

Exploring the Intersection of Machine Learning and Voice Pathology: A Comprehensive Review

1 Divya O M

Dept. of Computer Science
CHRIST University (Deemed to be)
Bangalore, India
divya.om@christuniversity.in

2 Sagaya Aurelia P

Dept. of Computer Science
CHRIST University (Deemed to be)
Bangalore, India
sagaya.aurelia@christuniversity.in

Abstract—Voice disorders may significantly impact an individual's quality of life and are often associated with occupational, social or health challenges. In recent years, the integration of machine learning techniques has helped in the field of voice pathology to make significant advancements. This comprehensive review explores the multifaceted relationship between machine learning and voice pathology, shedding light on how these innovative technologies improve the diagnosis, assessment, and understanding of voice disorders. The paper outlines various machine learning models deployed to distinguish between healthy and pathological voices. In this paper, a study is made on different machine learning techniques and feature extraction methods used in the literature for defining the characteristics of the voice and how they are used as input for machine learning algorithms for predicting the pathological voice from the normal one. The review addresses limitations such as the need for standardized datasets and the interpretability of machine learning models. It also highlights ongoing research directions, including integrating real-time monitoring devices to enhance the trustworthiness of machine learning-based diagnostic outcomes. Based on this study, it could be observed that classification and prediction techniques could be efficiently applied over the voice for pathology prediction with more accuracy and less human intervention.

Index Terms—Voice disorder, Machine learning, Classification, Prediction

I. INTRODUCTION

Technological advancements have the potential to significantly improve and support the various decision-making processes in the healthcare system. This may motivate people to adopt healthy habits and keep track of their health to promote early identification of any illness. Communication is essential in determining a person's physical and mental status, so voice pathology identification is vital in various healthcare applications[1]. With modern techniques, such as indirect laryngoscopy (Fig. 1), video laryngoscopy, stroboscope light, and the professional's ear, it is necessary to subjectively identify problems with the larynx and vocal folds, leading to a qualitative evaluation of these structures. Also, Medical professionals use grade, roughness, breathiness, asthenia, strain (GRBAS), and Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) scales to evaluate speech [3]. In the category of voice disorders, structural disorders are the most common. As technology advances, tools are developed to discern a normal voice from a pathological one through acoustical analysis[4].

The field of automatic detection has witnessed significant growth due to thorough study and analysis in acoustic recording. However, present automated voice pathology detection methods typically find it challenging to expand their scope beyond the limits of the initial training environment or incorporate additional related applications. However, machine learning algorithms have recently been extensively used to calculate the audio signal properties of the input signal to identify pathological voices[5]. Although some models claim to have accuracy rates of over 95%, it is yet to be determined if they can be applied across diverse datasets.

This paper aims to tackle this issue by utilizing the PVQD dataset, a relatively new resource, to investigate correlations among statistical parameters. In this work we tried to identify the relevant features for classifying and predicting normal and pathological voices using logistic regression models.

Despite ample research and data, finding an accurate solution is still challenging. The risk factor is combining audio signal processing and machine learning to produce accurate results. The essential elements include the data recording process, feature extraction and selection, and advancements in classification techniques[6].

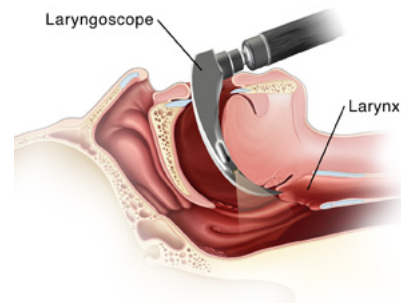


Fig. 1. Laryngoscopy[2]

A. Voice Disorders

Voice disorders like dysphonia require a thorough examination of the larynx and vocal folds for effective treatment[7]. They fall into three categories: functional, neurological, and structural.

B. Structural Disorders

Disorders can be caused by lesions, injuries, improper usage, infections, growths, or conditions like laryngitis and cysts[8] Fig. 2 depicts the examples of structural disorders. The sulcus on the mucosal surface of the vocal folds can also affect your voice [11]. Analyzing audio signals can help diagnose dysphonia.



