International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
ISSN:3048-7722
Volume 4, Issue 4, July-Aug-2026, PP: 01-6

Automatic Speech Disorder Detection Using Voice Signals and Machine Learning Techniques

K.Rajkumar[1], J.Akshaya[2]

[1]Assistant Professor, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana –, India.

[2]MCA Student, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana –, India.

***Abstract* – When a person has trouble communicating, it may have a negative impact on their social life, academic performance, and mental health, all of which contribute to a worse quality of life overall. Stuttering, vocal tremor, and dysarthria are three of the most common speech disorders; these conditions have different acoustic features and need to be properly diagnosed in order to provide effective therapeutic treatment. The assessment method is subjective, time-consuming, and reliant on clinical skill in conventional diagnostic processes since it mostly relies on perceptual evaluation undertaken by speech-language pathologists. This has led to a rise in the need for sophisticated AI-powered diagnostic tools that can accurately and consistently categorise speech disorders. Opportunities for the development of automated diagnostic frameworks that analyse complicated audio patterns with higher accuracy have arisen due to recent breakthroughs in digital voice processing and Machine Learning (ML). This research introduces a system that uses machine learning to automatically categorise speech problems based on audio signal analysis. The suggested method augments model generalisability and increases dataset variety by combining real-world voice recordings with synthetically produced speech signals produced by MATLAB-based mathematical modelling, thus overcoming the barrier of limited clinical speech datasets. Important acoustic parameters, such as fundamental frequency (F0), formant frequencies (F1 and F2), and signal augmentation and normalisation are performed on speech recordings during preprocessing in order to extract them. These properties characterise normal and disordered speech patterns, respectively. After that, several supervised machine learning algorithms, such as Gradient Boosting, Support Vector Machine (SVM), and Random Forest, are trained on the retrieved characteristics to categorise speech samples into various disorders. Traditional regression-based performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2) are used to assess the built models, in conjunction with an 80% training and 20% testing approach. Under the specified experimental circumstances, Random Forest and Gradient Boosting outperform Support Vector Machine in classification performance, with an overall accuracy of 62.50%, according to the experimental study. Based on the results, ensemble learning approaches outperform traditional classifiers when it comes to complicated pathological speech patterns. Objective speech disorder diagnosis, early clinical intervention, and enhanced telemedicine-based speech assessment systems for improved patient care are all possible outcomes of the suggested framework's integration of machine learning, acoustic signal processing, and synthetic data generation.**

***Keywords* - Speech Disorder Detection, Machine Learning, Voice Signal Analysis, Speech Signal Processing, Feature Extraction, Artificial Intelligence, Classification Algorithms**

I. INTRODUCTION

One of the most fundamental components of human communication is the capacity to put one's thoughts, emotions, and information into words. The effects of speech production disabilities on an individual's personal, academic, professional, and social life may be substantial. Speech disorders, such as stuttering, dysarthria, and vocal tremor, affect articulation, fluency, voice quality, and pronunciation; they are among the most common types of impaired communication. Many other physiological conditions may manifest in this way; they include, but are not limited to, disorders of the nervous system, developmental abnormalities, damage to the voice cords, stroke, Parkinson's disease, traumatic brain injury, and similar conditions. Timely treatment enhances the quality of life for those with speech disorders, but only if an accurate diagnosis is made. This is due to the fact that psychological health and communication abilities are often affected by speech disorders.

Historically, clinical assessments have been the mainstay for speech-language pathologists for diagnosing speech problems. Medical professionals use a battery of acoustic tests, which include pitch, articulation, rhythm, intensity, and fluency of speech, to determine the kind and severity of the condition. Despite their lengthy history of acceptability, these clinical evaluations are inherently subjective and may vary according to the clinician's level of training and expertise. There is a risk that manual screening methods may overlook subtle abnormalities in speech disorders at an early stage, delaying therapy. This highlights the pressing need for more advanced computer-assisted diagnostic tools capable of reliably and objectively classifying speech abnormalities.

