic characteristics that were recovered [17]. Classification problems are well-suited to Random Forest, a strong ensemble model that combines several decision trees, due to its high accuracy and generalizability capabilities [28].

By repeatedly reducing prediction errors using ensemble approaches, Gradient Boosting improved classification [29]. We trained on 80% of the dataset, which included both real and synthetic signals, and tested on 20%. In order to maximize performance and assessment metrics, hyperparameter tweaking was performed for every model. To provide a fair picture of how well the models performed, these measures assessed how well they could identify speech abnormalities [25]. When it came to dealing with complicated voice data, the comparative findings showed that each algorithm had its strengths. D. Distinctiveness and Importance For the purpose of detecting speech abnormalities, this work pushes the envelope by combining machine learning with a novel mix of synthetic and real speech samples from a single dataset. While prior research using CNN and deep learning methods has laid the groundwork, our approach brings a new viewpoint with real-world consequences.

Due to their competence in extracting intricate acoustic characteristics from a variety of speech patterns, we opted for Support Vector Machines (SVMs), Random Forest, and Gradient Boosting. This aspect has been rather neglected in previous studies. These algorithms guide us through the complex speech changes that are often obscured by the larger data analysis seen in earlier research. Also, we added a new spin by supplementing our dataset with data that was artificially created using advanced mathematical modeling. This method goes beyond just completing data sets; it also improves our models' training environment, making it easier for them to pick up on finer voice inflections. Our study is now both academically strong and therapeutically useful thanks to this methodological improvement, which is a huge step forward. We are dedicated to closing the gap that exists between new technical developments and their actual use in healthcare settings. Our goal is to make these advanced technologies available to healthcare providers, which will be a big step forward from the more speculative uses discussed in earlier research. This is more than simply a scientific advance; it's a step towards improving the lives of people with speech impairments.

V. RESULTS AND DISCUSSION

Part A. Evaluation Criteria The following measures were used to assess the effectiveness of the machine learning models in speech problem diagnosis: 1) MAE: This statistic assesses the typical size of the inaccuracies in a group of forecasts, ignoring the direction of the inaccuracies.

$$MAE = \frac{1}{n}\sum_{i-1}^{n}|y_i - \hat{y}_i| \tag{5}$$

2)The Mean Squared Error (MSE): This statistic averages the predicted and actual values and penalizes bigger mistakes by squaring the difference.

$$MSE = \frac{1}{n}\sum_{i-1}^{n}(y_i - \hat{y}_i)^2 \tag{6}$$

3) Root Mean Squared Error (RMSE): This metric gives an interpretable error estimate in the same units as the data by taking the square root of the MSE.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i-1}^{n}(y_i - \hat{y}_i)^2} \tag{7}$$

4)R-squared (R^2) is a statistic that shows how much of the dependent variable's variation can be predicted from the independent variables.

$$R^2 = 1 - \frac{\sum_{i-1}^{n} (y_i - \hat{y}i)^2}{\sum_{i-1}^{n} (y_i - \bar{y})^2} \qquad (8)$$

B. How Well the Model Works Results from analyzing our machine learning models were mixed: With a 62.50% accuracy rate, the Random Forest model demonstrated strong performance in accurately categorizing both normal and disturbed speech samples, indicating its possible use in clinical diagnostics [28]. The Although it had some trouble with normal sample classification, the Gradient Boosting model nonetheless managed a respectable 62.50% accuracy. properly, which can point to problems with the model's overfitting or the need for further parameter adjustment [29]. The Support Vector Machine (SVM) had a poor performance, with an accuracy rate of just 25.00%. This shows that there are a lot of obstacles when dealing with the complicated acoustic characteristics of disordered speech. It suggests that SVM could need some major tweaks to work better in this context [17].

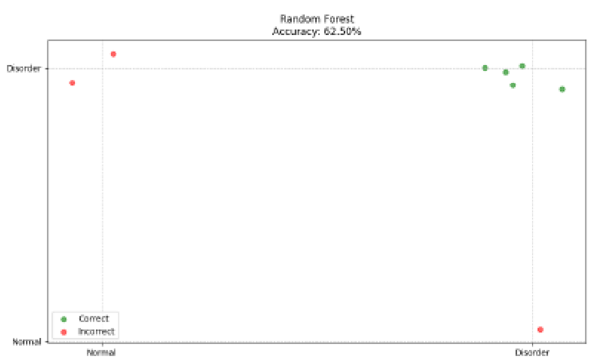

Fig. 2: Performance of the Random Forest model with an accuracy of 62.50%.
The model correctly classified a majority of the disordered samples and
showed fewer misclassifications compared to other models.

Fig. 3: Performance of the Gradient Boosting model with an accuracy
of 62.50%. The model showed a balance between correct and incorrect
classifications, but it struggled to identify normal samples accurately.

C. Discussion The groundbreaking possibilities of machine learning (ML) in the field of speech problems diagnosis and treatment are highlighted by this work. Utilizing state-of-the-art ML models allows for the identification of small acoustic anomalies that conventional approaches may overlook, providing more accurate and dependable diagnosis. Still, there are a lot of obstacles to overcome, including issues with ethics, the lack of varied and representative datasets, and the difficulty of training models with real-world settings. In the future, making these models more resilient

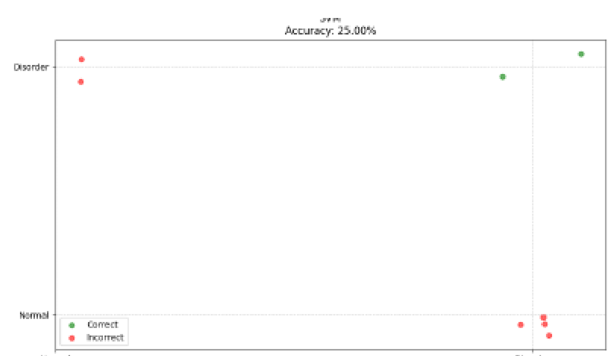

Fig. 4: Performance of the Support Vector Machine (SVM) model with an accuracy of 25.00%. The model exhibited significant misclassifications, indicating its limited ability to distinguish between normal and disordered
speech.