Fig. 2. Example of Structural Disorders

C. Functional Disorders

Functional voice disorders emerge as phonation-impairing non-organic abnormalities [12]. Muscle tension voice disorder (MTVD) and psychogenic voice disorder (PVD) are the main types of functional voice disorders[13].

D. Neurological Disorders

Neurological disorders include those arising from the central nervous system, the peripheral nervous system, and functional or behavioural[14]. Some examples are adductor palsy and spasmodic dysphonia[5].

E. Datasets

1) *Saarbruecken Voice Database(SVD)*: The SVD dataset has about 2000 German audio recordings of healthy and pathological subjects. The recordings were captured at a 16-bit resolution and 50 Hz frequency [21]. The dataset includes vowels [i, a, u] at normal, high, and low pitches and recordings of those three vowels with rising-falling pitch. There is also a recording of the sentence "Guten Morgen, wie geht es Ihnen?" However, there are fewer healthy samples than pathological ones[15].

2) *Massachusetts eye and ear infirmiry(MEEI)*: Developed by MEEI Voice and Speech Lab, this document includes the first part of the Rainbow passage and more than 1,400 vocal tests for the long vowel. The response frequency for standard samples was 25 or 50 kHz, whereas the sampling frequency was 50 kHz. However, concerns have been raised about its precision due to recordings from people with different health conditions and speaking environments, making it uncertain if it can distinguish between pathology and environmental factors[16].

3) *Arabic voice pathology database(AVPD)*: Data on three vowels (/a/, /u/, and /i/) was collected by recording voices reading a passage and three words while counting from 0-10. The Computerized Speech Lab model 4500, which operates at a frequency of 48 kHz, was used to record the dataset. The dataset includes 51% normal samples and pathological samples

TABLE I
DATASETS BASED ON VOICE PATHOLOGY

Dataset	Voice Samples	Phonemes	Language	Diseases
SVD[21]	1354 pathological 687 healthy	/a/,/i/,/u/, "Guten Morgen, wie geht es Ihnen?"	German	various pathologies
MEEI[22]	573 pathological 53 healthy	/a/	English	various pathologies
AVPD [23]	178 pathological 188 healthy	/a/,/u/ & /i/ 0-10,reading passage	Arabic	vocal fold cysts, nodules, paralysis,polyps &sulcus
VOICED [24]	150 pathological 58 healthy	/a/	English	hyper & hypo kinetic dysphonia, flux laryngitis
PVQD [25]	207 pathological 89 healthy	/a/,/i/, reading words	English	dysphonia
AVFAD[26]	346 pathological 363 healthy	/a/,/i/,/u/, reading of sentences	Portuguese	26 vocal pathologies

with conditions such as paralysis, sulcus, cysts, polyps, and nodules.[17].

4) *Voice ICarf E Dericoll Database (VOICED)*: A study recorded 5-second vowel sounds from 73 males and 135 females to identify voice disorders. The dataset includes 3 vocal diseases and covers individuals up to 80 years old.[18].

5) *Advanced Voice Function Assessment Database(AVFAD)*: AVFAD is an open access resource based on a sample of 346 patients clinically diagnosed with vocal pathology and 363 individuals with no vocal alterations. It contains sustaining vowels /a/, /i/, /u/, reading of sentences, reading of balanced text and spontaneous speech[26].

6) *Perceptuel Voice Qualit s Database (PVQD)*: The Voice Foundation's PVQD project comprises 296 audio files that feature continuous speech phrases requiring a thorough review and editing process for optimal machine learning use. The report includes pertinent information such as gender, age, diagnosis, and assessments utilizing the GRBAS and CAPE-V scales. These scales adhere to the CAPE-V standards for sustained /a/ and /i/ vowels. The audio files are encoded using a high-quality 16-bit at a 44.1 kHz sample rate[19].