Automation in voice recognition has taken a giant leap forward because to advancements in digital signal

processing, artificial intelligence, and machine learning. Machine learning algorithms have replaced more laborious human clinical assessments by deciphering complex acoustic data from audio recordings, exposing hidden speech patterns. The capacity of intelligent diagnostic systems to analyse several speech characteristics simultaneously improves their ability to reliably distinguish between healthy and sick speech. These technological advancements have further established machine learning as a promising tool for physicians to use in the diagnosis of communication disorders.

One of the biggest challenges in creating reliable algorithms for speech disorder classification is the absence of easily available, large-scale, annotated clinical speech datasets. Speech recordings from patients diagnosed with specific disorders might be challenging to get for a variety of reasons, including privacy concerns, limited patient groups, and the substantial effort required for clinical annotation. Machine learning algorithms struggle to generalise their findings to other patient populations and learn best with large datasets. Researchers have used synthetic speech generation approaches to statistically simulate aberrant speech characteristics in attempt to get around this limitation. Combining synthetic voice samples with real clinical recordings allows for the construction of more comprehensive training datasets. Because of this, robust machine learning models can deal with varying degrees of voice variation.

This study introduces a machine learning system that can classify speech abnormalities automatically using data from both real and synthetic voices. Create synthetic speech samples that imitate the acoustic aspects of dysarthria, vocal tremor, and stuttering using mathematical models in MATLAB. The processed speech signals are analysed for important acoustic features including fundamental frequency (F0) and formant frequencies (F1 and F2) in order to perform categorisation. Three examples of supervised machine learning algorithms that are trained and evaluated with a training-to-test ratio of 80% to 20% include gradient boosting, supported vector machine (SVM), and random forest. In the end, we want to build a sophisticated diagnostic system that will allow telemedicine to flourish and clinicians to treat speech issues more quickly and individually.

II. LITERATURE SURVEY

Automatic speech disorder identification is a growing area of research in AI, biomedical engineering, and speech processing because of the objective and rapid clinical diagnosis it can give. The use of deep learning and machine learning to evaluate audio data retrieved from voice recordings for the purpose of diagnosing aberrant speech has been the subject of much study. There is encouraging evidence that these AI systems may improve diagnostic consistency while reducing the need for subjective clinician assessment.

Researchers mostly used conventional speech signal processing techniques when first studying vocal abnormalities associated with neurological and communicative disorders. After extracting the acoustic properties from the recorded voice, such as fundamental and formant frequencies, jitter, shimmer, energy, and speech rate, the data was statistically analysed. The ability of doctors to distinguish between healthy and abnormal speech using these handcrafted acoustic signals was contingent upon feature extraction methods developed by specialists and limited computational resources.

As a result of advancements in machine learning, supervised learning algorithms such as Decision Trees, Support Vector Machines (SVMs), Artificial Neural Networks, Gradient Boosting, and Random Forests were created to categorise speech disorders. These algorithms demonstrated improved performance by automatically detecting complex relationships among various auditory features. The ability of ensemble learning approaches such as Random Forest and Gradient Boosting to simulate nonlinear acoustic fluctuations and enhance classification resilience across various speech disorders is shown by their persistent outperformance of traditional classifiers in comparative research.

Recently, there has been an uptick in studies investigating the feasibility of using deep learning techniques to detect anomalies in speech. Deep learning models trained on spectrograms, Recurrent Neural Networks (RNNs), and Convolutional Neural Networks (CNNs) have successfully learned features automatically from audio sources. Unlike conventional machine learning methods, which need hand-crafted acoustic descriptors, deep learning architectures may autonomously construct hierarchical feature representations capable of capturing nuanced pathological speech characteristics. Several investigations on the use of deep learning algorithms for the diagnosis of dysarthria, Parkinson's disease, and vocal tremor have shown encouraging findings.

Another major challenge, according to literature assessments, is the absence of sufficiently large clinical speech datasets. Inadequate patient recordings in many publicly available datasets of speech disorders reduce the generalisability of machine learning models. Data augmentation and synthetic data creation methods have been used by researchers to statistically imitate aberrant speech features in an attempt to address this issue. Using synthetic speech samples improves model robustness, dataset diversity, and data imbalance mitigation. Artificial voices may supplement machine learning in situations when high-quality, real-world clinical recordings are unavailable.

