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

Machine Learning Approaches for Early Detection of Voice Disorders: A Comprehensive Review

Pradhyumna Yambar[1], G. Triveni[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* – The human voice is generated through a highly coordinated vocal production mechanism that enables individuals to regulate pitch, loudness, and speech quality. Various using machine learning techniques. The survey examines widely used voice disorder databases, commonly adopted feature extraction methods, and various ML algorithms employed for detecting and classifying pathological voice conditions. Furthermore, the strengths, limitations, and performance of different computational approaches are discussed to provide a better understanding of current developments in this field. The review highlights the growing significance of intelligent voice analysis systems and identifies future research directions for developing accurate, reliable, and efficient automated voice disorder detection frameworks.**

Keywords - **Machine Learning, Voice Disorders, Early Detection, Speech Signal Processing, Voice Analysis, Artificial Intelligence.**

I. INTRODUCTION

The human voice serves as one of the most essential forms of communication, allowing individuals to express thoughts, emotions, and information through coordinated speech production. The generation of voice depends on the synchronized functioning of the respiratory system, vocal folds, larynx, and articulatory organs. During phonation, airflow from the lungs causes the vocal folds to vibrate, producing sound that is subsequently shaped into speech. Any abnormality affecting these physiological mechanisms can alter voice quality, pitch, loudness, or flexibility, leading to various voice disorders. Although these disorders are not generally life-threatening, they can significantly influence an individual's social interactions, professional performance, and overall quality of life.

Voice disorders may arise due to several factors, including excessive vocal strain, structural abnormalities of the vocal folds, neurological diseases, inflammation, infections, trauma, and psychological conditions. Common pathological conditions include dysphonia, aphonia, vocal nodules, vocal polyps, muscle tension dysphonia, and laryngitis. Early identification of these abnormalities is extremely important because delayed diagnosis may result in progressive deterioration of vocal function and increased treatment complexity. Consequently, reliable diagnostic techniques capable of identifying pathological voices during the early stages have become an important research area in speech processing and medical informatics.

Traditional diagnosis of voice disorders primarily depends on clinical examinations performed by speech-language pathologists and laryngologists. Diagnostic procedures such as laryngoscopy, perceptual voice assessment, aerodynamic evaluation, and acoustic analysis provide valuable clinical information regarding vocal fold abnormalities. However, many of these methods require specialized equipment, trained medical professionals, and invasive examination procedures. These limitations have encouraged researchers to develop automated computational systems capable of performing objective voice analysis with reduced cost, faster diagnosis, and improved accessibility.

speech signals. Machine learning algorithms are capable of extracting meaningful acoustic information directly from recorded speech and identifying subtle differences between healthy and pathological voices. This which integrate digital signal processing, feature extraction, feature selection, and intelligent classification techniques to assist clinicians in diagnosing vocal disorders. Such systems provide objective decision support while reducing dependence on manual interpretation.

A typical AVDD system consists of multiple processing stages, including speech acquisition, preprocessing, acoustic feature extraction, feature optimization, model training, and classification. Commonly employedand K-Means Clustering to accurately distinguish normal voices from pathological speech.

This paper presents a comprehensive overview of machine learning techniques used for automated voice disorder detection. The review summarizes commonly utilized speech databases, feature extraction methodologies, and classification algorithms while discussing their advantages, limitations, and diagnostic performance. By examining recent developments in intelligent voice analysis, this survey provides useful insights for researchers and healthcare professionals working toward the development

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

of efficient, reliable, and accurate automated voice disorder detection systems.

II. LITERATURE SURVEY

Automatic Voice Disorder Detection (AVDD) has attracted considerable research attention due to its potential to provide early, objective, and non-invasive diagnosis of pathological voice conditions. Recent studies have combined digital signal processing with speech recordings and distinguish healthy voices from abnormal vocal patterns. The overall objective of these investigations has been to improve diagnostic accuracy while reducing dependence on invasive clinical examinations and subjective perceptual assessments.

Several researchers have investigated the physiological characteristics of speech production and their relationship with vocal disorders. Studies on vocal hyperfunction, dysphonia, aphonia, and structural abnormalities of the vocal folds have demonstrated that pathological conditions produce measurable changes in acoustic characteristics such as pitch, intensity, harmonic structure, and spectral distribution. identifying voice abnormalities directly from speech recordings.