II. A COMPREHENSIVE REVIEW OF RECENT STUDIES USING THE SVD DATASET

Table II shows the summary of the recent IEEE published works in voice pathology detection and classification using

TABLE II
SUMMARY OF RECENT WORKS IN VOICE PATHOLOGY CLASSIFICATION USING SVD DATASET

No	Ref paper id	Objective	Selected Features	Models Used	Accuracy
1	[22]	Voice Quality analysis using MDVP parameters (IEEE) 2023	MDVP	XGBoost classifier	98%
2	[23]	Using SVM classify the voice pathology	Peaks, Zero Crossing Rate (ZCR), Fundamental Frequency (F0), Jitter, Max	SVM	92 %
3	[24]	Using Bi-LSTM classify the pathological voice	All-Pole Group Delay Function (APGDF), MFCC, CQCC	Bidirectional long short-term memory (Bi-LSTM)	92.80 %
4	[25]	Develop an automated voice detection system	bark spectrum	KNN,SVM, ensemble model, and NNmodel	94.45%(KNN), 90.00 %(SVM), 95.60 %(Ensemble), 93.35%(NN)
5	[6]	Detection of pathological speech using DNN	Spectrogram	DNN, VGG16 CNN based transfer learning	82%
6	[9]	Assess the Dysarthrophic Voice	RASTA-PLP	ANN	100%
7	[27]	Using CNN identify the voice disorder	CNN	DNN, VGG16 CNN transfer learning, CaffeNet Models	94.40%
8	[28]	AVDD using self supervised methods		DNN	80.71% - Phrases 82.8% - Vowels
9	[29]	Conversion from Pathological to Normal Voice using E-DGAN	Spectrogram	ED GAN	18.67%
10	[30]	Using Multiresolution Time Series Classify the voice pathology	deep feature extraction	SVM, XCM, MRTSCN	90%
11	[31]	Pathological VDD Using CNN [31]	MFCC	CNN	23% more than the existing results
12	[32]	Using Naive Bayes Detect Dysphonia	MFCC	Naive Bayes	81.48%
13	[33]	Using multiple features analyse the voice pathology	Chroma, mel spectrogram, MFCC	DNN	96.77%
14	[34]	Using OSELM detect and classify the voice pathology	MFCC	OSELM	89.47%
15	[26]	Combining various signal processing techniques classify the voice disease	MFCC and RASTA-PLP	FFNN classifier with SCG algorithm	80.7%

SVD dataset. Liza et al. used Python to create MDVP parameters and differentiate between healthy and pathological voices using XGBoost. They achieved an accuracy rate of 98% and focused on polyps, nodules, and Reinke Edema [22].

Mohamed et al. used a set of various voice features. These voice features were extracted from the voice samples, and then the classification was performed using SVM [23].

Anilkumar et al. makes use of three different feature extraction methods. With the CQCC feature extraction approach and the BI-LSTM classification model, the system achieved a remarkable accuracy rate of 92.7% [24].

Kumar et al. presented in this study a novel method which compared various classification models in discriminating pathological voices from normal voices. The authors claim that the accurate model for classifying voice signals is the ensemble model, which places significant emphasis on the back spectrum [25].

Krishnan et al., in their research, uses DNN to identify pathological speech through the VGG16 convolutional neural network and transfer learning method. The networks, trained on different vowel subsets, are merged into an ensemble model, achieving 82% accuracy. Pre-trained CNNs are effective, eliminating the need for extensive data and language level restrictions [6].

With the help of a case study Muhammad et al. explained the role of AI and IOT [9].

Ribas et al. explore using Self-Supervised representation learning in an AVDD system to distinguish between normal and pathological speech. The study achieved accuracy levels of up to 93.36%, providing valuable insights for improving voice disorder diagnosis and treatment [27].

The E-DGAN technique has been developed by Chu et al. to treat voice disorders. It incorporates an encoder-decoder structure and a new assessment measure known as "content similarity". Its successful testing proves its potential for personalized voice conversion[28].

Zhao et al. have developed a network that uses neural networks to identify and classify vocal disorders. The system analyzes speech data and has shown a 17% improvement in distinguishing healthy voices from those with issues[29].

The developed method by Islam et al. successfully distinguished between dysphonic and normal voices with over 23% accuracy, outperforming existing methods[30].

