Machine Learning Approaches to Laryngeal Pathologies Detection and Classification: A Comprehensive Literature Review

Hassan Ezzahori*, **Abdelkrim Hammimou, Abdelghani Boudaoud, Mounaim Aqil**

EAPT, Superior School of Technology
Sultane Moulay Slimane University
Beni Mellal, 23000, Morocco
E-mails: hasan.ez-zahori@usms.ma,
abdelkrim.hamimmou@usms.ma,
abdelghani.boudaoud@usms.ma,
mounaim.aqil@usms.ma

**Corresponding author*

Received: May 29, 2025 **Accepted: October 06, 2025**

Published: December 31, 2025

Abstract: *Voice alterations are the most frequent early sign of laryngeal pathologies. Implementing voice screening protocols could help to detect early laryngeal diseases, such as cancer. This field is increasingly turning to machine learning (ML) approaches for diagnostic purposes, using analysis of vocal patterns and speech characteristics to identify these disorders. This review aims to synthesize and evaluate the literature on laryngeal disease detection and classification using ML. A comprehensive search of five major multidisciplinary and specialized databases was conducted to identify articles published between 2015-2024, yielding 102 relevant studies. Data extraction and analysis were conducted using the "preferred reporting items for systematic reviews and meta-analyses" system. The included studies utilize deep learning or ML algorithms to analyze the speech signal. The review reveals that the Saarbrücken voice database remains the most coveted by the researchers, as it was used in 53% of the studies. It shows that mel-frequency cepstral coefficients are the most commonly used features, appearing in 54% of included studies, alongside the support vector machine algorithm, which is the most commonly used classifier (50% of the studies). The review demonstrates that traditional ML techniques are constantly being overtaken by deep learning ones. This review serves as a roadmap for future research, guiding the development of ML and deep learning-based algorithms for laryngeal disease detection.*

Keywords: *Voice disorders, Neural network, Machine learning, Deep learning, Laryngeal disease.*

Abbreviations

Amplitude perturbation quotients	APQ
Arabic voice pathology database	AVPD
Artificial intelligence	AI
Artificial neural networks	ANN
Convolutional neural network	CNN
Decision tree	DT
Deep neural network	DNN
Dual-tree complex wavelet transform	DTCWT
Extreme gradient boosting	XGBoost
Extreme learning machine	ELM

Gammatone cepstral coefficients	GTCC
Gammatone frequency cepstral coefficients	GFCC
Gammatone spectral latitude	GTSL
Gaussian mixture models	GMM
Gradient boosting machine	GBM
Harmonics to noise ratio	HNR
Histogram-kullback-leibler divergence	H-KLD
Hospital Universitario Principe de Asturias database	HUPA
Interlaced derivative pattern	IDP
K-nearest neighbors	KNN
Light gradient-boosting machine	LightGBM
Linear discriminant analysis	LDA
Linear predictive cepstral coefficients	LPCC
Linear predictive coefficients	LPC
Logistic regression	LR
Long short time memory	LSTM
Machine learning	ML
Massachusetts eye and ear infirmary database	MEEI
Mel-frequency cepstral coefficients	MFCC
Multilayer perceptron	MLP
Naive bayes	NB
Neural network	NN
Online sequential extreme learning machine	OSELM
Pitch perturbation quotients	PPQ
Preferred reporting items for systematic reviews and meta-analysis	PRISMA
Random forest	RF
Root mean squared energy	RMSE
Saarbrücken voice database	SVD
Short-time Fourier transform	STFT
Stationary wavelet transform	SWT
Stochastic gradient descent	SGD
Support vector machine	SVM
Voice Icar Federico II	VOICED
Zero-crossing rate	ZCR

Introduction

Laryngeal disorders can affect speech production ability, including vocal cord paralysis, laryngitis, and spasms. There are multiple causes of vocal disorders, including physiological, neurological, and psychological factors. Studies have shown that voice problems may affect 6-15% of the general population, increasing to 20-50% and may reach as high as 80% for teachers [5]. Prevalence of vocal complaints reported by 59% of call center operators [81]. For this reason, great importance is given to the diagnosis of vocal pathologies. The detection of voice pathologies using AI and ML has continued to grow over the last two decades, thanks to the evolution of the computing power of machines [27].