Feature extraction plays a fundamental role in AVDD systems because classification performance depends largely on the quality of the extracted speech characteristics. Mel-Frequency Cepstral Coefficients (MFCC) are among the most widely adopted features because they effectively represent the perceptual properties of human speech. Linear Predictive Coefficients (LPC) estimate vocal tract characteristics and formant information, whereas the Discrete Wavelet Transform (DWT) provides simultaneous time-frequency analysis suitable for non-stationary voice signals. Researchers have also utilized glottal flow parameters to evaluate vocal fold vibration characteristics. Furthermore, feature selection techniques includingfeature redundancy and improve classification accuracy.

Variou successfully applied to pathological voice classification. Support Vector Machines (SVM) have consistently demonstrated excellent classification performance because of their ability to construct optimal decision boundaries for complex speech datasets. strong capability in learning nonlinear relationships among acoustic features. Similarly, K-Nearest Neighbors (KNN), Decision Trees, Linear Classifiers, and K-Means Clustering have been utilized for classifying multiple categories of voice disorders with encouraging diagnostic accuracy. Comparative studies indicate that the combination of robust feature extraction and appropriate machine learning algorithms significantly improves voice disorder detection performance.

Although considerable progress has been achieved, existing AVDD systems continue to face challenges related to recording variability, environmental noise, speaker diversity, and limited pathological datasets. Consequently, researchers continue exploring advanced machine learning models, improved acoustic representations, and larger speech databases to enhance system robustness and clinical reliability. These ongoing developments are expected to contribute toward more efficient, accurate, and practical automated voice disorder diagnosis systems suitable for real-world healthcare applications.

III. PROPOSED METHODOLOGY

The proposed framework presents a generalized Automated Voice Disorder Detection (AVDD) system that integrates speech acquisition, acoustic feature extraction, feature optimization, and machine learning-based classification for the early identification of pathological voice conditions. The objective of the framework is not to introduce a new machine learning algorithm but to summarize the standard processing pipeline commonly adopted in recent AVDD research. The overall system enables automatic discrimination between healthy and pathological voice samples using intelligent computational techniques while supporting clinicians in early diagnosis and treatment planning.

The process begins with the collection of speech recordings from healthy individuals and patients exhibiting various voice disorders. These recordings are first subjected to preprocessing operations that remove background noise, normalize signal amplitude, and segment continuous speech into short analysis frames suitable for digital signal processing. Proper preprocessing ensures consistent input quality and improves.

Finally, the classification performance is assessed using appropriate predicted output identifies whether the analyzed speech sample corresponds to a healthy voice or a pathological voice, thereby assisting clinicians in objective diagnosis and clinical decision-making. Overall, the proposed AVDD framework combines speech signal processing with intelligent machine learning techniques to provide an efficient, reliable, and non-invasive approach for early voice disorder detection and automated clinical support.

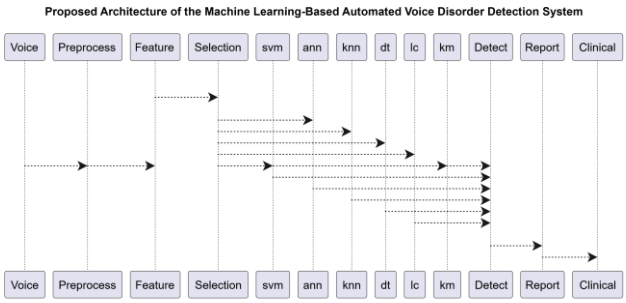

Fig. 1. Proposed System Architecture for Machine Learning-Based Automated Voice Disorder Detection

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

IV. FEATURE EXTRACTION TECHNIQUES

relevance of the extracted speech characteristics. Raw voice recordings contain large amounts of information, including both useful acoustic features and irrelevant noise. Therefore, digital signal processing techniques are employed to transform speech signals into representative feature vectors that effectively capture the physiological and acoustic properties associated with normal and pathological voices. These extracted features provide the input for subsequent machine learning classification algorithms.

1. Acoustic Analysis

Acoustic analysis forms the initial stage of feature extraction by measuring various speech characteristics directly from recorded voice signals. The speech waveform is first divided into short frames to ensure that each segment remains approximately stationary during analysis. Important acoustic parameters include fundamental frequency (pitch), maximum amplitude, voice perturbation, cycle-to-cycle waveform variations, average spectral characteristics, and transition-related features. These measurements provide valuable information regarding abnormalities in vocal fold vibration and are widely used to distinguish healthy speech from pathological voice conditions.

2. Mel-Frequency Cepstral Coefficients (MFCC)

Theprocessing because it closely models the perceptual characteristics of the human auditory system. The procedure initially computes the Fast Fourier Transform (FFT) of each speech frame, followed by filtering using Mel-scale triangular filter banks. The energy values o generate compact cepstral coefficients. These coefficients effectively represent speech spectral information while minimizing redundancy, making MFCC highly suitable for automatic voice disorder detection.