Another similar work, presented by Al-Dhief et al., detects Dysphonia Disease from the voice samples using Naïve Bayes classifier. The proposed method achieved an accuracy of 81.48%, 65% of sensitivity, a specificity of 91.17%, and a 76.98% G-mean when used with the MFCCs as voice features [31].

A study created a precise model for predicting voice-related disorders using a DNN. The model achieved 77.49% accuracy and 96.77% when trained on specific illnesses. Gender information improved accuracy to 88.01%. The study focused on vowels and "continuous sentence" audio files [32].

Using MFCCs, OSELM has shown an accuracy rate of 91.17%, with impressive precision and recall rates. OSELM performs well in terms of f-measure (87%), G-mean (87.55%), and specificity (97.67%) as well [33].

Jayasree et al. focus on pathological voice detection and classification in Autism Spectrum Disorder and Down Syn-

drome disorders using signal processing based methodologies and Feed Forward Neural Network. The proposed methodology compared the data with perturbation measures, noise parameters and cepstral features. [34].

Fang, Shih-Hau, et al. evaluated the performance of DNN in collected voice samples from private hospitals and publically available MEEI datasets. Three variants of MFCCs (MFCC, MFCC + delta and MFCC(N) + delta) were used for the feature extraction. DNN with MFFCC(N) + Delta achieved the highest accuracy of 94.26%.

Studies have shown impressive accuracy rates for identifying vocal disorders using machine learning. One study found a 98.3% accuracy rate using an LSTM classifier, while another saw promising results with a 94.4% accuracy rate using pre-existing CNN models and transfer learning. Combining MFCCs and glottal signal parameters also produced better outcomes [26].

A. Derived Features from the literature review and its implications

In this section, we discuss the findings of our literature study on the features extracted using audio signal processing and their implications. Jitter is determined by identifying the timing of glottal pulses and calculating the average absolute difference between consecutive periods over a period of N. This can be used to detect voice pathology by comparing it to a threshold value of 83.2 μ s in adults.

$$\text{Jitter} = \frac{1}{N-1} \sum_{i=1}^{N-1} |\text{Timestamp}[i] - \text{Timestamp}[i-1]| \quad (1)$$

Shimmer is the average absolute difference, expressed as a fraction of the average amplitude, between the amplitudes of two successive periods T_i and T_{i+1} . The threshold for finding pathologies for this metric is set at 3.81%.

$$\text{Shimmer (local)} = \frac{1}{N} \sum_{i=1}^{N-1} \frac{|A_i - A_{i+1}|}{\frac{1}{N} \sum_{i=1}^N A_i} \quad (2)$$

Where:

Shimmer (local) : Local shimmer measure

N : Number of consecutive periods

A_i : Amplitude of the i th period

Let $x(t)$ represent the voice signal. The maximum value of the voice signal is given by:

$$\max(x(t)) = \max_t x(t) \quad (3)$$

Where:

$\max(x(t))$: Maximum value of the voice signal

t : Time

Zero Crossing Rate(ZCR)

Voiced speech has energy concentrated below 3 kHz, while unvoiced speech is more common in higher frequencies. Frequency is related to zero-crossing rate and energy distribution, with higher frequencies having higher zero-crossing rates and lower frequencies having lower ones.

$$\text{ZCR}_i = \frac{1}{2(N-1)} \sum_{n=1}^{N-1} |x_i[n] - x_i[n-1]| \quad (4)$$

Linear Frequency Cepstral Coefficient (LFCC)

A filter bank with 40 equally spaced filters is used to achieve the desired frequency range of 133-6857 Hz. Each filter has a bandwidth of 164 Hz. The signal undergoes DFT, triangle filtering, logarithmic compression, and DCT to obtain LFCC FB-40 parameters.