Combining machine learning, synthetic data creation, and speech signal processing is one effective technique to automatically categorise speech abnormalities, according to the current corpus of research. Problems with different

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
ISSN:3048-7722
Volume 4, Issue 4, July-Aug-2026, PP: 01-6

datasets, real-time deployment, generalising models, and clinical validation still need more research. The ongoing integration of state-of-the-art signal processing techniques with smart machine learning models is anticipated to greatly improve the precision, usability, and dependability of computer-assisted speech problem diagnosis.

III.PROPOSED METHODOLOGY

In order to automatically detect problematic speech via acoustic analysis of voice recordings, the suggested framework offers a smart Machine Learning-based Speech Disorder Classification System. Stuttering, vocal tremor, and dysarthria are the three main speech disorders that the system intends to categorise using supervised machine learning and speech signal processing. Acquisition of voice signals, preprocessing of speech, production of synthetic data, extraction of acoustic features, classification using machine learning, and assessment of performance are the six main processes that make up the entire framework.

The first step is to acquire voice signals, which include both naturally occurring clinical speech recordings and signals produced artificially. The suggested system increases the variety of accessible pathological speech datasets by using mathematical modelling methods to create synthetic speech signals that stand in for stuttering, vocal tremor, and dysarthria, as there aren't many of these available publically. These synthetic voice samples enhance the training dataset's resilience by mimicking the acoustic features of actual disordered speech.

Before feature extraction, all audio recordings are subjected to speech preprocessing to enhance the signal quality, after data gathering. In order to eliminate undesired background interference while keeping relevant auditory information, preprocessing procedures include noise reduction, signal normalisation, segmentation, and filtering. These procedures provide consistent audio signals that may be used for accurate feature extraction.

The next step is to apply acoustic feature extraction to the processed speech data. This will extract crucial speech properties such the fundamental frequency (F0) and formant frequencies (F1 and F2). These acoustic parameters characterise the features of normal and abnormal speech, including vibration of the vocal folds, articulation, and resonance. To further increase the variety of training data, mathematical models are used to imitate the frequency, amplitude, and temporal fluctuations associated with various speech disorders.

Support Vector Machine (SVM), Random Forest, and Gradient Boosting are three supervised machine learning algorithms that are fed the extracted feature vectors. The models are trained to categorise voice recordings into Normal Speech, Stuttering, Vocal Tremor, and Dysarthria. The dataset is split into 80% training data and 20% testing data. Classification accuracy and model generalisability are both enhanced by hyperparameter optimisation.

Last but not least, the trained models are assessed using R^2, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Squared Error (MSE). According to the results of the experiments, the Support Vector Machine has a total accuracy of 25.00%, while Random Forest and Gradient Boosting get the best results (62.50%). Supporting intelligent clinical decision-making and telemedicine applications, the proposed framework shows that automated speech disorder diagnosis can be achieved by integrating synthetic data generation, machine learning, acoustic feature extraction, and speech signal processing.

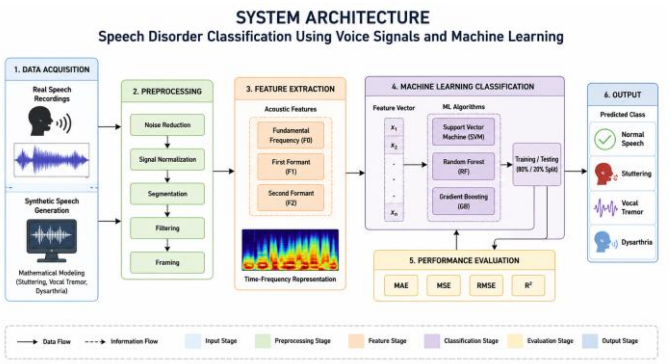

Fig.1 System Architecture

IV. METHODOLOGY

As an example of a structured Machine Learning (ML) architecture, the method demonstrates how to automatically classify speech abnormalities by analysing audio signals. The solution unifies a single diagnostic procedure that integrates speech signal preprocessing, synthetic data generation, acoustic feature extraction, machine learning classification, and performance evaluation. Despite working with small clinical speech datasets, the primary objective is to consistently distinguish between normal speech and disorders such dysarthria, stuttering, and vocal tremor. The proposed method combines real-world audio with synthetic speech signals to improve the in-built classification models and expand the dataset diversity.