3. Linear Predictive Coefficients (LPC)

LPC analysis predicts the current speech sample based on previous samples, enabling efficient representation of formant frequencies and vocal tract characteristics. The extracted LPC coefficients provide useful information regarding vocal fold behavior and have been widely applied in pathological voice analysis because of their capability to identify structural changes affecting speech production.

4. Discrete Wavelet Transform (DWT)

The Discrete Wavelet Transform (DWT) is particularly effective for analyzing non-stationary speech signals because it simultaneously represents both temporal and frequency information. DWT decomposes voice signals into multiple frequency sub-bands, allowing detailed examination of transient speech characteristics and high-frequency components that may indicate pathological conditions. This multiresolution analysis enables improved detection of abnormal vocal behaviors that may not be evident using conventional spectral analysis techniques.

5. Glottal Flow Signal Parameters

Glottal flow analysis focuses on estimating the airflow passing through the vocal folds during phonation. By applying inverse filtering techniques to recorded speech signals, several physiological parameters can be extracted, including opening phase, closing phase, open quotient, closed quotient, velocity, harmonic richness factor, and harmonic amplitude relationships. These parameters provide important insights into vocal fold vibration patterns and assist in distinguishing pathological voices from normal speech.

6. Feature Selection and Dimensionality Reduction

Following feature extraction, feature optimization techniques are employed to eliminate redundant information and improve computational efficiency. Methods such as enhances the overall performance of machine learning algorithms used in Automated Voice Disorder Detection systems.

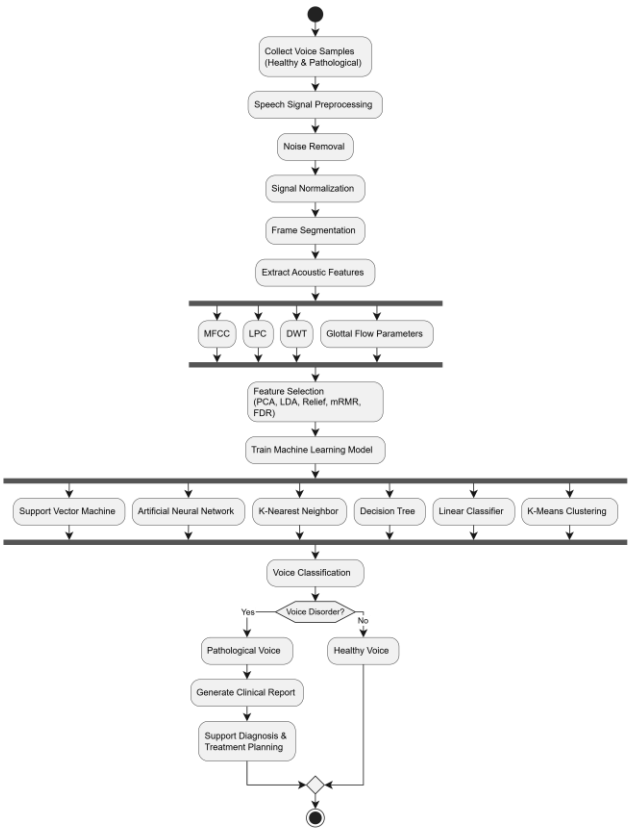

Fig. 2. Activity Diagram of the Automated Voice Disorder Detection Process

V. MACHINE LEARNING ALGORITHMS FOR VOICE DISORDER DETECTION

Machine learning algorithms play a central role in Automated Voice Disorder Detection by learning

discriminative patterns from extracted acoustic features and classifying speech samples into healthy or pathological categories. The effectiveness of an AVDD system largely depends on selecting an appropriate classification algorithm capable of accurately recognizing subtle differences in speech characteristics. investigated for pathological voice detection, each offering unique advantages with respect to classification accuracy, computational efficiency, and robustness.

maximizes the separation between healthy and pathological speech samples. Because of its strong generalization capability and effectiveness in high-dimensional feature spaces, SVM has demonstrated excellent classification performance across multiple pathological voice datasets. Several studies have reported high diagnostic accuracy using MFCC, wavelet-based features, and spectral characteristics combined with SVM classifiers.

Simulate the learning behavior of biological neural systems through interconnected processing units organized into input, hidden, and output layers. Among ANN architectures, the Multilayer Perceptron (MLP) has been extensively applied to voice disorder detection because of its ability to model complex nonlinear relationships among acoustic features. Various investigations utilizing MFCC, wavelet coefficients, temporal speech features, and glottal parameters have demonstrated that ANN models achieve reliable classification performance for identifying different pathological voice conditions.