Mel-frequency cepstral coefficients (MFCC)

Because the MFCC transforms a logarithmic spectral representation, it is closely related to the calculation of the cepstrum. The steps to determine MFCC coefficients are as follows. The signal is divided into brief frames, multiplying by a Hamming Window. Calculate the power spectrum for each windowed frame, the FFT for the windowed frames, and the Mel filter bank. Calculate the logarithm, the Mel filter bank, and the discrete cosine transform (DCT) before applying them to the power spectra.

$$\text{MFCC}(n) = \text{DCT} \left(\log \left(\sum_{m=0}^{N-1} |X(n, m)|^2 \cdot H(m) \right) \right) \quad (5)$$

$$H(m) = \begin{cases} 1, & \text{if } m = 0 \\ 2, & \text{otherwise} \end{cases}$$

The Continuous Wavelet Transform (CWT)

of a finite energy signal $x(t)$; $w_x(a, b)$ represents the Continuous Wavelet Transform at scale a and position b .

$$W_x(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t) \cdot \Psi^* \left(\frac{t-b}{a} \right) dt \quad (6)$$

where b is called the translation parameter, which determines the time location of the wavelet, and a is the scaling parameter.

$$W_x(a, b) = \langle x(t), \phi_{a,b}(t) \rangle \quad (7)$$

where $w_x(a, b)$ are many wavelet coefficients and is a function of scale and position.

$$x(t) = \sum_n c_{j,n} \phi_{j,n}(t) + \sum_n d_{j,n} \psi_{j,n}(t) \quad (8)$$

B. Applied Machine Learning Models and Algorithms

Speech samples are classified as normal or pathological through feature extraction, selection, and classification using a classifier like k-NN or SVM. SVMs are preferred for creating an optimal boundary between the two classes in the feature space. Nonlinear classifiers like HMM and neural networks are also utilized[37].

III. UNCOVERING PATTERNS AND DIAGNOSTIC SIGNIFICANCE OF PVQD DATASET

After analyzing the PVQD dataset, we found diagnostic patterns that clarify the relationship between acoustic aspects of the voice and specific traits. These findings not only deepen our understanding of voice-related illnesses but also hold the promise of profoundly altering medical diagnostic procedures.

Pearson's Correlations

Variable	Age	Severity	Breathiness	Loudness	Pitch	Roughness	Strain
1. Age	Pearson's r —						
2. Severity	Pearson's r 0.511	—					
3. Breathiness	Pearson's r 0.396	0.872	—				
4. Loudness	Pearson's r 0.450	0.888	0.893	—			
5. Pitch	Pearson's r 0.478	0.860	0.776	0.863	—		
6. Roughness	Pearson's r 0.457	0.811	0.625	0.660	0.749	—	
7. Strain	Pearson's r 0.455	0.889	0.695	0.803	0.827	0.753	—

Fig. 3. Correlation between the variables in PVQD dataset

We will highlight the main conclusions, relationships, and statistical analyses that form the foundation of our study in this section.

The correlation table clearly displays the relationship between Age, Severity, Breathiness, Loudness, Pitch, Roughness, and Strain. Of particular significance, Loudness and Severity share a strong positive correlation. This implies that an increase in volume is directly linked to a more severe condition. Additionally, severity has moderate positive correlations with Pitch ($r \approx 0.860$) and Strain ($r \approx 0.889$), suggesting that as severity increases, pitch and strain tend to increase as well.

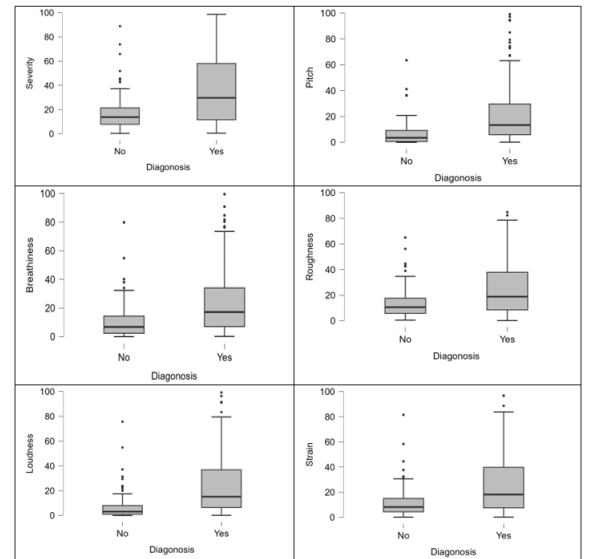


Fig. 4. Diagnosis ratio among specific acoustic features

The presence of a disease is accompanied by an increase in specific acoustic features of the voice. From Fig 4 it is clear that the study conclusively demonstrates discernible variations in these features when specific attributes are either present or absent in the voice.