Obtaining voices from both real-life clinical datasets and synthetically generated speech samples is the first step in the process, known as voice signal collection. Since it is challenging to acquire large enough pathological speech datasets due to privacy concerns and a lack of patient availability, the proposed method employs mathematical modelling in MATLAB to generate synthetic speech signals that represent different speech disorders. These synthetic recordings of voices imitating dysarthria, vocal tremor, and stuttering have expanded the training data available to machine learning algorithms.

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
ISSN:3048-7722
Volume 4, Issue 4, July-Aug-2026, PP: 01-6

Following data acquisition, all voice recordings undergo signal preprocessing to improve recording quality and prepare the data for feature extraction. After the background noise is reduced using noise reduction techniques, the signal strength is adjusted to be consistent across all recordings by altering the amplitude. It is possible to further enhance the quality of the voice signals while preserving critical acoustic information needed for accurate disorder classification by using speech segmentation and filtering procedures. If you follow these guidelines, your speech data will be ready for feature extraction with no surprises.

After the first processing is complete, the system uses acoustic feature extraction to distinguish between typical and non-standard speech patterns. The fundamental frequency (F0) is one of the acoustic characteristics that was recovered; it reflects the vibration of the vocal folds. The first and second formant frequencies (F1 and F2) are another; they indicate the resonance features of the vocal tract. These auditory features could provide light on how speech is articulated, phonated, frequency-shifted, and stabilised. Synthetic samples that are strikingly comparable to real-life speech disorders are produced using mathematical models that vary the amplitude, frequency, and temporal elements of speech in order to simulate sick speech patterns.

In order to categorise the data, the acquired acoustic feature vectors are then entered into several supervised machine learning algorithms. Three classifiers—Gradient Boosting, Support Vector Machine (SVM), and Random Forest—were selected because of their ability to depict complex nonlinear connections on voice traits. There are a total of 80% training samples and 20% testing samples in the dataset, which include both real and synthetic speech. When developing a model, optimising the hyperparameters is a crucial step in ensuring the model can accurately identify speech impairments and improving the model's prediction performance. A trained classifier may use an input speech sample to differentiate between normal speech, stuttering, vocal tremor, and dysarthria.

Finally, R^2, RMSE, MAE, and MSE are widely used performance measures in regression analysis that are used to evaluate the constructed machine learning models. Comparative analysis provides an objective way to evaluate an algorithm's classification skills. The proposed approach combines supervised machine learning, acoustic feature extraction, synthetic data creation, and speech signal processing to help with accurate telemedicine-based clinical decision-making and diagnostic applications. The outcome is an intelligent and organised system for identifying speech problems.

V. EXPERIMENTAL RESULTS AND DISCUSSION

Through experimental assessment, we look at how well the suggested machine learning architecture can automatically diagnose speech abnormalities based on acoustic features extracted from audio recordings. The research employs a hybrid dataset consisting of both real clinical speech samples and artificially generated voice signals to extensively evaluate the built classification models in different speech contexts. Machine learning algorithms' ability to generalise across different problematic speech patterns is improved when synthetic data is generated since it increases the quantity of training data available to them.
All audio recordings undergo preprocessing steps prior to model training, including noise reduction, signal normalisation, segmentation, and acoustic feature extraction. A number of crucial speech metrics, including the fundamental frequency (F0), formant frequencies (F1 and F2), and others, are used to construct feature vectors that may distinguish between different types of speech disorders. The whole dataset is divided into two halves: 80% for training and 20% for testing. This way, the machine learning models can be evaluated fairly.

The three supervised ML algorithms—Gradient Boosting, Support Vector Machine (SVM), and Random Forest—are trained using the recovered audio features. Optimising hyperparameters improves classification performance while reducing prediction mistakes. Finally, the conditioned models categorise speech recordings into four categories: Normal Speech, Stuttering, Vocal Tremor, and Dysarthria. Performance assessments are conducted using several metrics like as MAE (Mean Squared Error), MSE (Mean Absolute Error), and R^2 (Coefficient of Determination) to provide a comprehensive picture of the reliability of the model and the precision of its predictions.