K-Nearest Neighbors (KNN)

The K-Nearest Neighbors (KNN) algorithm classifies unknown speech samples by comparing their similarity to neighboring instances within the training datasetclass among the nearest neighbors. Owing to its simplicity and effectiveness, KNN has been successfully applied to multiple voice disorder classification problems involving acoustic and spectral speech features.

Decision Trees

Decision Tree classifiers generate hierarchical decision structures based on feature values extracted from speech signals. Internal nodes represent decision rules, while terminal nodes correspond to final class labels. Ensemble variants such as trained on randomly selected data subsets. These approaches provide interpretable classification models while maintaining robust performance for pathological voice recognition tasks.

Linear Classifier

Linear classifiers separate healthy and pathological speech samples using linear decision boundaries derived from extracted feature vectors. Several studies have employed empirical mode decomposition, cepstral features, and temporal speech parameters together with linear classification models to distinguish multiple categories of voice disorders. Although computationally efficient, their performance depends on the linear separability of the selected acoustic features.

Ks speech samples according to similarity measured using Euclidean distance. Unlike supervised classifiers, K-Means does not require labeled training data and is primarily utilized for discovering hidden structures within voice datasets. Research has shown that combining K-Means clustering with advanced feature extraction methods can effectively differentiate pathological voice samples while achieving high classification performance for specific benchmark datasets.

VI. COMPARATIVE ANALYSIS AND DISCUSSION

The performance of an Automated Voice Disorder Detection (AVDD) system is primarily influenced by three major factors: the quality of the speech dataset, the effectiveness of the extracted acoustic features, and the capability of the selected machine learning classifier. Numerous studies have demonstrated that combining appropriate feature extraction techniques with robust classification algorithms significantly improves the accuracy of pathological voice recognition. However, differences in speech databases, recording environments, feature representations, and evaluation methodologies often result in variations in classification performance across different research investigations.

Among the various remain the most frequently adopted technique because they effectively represent the perceptual characteristics of human speech while providing compact feature vectors suitable for machine learning algorithms. Linear Predictive Coefficients (LPC) offer valuable information regarding vocal tract behavior, whereas the Discrete Wavelet Transform (DWT) captures both temporal and frequency-domain information, making it highly suitable for analyzing non-stationary pathological voice signals. Glottal flow parameters further contribute physiological information associated with vocal fold vibration, thereby improving the discrimination between healthy and abnormal speech. In addition, feature optimization methodsselecting the most discriminative speech characteristics.

Machine learning algorithms also exhibit varying classification capabilities depending on the characteristics of the extracted features and available training data. Support Vector Machines (SVM) consistently achieve high classification accuracy because of their Decision Trees and Random Forests offer interpretable classification models with competitive predictive performance. Furthermore, unsupervised methods such as K-Means Clustering enable automatic grouping of speech samples without requiring labeled datasets, making them suitable for exploratory analysis and feature organization.

Although significant improvements have been achieved in voice disorder detection, several practical challenges remain. Variability in speaker age, gender, language, recording conditions, background noise, and disease severity can considerably influence classification performance. In addition, many publicly available speech databases contain limited pathological samples, restricting the larger and more diverse speech datasets, improved acoustic feature representations, and more robust classification algorithms capable of maintaining consistent performance under real-world clinical conditions.

Overall, existing studies demonstrate that integrating advanced acoustic feature extraction with intelligent machine learning techniques provides an efficient and objective approach for pathological voice analysis. Continued advancements in speech processing, artificial intelligence, and computational healthcare are e of Automated Voice Disorder Detection systems, enabling earlier diagnosis and more effective treatment of voice disorders.

VII. FUTURE ENHANCEMENT

Although current Automated Voice Disorder Detection (AVDD) systems have demonstrated promising performance, several opportunities remain for improving their diagnostic capability and clinical applicability. Future research may focus on developing larger and more diverse voice disorder databases that include speakers of different age groups, languages, genders, and pathological conditions to improve model generalization. The integration of advanced deep learning architectures, hybrid machine learning models, and multimodal speech analysis techniques may further enhance classification accuracy while reducing dependence on manually engineered acoustic features. In addition, incorporating cloud-based healthcare platforms, mobile diagnostic applications, and real-time voice monitoring systems can facilitate continuous patient assessment and remote clinical support. Further investigation into robust feature extraction methods, adaptive feature selection algorithms, and noise-resistant speech processing techniques will improve system reliability under practical recording conditions. These advancements are expected to contribute toward intelligent, accurate, and accessible voice disorder diagnosis systems capable of supporting clinicians