Based on the model's performance metrics, shown in the confusion matrix in Fig 7, which include accuracy, precision, recall, specificity, and F1-score, we can assess its effectiveness in binary classification. The model's accuracy is around 76.35%, which means it can make correct predictions for most cases. All the combinations were conducted with an independent

sample T-test and confirmed the significant difference among the pairs (for all $p < 0.05$).

Logistic regression analysis which is shown in Fig.5 and 6 yields valuable insights into the diagnostic practices of medical professionals, shedding light on the key factors that influence a positive diagnosis.

Model Summary - Diagnosis

Model	Deviance	AIC	BIC	df	X ²	p	Adjusted R ²
H ₀	389.544	391.544	395.235	295			
H ₁	289.574	307.574	340.787	287	99.971	< .001	0.392

Fig. 5. Logistic Model Summary on PVQD dataset

Coefficients

	Estimate	Standard Error	Odds Ratio	z	Wald Test		
					Wald Statistic	df	p
(Intercept)	-1.656	0.360	0.191	-4.598	21.139	1	< .001
Age	0.033	0.008	1.033	4.102	16.828	1	< .001
Severity	-0.048	0.029	0.953	-1.622	2.630	1	0.105
Breathiness	0.009	0.023	1.009	0.399	0.159	1	0.690
Loudness	0.069	0.025	1.071	2.778	7.720	1	0.005
Pitch	0.028	0.027	1.028	1.045	1.091	1	0.296
Roughness	0.004	0.021	1.004	0.187	0.035	1	0.852
Strain	0.018	0.024	1.018	0.744	0.553	1	0.457
Gender (Male)	0.644	0.319	1.904	2.020	4.079	1	0.043

Note. Diagnosis level 'Yes' coded as class 1.

Fig. 6. Coefficients

As per the Fig 8, Age, Loudness, and Gender (Male) significantly impact our logistic regression model for classifying voice disorders, as indicated by their odds ratios and Wald statistics. The severity is also can be considered.

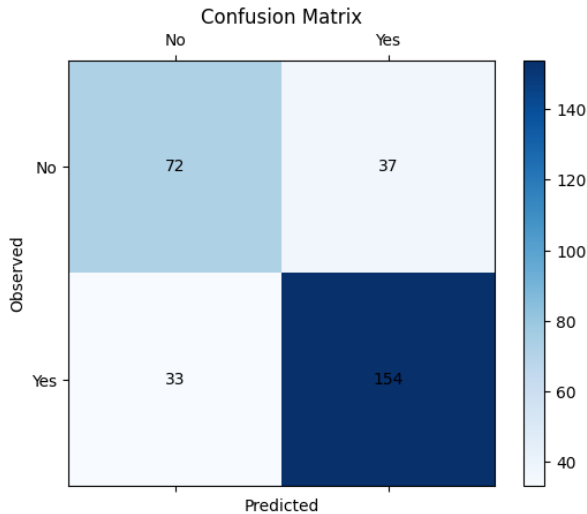


Fig. 7. Confusion Matrix

IV. INFERENCES AND CONCLUSION

The literature review revealed a diversity of voice databases, ranging in size and shape. The controlling environmental