Based on experimental findings, the Random Forest classifier has the highest overall classification ability, with an accuracy of 62.50%. Most aberrant speech samples can be correctly identified by the model, and it maintains its balanced classification capability across different illness categories. Thanks to its ensemble learning strategy, which effectively handles nonlinear interactions among acoustic variables, it becomes more reliable and resilient in diagnostics. The impressive predictive potential of the Gradient Boosting classifier is shown by its 62.50% overall accuracy, which is comparable to other methods used for complex speech patterns. Experimental findings reveal that Gradient Boosting performs quite badly when categorising normal speech samples, suggesting a need for improvement in parameter tuning and training dataset size.
The Support Vector Machine (SVM) has a much worse overall classification accuracy (25.00%) compared to the ensemble learning models. Even while SVM works well for linearly separable datasets, the intricate nonlinear auditory characteristics of disordered speech reduce its classification skills in the experimental circumstances

used. When analysing heterogeneous speech signals with modest pathogenic variations, ensemble learning approaches seem to be more robust.

The comparative evaluation showed that Support Vector Machine was not as effective as Random Forest and Gradient Boosting in identifying speech problems using the collected acoustic features. When you combine machine learning with synthetic voice production, you get more diversified datasets and more reliable categorisation results. These findings demonstrate that mathematical modelling, machine learning, and speech signal processing may be used to create intelligent computer-assisted diagnostic systems that can help speech-language pathologists. These systems are able to objectively, accurately, and scalablely evaluate speech abnormalities.

VI. FUTURE ENHANCEMENT

There are a number of ways to make the suggested machine learning framework even more accurate, resilient, and useful for real-world use, even if it shows promise in the classification of speech problems. It is possible that in the future researchers may try to increase the size and diversity of the speech dataset by adding recordings from people of varying ages, sexes, languages, accents, and degrees of speech impairment severity. Machine learning models would be better able to learn and generalise to true clinical settings if they had access to a bigger dataset.

Extra acoustic characteristics beyond formant frequencies (F1 and F2) and fundamental frequency (F0) extraction is another significant improvement. Mel-Frequency Cepstral Coefficients (MFCCs), shimmer, jitter, harmonic-to-noise ratio (HNR), pitch fluctuation, energy, and speech pace are some of the speech features that may provide more thorough depictions of speech disorders. Classification accuracy and the ability to detect small speech anomalies might both be enhanced by using these state-of-the-art acoustic characteristics.

More complex Deep Learning architectures like Transformer-based speech models for automated feature learning from voice data, as well as RNNs, CNNs, and LSTM networks could be the subject of future research. Traditional machine learning algorithms could miss certain intricate spectral and temporal correlations that these models might reveal. In addition, hybrid methods that integrate deep learning with ensemble ML approaches have the potential to enhance classification accuracy and resilience even more.

Further integration of cloud computing, mobile healthcare apps, and telemedicine platforms allows the suggested architecture to be expanded into a real-time clinical decision-support system. Patients could record their voices from afar using these technologies, and doctors could listen to them to get a feel for their speech before they ever set foot in the clinic. In addition, XAI approaches may be used to explain model predictions in a clear and understandable way, which helps clinicians have more faith in the results and makes better diagnostic judgements. In the long run, these upgrades could turn the suggested framework into an all-inclusive AI speech analysis system that can diagnose speech disorders accurately, in real-time, and easily accessible, all while enhancing healthcare delivery and facilitating individualised speech therapy.

VII. CONCLUSION

Stuttering, vocal tremor, and dysarthria may all be automatically detected by this study's Machine Learning-based Speech Disorder Classification Framework, which analyses speech signals acoustically. To provide an efficient and objective diagnostic framework for abnormal speech classification, the suggested approach integrates supervised machine learning methods, acoustic feature extraction, synthetic data production, and speech signal preprocessing. A more diverse dataset and more robust classification model are the goals of the suggested method, which makes use of both naturally occurring and artificially produced speech samples.

Three supervised machine learning algorithms—Support Vector Machine (SVM), Random Forest, and Gradient Boosting—were trained using crucial acoustic characteristics collected from voice recordings. These parameters included fundamental frequency (F0), formant frequencies (F1 and F2), and more. The models were tested using metrics such as R^2, Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Their development followed an 80% training and 20% testing method. The experimental results showed that ensemble learning approaches performed better for pathological speech classification than support vector machines; the top classification accuracy was 62.50% for Random Forest and Gradient Boosting, and 25.00% for Support Vector Machine.