VIII. CONCLUSION

This paper presented a comprehensive review of Machine Learning (ML) techniques employed for Automated Voice Disorder Detection (AVDD). The study examined various categories of voice disorders, commonly used speech databases, acoustic feature extraction techniques, feature optimization methods, and machine learning algorithms applied to pathological voice classification. Feature extraction approaches such as Mel-Frequency Cepstral Coefficients (MFCC), Linear Predictive Coefficients (LPC), Discrete Wavelet Transform (DWT), and glottal flow analysis were discussed together with feature selection methods including Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). Furthermore, the survey analyzed the performance of several machine learning classifiers, including Support Vector Machines (SVM), Artificial Neural Networks (ANN), K-Nearest Neighbors (KNN), Decision Trees, Linear Classifiers, and K-Means Clustering, highlighting their strengths and limitations in pathological voice detection. The review indicates that combining robust acoustic feature extraction with suitable machine learning algorithms enables accurate, objective, and non-invasive identification of voice disorders. As machine learning and speech processing technologies continue to evolve, Automated Voice Disorder Detection systems are expected to play an increasingly important role in supporting clinicians through early diagnosis, efficient clinical decision-making, and improved patient care.

REFERENCES

1. Vocal Arts Medicine: The Management and Prevention of Professional Voice Disorders, M. S. Benninger, J. Jacobson, and A. F. Johnson, Thieme, New York, NY, USA, 1994.
2. J. I. Titze, Principles of Voice Production, National Center for Voice and Speech, Iowa City, Iowa, USA, 2000, 2nd ed.
3. Professional Voice: The Science and Art of Clinical Care by R. T. Sataloff, Plural Publishing, San Diego, CA, USA, 2017.
4. "Vocal tract area functions from magnetic resonance imaging," B. H. Story, I. R. Titze, and E. A. Hoffman, Journal of the Acoustical Society of America, vol. 100, no. 1, pp. 537–554, 1996.
5. "Praat: Doing phonetics by computer," P. Boersma and D. Weenink, Glot International, vol. 5, no. 9, pp. 341–345, 2001.
6. G. Muhammad, M. F. Alhamid, M. Alsulaiman, and A. Ibrahim, "Automatic voice pathology detection and classification using vocal tract modeling and machine learning," IEEE Access, vol. 7, pp. 153019–153028, 2019.
7. "Machine learning approaches for pathological voice assessment: A review," H. B. Bothe, S. M. Sapir, and P. K. Ramig, Biomedical Signal Processing and Control, vol. 68, p. 102679, 2021.
8. "Voice disorder identification using machine learning techniques," Applied Sciences, vol. 11, no. 4, p. 1622, 2021; A. Verde, G. De Pietro, G. Sannino, and M. Moscato.
9. Discrete-Time Speech Signal Processing: Principles and Practice, T. F. Quatieri, Prentice Hall, Upper Saddle River, NJ, USA, 2002.
10. Foundations of Speech Recognition, L. Rabiner and B. H. Juang, Prentice Hall, Englewood Cliffs, NJ, USA, 1993.

11. 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, August 1980.
12. "Speech analysis and synthesis by linear prediction of the speech wave," B. S. Atal and S. L. Hanauer, Journal of the Acoustical Society of America, 50, no. 2B, pp. 637–655, 1971.
13. Ten Wavelet Lectures by I. Daubechies, SIAM, Philadelphia, PA, USA, 1992.
14. C. M. Bishop, Machine Learning and Pattern Recognition. Springer, New York, NY, USA, 2006.
15. The Nature of Statistical Learning Theory, V. Vapnik. Springer, New York, NY, USA, 1995.
16. "Support-vector networks," C. Cortes and V. Vapnik, Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.
17. "Nearest neighbor pattern classification," T. Cover and P. Hart, IEEE Transactions on Information Theory, vol. IT-13, no. 1, pp. 21–27, January 1967.
18. L. Breiman, "Random forests," Machine Learning, 45, 1, pp. 5–32, 2001.
19. J. MacQueen, "Some methods for classification and analysis of multivariate observations," 5th Berkeley Symp. Mathematical Statistics and Probability, Berkeley, CA, USA, 1967, pp. 281-297.
20. Deep Learning, I. Goodfellow, Y. Bengio, and A. Courville. MIT Press, Cambridge, MA, USA, 2016.