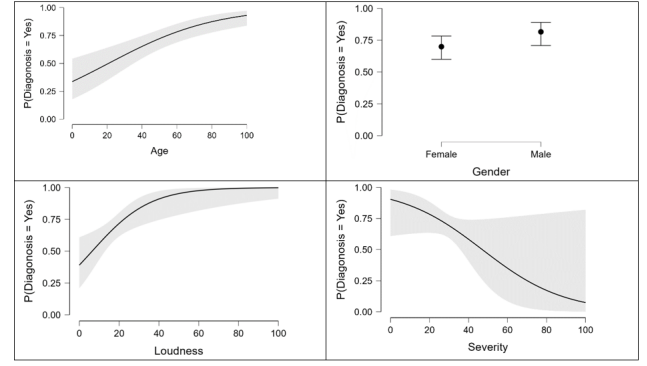


Fig. 8. Predicted features

factors impact voice recording for data collection, making using standardised acquisition techniques necessary. Before recording, the participants must be trained on various aspects like clear pronunciation, constant sound intensity, and recording protocols. It is impertinent to extract appropriate and relevant information from speech signals. Machine learning models use these features as their input, and the prediction directly impacts how effectively they function. It is essential to select a suitable algorithm, for it influences the success of the classification of vocal pathology. Anti-interference verification is essential to ensure the accuracy and dependability of the data.

Although the analysis of PVQD dataset emphasizes the importance of the correlations found between acoustic features and specific traits, the model needs improvement to discover more. However, this research effort aims to harness the findings and methodologies developed herein to apply these features for real-time voice classification. Furthermore, we endeavour to shed light on assessing the severity of these disorders in real-world datasets. This work represents a significant step towards effectively enhancing our ability to address voice-related conditions in clinical practice. Future studies should explore extracting more dimensions and traits from the dataset, experimenting with different machine learning models and training methods, and allocating evenly among vocal problems. Accuracy is crucial in diagnosing voice pathology, so selecting optimal feature extraction and classifiers is essential.

REFERENCES

- [1] Milani, MG Manisha, Murugaiya Ramashini, and Murugiah Krishani. "A real-time application to detect human voice disorders." DASA. IEEE, 2020.
- [2] <https://mountnittany.org/wellness-article/laryngoscopy-direct-rigid>
- [3] Sharma, Garima, Kartikeyan Umaphathy, and Sridhar Krishnan. "Trends in audio signal feature extraction methods." Applied Acoustics 158 (2020): 107020.
- [4] de Oliveira Rosa, Marcelo, José Carlos Pereira, and Marcos Grellet. "Adaptive estimation of residue signal for voice pathology diagnosis." IEEE Transactions on Biomedical Engineering 47.1 (2000): 96-104.