The results show that a computational method using machine learning may help speech-language pathologists diagnose communication impairments. Faster and more consistent detection of problematic speech patterns is made possible by automated analysis of voice signals, which also lowers the subjectivity involved with human clinical assessment. There is great promise for the use of intelligent diagnostic systems to facilitate telemedicine, early clinical intervention, and individualised treatment planning.

Finally, by combining acoustic feature analysis with supervised learning approaches, the suggested machine learning framework creates a solid basis for intelligent speech disorder detection. Future computer-aided speech evaluation systems should be able to improve diagnosis accuracy, accessibility, and patient care as a whole, and this research gives important insights into how AI is becoming more important in healthcare.

REFERENCES

1. L. R. Rabiner and R. W. Schafer, Digital Processing of Speech Signals. Englewood Cliffs, NJ, USA: Prentice Hall, 1978.
2. J. H. L. Hansen and T. Hasan, “Speaker recognition by machines and humans: A tutorial review,” IEEE Signal Processing Magazine, vol. 32, no. 6, pp. 74–99, Nov. 2015.
3. T. F. Quatieri, Discrete-Time Speech Signal Processing: Principles and Practice. Upper Saddle River, NJ, USA: Pearson, 2002.
4. L. Deng and D. O'Shaughnessy, Speech Processing: A Dynamic and Optimization-Oriented Approach. New York, NY, USA: Marcel Dekker, 2003.
5. H. Bourlard and N. Morgan, Connectionist Speech Recognition: A Hybrid Approach. Boston, MA, USA: Springer, 1994.
6. S. Davis and P. Mermelstein, “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences,” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-28, no. 4, pp. 357–366, Aug. 1980.
7. J. P. Campbell Jr., “Speaker recognition: A tutorial,” Proceedings of the IEEE, vol. 85, no. 9, pp. 1437–1462, Sep. 1997.
8. C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006.
9. L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
10. J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.
11. C. Cortes and V. Vapnik, “Support-vector networks,” Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.
12. F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
13. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.
14. D. Jurafsky and J. H. Martin, Speech and Language Processing, 3rd ed. Upper Saddle River, NJ, USA: Pearson, 2024.
15. B. H. Juang and L. R. Rabiner, “Automatic speech recognition – A brief history of the technology development,” Elsevier Encyclopedia of Language and Linguistics, 2005.
16. H. Hermansky, “Perceptual linear predictive (PLP) analysis of speech,” Journal of the Acoustical Society of America, vol. 87, no. 4, pp. 1738–1752, 1990.
17. M. A. Anusuya and S. K. Katti, “Speech recognition by machine: A review,” International Journal of Computer Science and Information Security, vol. 6, no. 3, pp. 181–205, 2009.
18. S. Furui, Digital Speech Processing, Synthesis, and Recognition, 2nd ed. Boca Raton, FL, USA: CRC Press, 2018.
19. A. Graves, A. Mohamed, and G. Hinton, “Speech recognition with deep recurrent neural networks,” in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), Vancouver, BC, Canada, 2013, pp. 6645–6649.
20. A. Vaswani et al., “Attention is all you need,” in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017.
21. D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in Proc. International Conference on Learning Representations (ICLR), 2015.
22. R. B. Dannenberg and M. Goto, “Music structure analysis from acoustic signals,” IEEE Signal Processing Magazine, vol. 31, no. 2, pp. 18–27, 2014.
23. M. Gales and S. Young, “The application of hidden Markov models in speech recognition,” Foundations and Trends in Signal Processing, vol. 1, no. 3, pp. 195–304, 2008.
24. D. P. Ramesh Babu and P. Krishna Rao, “Machine learning approaches for intelligent speech signal analysis and healthcare applications,” International Journal of Intelligent Systems and Applications in Engineering, vol. 12, no. 5, pp. 2618–2630, 2024.
25. S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland, The HTK Book (for HTK Version 3.4). Cambridge, U.K.: Cambridge University Engineering Department, 2009.