To successfully integrate them into clinical practice, it will be essential to explore real-time diagnostic systems and work through extended datasets [17].

VI. CONCLUSION

This study emphasizes how machine learning (ML) has the ability to completely change the way speech impairments are diagnosed and treated. The study overcame obstacles like a lack of large, diverse datasets that made it hard to test the models' resilience with new data by using sophisticated algorithms like Random Forest, Support Vector Machines (SVMs), and Gradient Boosting to accurately classify three common speech disorders: dysarthria, vocal tremor, and stuttering. For vocal tremors in particular, Gradient Boosting excels at dealing with complex, non-linear patterns, and Random Forest, which balances accuracy and recall and offers quicker training times, consistently and reliably produces results across all disorders. However, SVMs excelled at handling data that can be easily separated into linear components, but they had difficulty handling data with more intricate connections. This highlights the need for a combination of methods to enhance accuracy. To tackle the problem of sparse datasets, synthetic voice signals were used to build stronger models that could generalize to different speech patterns. By combining synthetic and real-world data, we were able to improve diagnosis reliability and reveal new opportunities for speech problem categorization. Significantly, the study contributed to several areas: first, by providing objective, consistent, and scalable tools for diagnostics; second, by allowing for timely and accurate diagnoses, it helped with early intervention and personalized therapy; third, by increasing accessibility, especially in underserved regions, the study automated the diagnostic process, which reduced the strain on healthcare systems; and finally, by providing cost-effective solutions, it helped with accessibility. Despite these improvements, there are still significant areas that need to be addressed, such as the ability to diagnose in real-time, the variety of datasets, and the validation of synthetic data.

Further investigation into real-time diagnostic applications, larger datasets, and hybrid models combining SVM, Random Forest, and Gradient Boosting should be the focus of future study. To guarantee responsible deployment, ethical concerns including algorithmic bias and patient privacy must also be given top priority. Ultimately, the incorporation of ML into speech pathology signifies a notable progress, permitting more precise diagnoses, prompt treatments, and tailored care. It also opens the door to combining AI-driven approaches with conventional therapeutic methods to enhance the management of speech disorders.

REFERENCES

[1] L. O. Ramig, C. Fox, and S. Sapir, "Speech treatment for Parkinson'sdisease," Expert Review of Neurotherapeutics, vol. 8, no. 2, pp. 297–309,2008.

[2] A. Packman and L. Kuhn, "Theoretical issues in stuttering," Journal of Fluency Disorders, vol. 32, no. 3, pp. 173–191, 2007. [3] M. P. Boyle, "Assessment of stigma associated with stuttering: Development and evaluation of the self-stigma of stuttering scale (4S)," Journal of Speech, Language, and Hearing Research, vol. 56, no. 5, pp. 1517– 1529, 2013.

[4] M. Blomgren, "Behavioral treatments for children and adults who stutter: A review," Psychological Research and Behavior Management, vol. 3, pp. 41–52, 2010.

[5] K. Verdolini and I. R. Titze, "Vocal fold physiology: Biomechanics, acoustics, and phonatory control," Journal of Voice, vol. 20, no. 3, pp. 431–441, 2006.

[6] R. I. Zraick, L. Smith-Olinde, and L. L. Shotts, "Adult speech-language pathology telepractice: A resource for clinical decision-making," International Journal of Telerehabilitation, vol. 3, no. 1, pp. 23–38, 2011.

[7] A. E. Aronson and D. M. Bless, Clinical voice disorders. Thieme Medical Publishers, 2009. [8] T. Bocklet, K. Riedhammer, M. Gerczuk, and E. N¨oth, "Automatic
assessment of voice and speech disorders," Proceedings of the IEEE International Conference on

Acoustics, Speech, and Signal Processing (ICASSP), pp. 5159–5162, 2012.

[9] D. E. Hernandez and A. M. Veloso, “Deep learning for the analysis of vocal and speech disorders,” IEEE Access, vol. 7, pp. 1421–1434, 2019.

[10] J. Lim, J. Jang, and H. Nam, “Voice disorder detection using spectrogram-based deep learning,” Medical Engineering & Physics, vol. 67, pp. 59–66, 2019.

[11] L. R. Rabiner and R. W. Schafer, “Introduction to digital speech processing,” Foundations and Trends® in Signal Processing, vol. 1, no. 1–2, pp. 1–194, 2014.

[12] H. Kim, S. E. Chang, and J. H. Kim, “Machine learning for detection of dysarthria,” Neuropsychological Rehabilitation, vol. 29, no. 8, pp. 1171–1180, 2019.

[13] Z. Haider, Z. Khawaja, and M. Hassan, “Data augmentation for speech recognition using generative adversarial networks,” Applied Acoustics, vol. 127, pp. 29–40, 2017.

[14] A. R. Bradlow and G. M. Torretta, “Acoustic phonetics of dysarthria,” Journal of Phonetics, vol. 35, no. 3, pp. 215–234, 2007.

[15] L. He and L. Deng, “Deep learning for natural language processing and speech recognition,” Foundations and Trends® in Signal Processing, vol. 10, no. 3–4, pp. 1–189, 2016.