- [5] Hegde, Sarika, et al. "A survey on machine learning approaches for automatic detection of voice disorders." *Journal of Voice* 33.6 (2019): 947-e11.
- [6] Krishnan, Sridhar, and Yashodhan Athavale. "Trends in biomedical signal feature extraction." *Biomedical Signal Processing and Control* 43 (2018): 41-63.
- [7] Al-Nasheri, Ahmed, et al. "Investigation of voice pathology detection and classification on different frequency regions using correlation functions." *Journal of Voice* 31.1 (2017): 3-15.
- [8] Orozco-Arroyave, Juan Rafael, et al. "Characterization methods for the detection of multiple voice disorders: neurological, functional, and laryngeal diseases." *IEEE journal of biomedical and health informatics* 19.6 (2015): 1820-1828.
- [9] Muhammad, Ghulam, and Musaed Alhussein. "Convergence of AI and IOT in smart healthcare: a case study of voice pathology detection." *Ieee Access* 9 (2021): 89198-89209.
- [10] Al-Nasheri, Ahmed, et al. "An investigation of MDVP parameters in three different databases for voice pathology detection and classification." *Journal of Voice* 31.1 (2017): 113-e9.
- [11] <https://my.clevelandclinic.org/health/body/24456-vocal-cords>
- [12] Duffy, J. R. "Functional speech disorders: clinical manifestations, diagnosis, and management." *Handbook of clinical neurology* 139 (2016): 379-388.
- [13] Chung, David S., et al. "Functional speech and voice disorders: case series and literature review." *Movement disorders clinical practice* 5.3 (2018): 312-316.
- [14] Wang, Tiffany V., and Phillip C. Song. "Neurological voice disorders: a review." *International Journal of Head and Neck Surgery* 13.1 (2022): 32-40.
- [15] Saarbruecken Voice Database—Handbook, Stimmdatenbank.coli.uni-saarland.de. [Online].
- [16] M. OpenCourseWare, Lab Database — Laboratory on the Physiology, Acoustics, and Perception of Speech — Electrical Engineering and Computer Science — MIT OpenCourseWare, Ocw.mit.edu. [Online].
- [17] <https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-542j-laboratory-on-the-physiology-acoustics-and-perception-of-speech-fall-2005/lab-database>
- [18] <https://www.hindawi.com/journals/jhe/2017/8783751> [19] <https://doi.org/10.1016/j.jvoice.2020.10.001>
- [20] Mesallam, Tamer A., et al. "Development of the AVPD and its evaluation by using speech features and machine learning algorithms." *Journal of healthcare engineering* (2017).
- [21] Martínez, David, et al. "Voice pathology detection on the Saarbrücken voice database with calibration and fusion of scores using multifocal toolkit." *IberSPEECH 2012 Conference, Madrid, Spain*
- [22] Liza, Rahmi, and Chen-Kun Tsung. "Analyzing Voice Quality with MDVP for Disease Determination." *IS3C. IEEE*, 2023.
- [23] <https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-542j-laboratory-on-the-physiology-acoustics-and-perception-of-speech-fall-2005/lab-database/>
- [24] AnilKumar, V., and R. Venkata Siva Reddy. "Classification of voice pathology using different features and Bi-LSTM." *ICSSSES.IEEE*, 2023.
- [25] Kumar, Deepak, Udit Satija, and Preetam Kumar. "Automated Classification of Pathological Speech Signals." *INDICON. IEEE*, 2022.
- [26] Zakaria, Salman, S. Thanush, and M. Mugilan. "Voice Disorder identification Using Convolutional Neural Network." *ICCST. IEEE*, 2022.
- [27] Ribas, Dayana, et al. "Automatic Voice Disorder Detection Using Self-Supervised Representations." *IEEE Access* 11 (2023): 14915-14927.
- [28] Chu, Minghang, et al. "E-DGAN: An Encoder-Decoder Generative Adversarial Network Based Method for Pathological to Normal Voice Conversion." *IEEE Journal of Biomedical and Health Informatics* (2023).
- [29] Zhao, Denghuang, et al. "Pathological Voice Classification Using Multiresolution Time Series Classification Network." *ICSMD. IEEE*, 2022.
- [30] Islam, Rumana, Esam Abdel-Raheem, and Mohammed Tarique. "Deep Learning Based Pathological Voice Detection Algorithm Using Speech and Electroglottographic (EGG) Signals." *(ICECTA). IEEE*, 2022.
- [31] Al-Dhief, Fahad Taha, et al. "Dysphonia Detection Based on Voice Signals Using Naive Bayes Classifier." *ISTT. IEEE*, 2022.
- [32] Zakariah, Mohammed, et al. "An Analytical Study of Speech Pathology Detection Based on MFCC and Deep Neural Networks." *Computational and Mathematical Methods in Medicine*(2022).
- [33] Al-Dhief, Fahad Taha, et al. "Voice pathology detection and classification by adopting online sequential extreme learning machine." *IEEE Access* 9 (2021): 77293-77306.
- [34] Jayasree, T., and S. Emerald Shia. "Combined Signal Processing Based Techniques and Feed Forward Neural Networks for Pathological Voice Detection and Classification." *Sound & Vibration* 55.2 (2021).
- [35] Islam, Rumana, Mohammed Tarique, and Esam Abdel-Raheem. "A survey on signal processing based pathological voice detection techniques." *IEEE Access* 8 (2020): 66749-66776.
- [36] Wahed, Manal Abdel. "Computer aided recognition of pathological voice." *2014 31st National Radio Science Conference (NRSC). IEEE*, 2014.
- [37] Abdulmajeed, Nuha Qais, Belal Al-Khateeb, and Mazin Abed Mohammed. "A review on voice pathology: Taxonomy, diagnosis, medical procedures and detection techniques, open challenges, limitations, and recommendations for future directions." *Journal of Intelligent Systems* 31.1 (2022)