During the last decade, ML algorithms have achieved very good results [48, 114]. However, neither the biomedical industry nor the clinical community has exploited research results. This is essentially due to the great disparity of the results obtained. The absence of standards and standardization of terminology also contributes to this situation.

Other important issues are related to the possibility of putting it into practice and verifying the results for possible industrialization. This can be ensured by providing detailed descriptions in the publications of the AI algorithms, platforms, software, and libraries. Few authors share the code of their programs and set up an executable version [46, 97]. All of these factors hinder the development and collaboration of the research community. ML algorithms utilized for classification use different databases such as MEEI [75], SVD [93], or HUPA [42]. Since datasets are different, comparing the performance of vocal pathology prediction systems is far from being objective. In this perspective, we propose to review recent works that use ML to classify or detect laryngeal diseases.

Several reviews have been conducted in recent years [14, 54, 100, 110], but none have covered as many articles. Additionally, this review is unique as it covers an entire decade and provides insight into the evolution of research in this field. In 2019, the authors of [43] established a review of automatic voice condition analysis systems. They presented concepts, categorization of different aspects of voice pathology, and the methodologies and methods most commonly used in these systems. The authors described preprocessing methods, including accentuation, filtering, framing, and windowing techniques. They examined different voice features such as temporal, spectral, and cepstral characteristics. They analyzed and criticized the most frequently used classification and decision-making algorithms. In 2020, the authors of [110] conducted a meta-analysis review using 45 studies published between 2002-2020 in reputable journals and conference proceedings. The authors of [110] aimed to compare and reveal the weaknesses and strengths of the SVD, MEEI, and AVPD [77]. The authors established that, until then, the MEEI was the most frequently used database, appearing in 54% of articles. The study discussed classification techniques and concluded that SVM was the most commonly used classification algorithm during this period.

Another study categorize and compares ML techniques in voice disease diagnosis [100]. In [100] authors analyzed 48 articles published between 2002-2022. The review classified voice disorders into 3 categories: mental illness (30 articles), laryngeal diseases (6 articles), and neurological diseases (12 articles). The authors provided targeted diseases, classification algorithms, features, and metrics. However, with only 6 articles addressing laryngeal diseases, the sample size was too small to reach meaningful conclusions about databases or algorithms usage frequency and their effectiveness.

Earlier comprehensive review of 98 articles published between 1996-2017 covering various categories of voice disorders: structural lesions, inflammatory conditions, trauma or injury-based disorders, systemic conditions, psychiatric and psychological disorders, neurological conditions, etc. [54]. In this review, the authors presented all stages of classification, discussed, and compared the effectiveness of multiple features and classifiers. Common ML algorithms have been described and grouped into 8 categories: ANN, DTs, hidden Markov models, GMMs, k-means clustering, SVM, linear classifiers, and combined classifiers.

Despite the substantial research generated in this area, there remains a critical need for systematic synthesis and integration of the accumulated knowledge. Publications are scattered across various disciplines (otorhinolaryngology, phoniatrics, signal processing, and ML), making it challenging to gain an overview of the progress made. This systematic review aims to fill this gap by providing a comprehensive overview of works published over the past 10 years. Its main objectives are to provide a comprehensive analysis of the techniques, databases, and voice features used, to consolidate the progress made, and to identify unexplored fields. This question can be answered through three objectives:

- To identify and characterize voice databases related to laryngeal pathologies, based on their size, composition, and available clinical information, as reported in scientific literature.
- To conduct a comprehensive analysis of the techniques and methodologies employed in the study of voice disorders associated with laryngeal pathologies, including a review of speech signal processing methods and analytical models.
- To compile an inventory of the most frequently extracted voice features relevant for detecting and classifying laryngeal diseases to consolidate advances in acoustic biomarkers.

Materials and methodology

To achieve the previously mentioned objectives, a search for quality scientific articles is necessary. PRISMA is selected as the screening method [72].

Databases concerned

The literature search was conducted across five major databases, including three specialized repositories: PubMed for biomedical and life sciences, IEEE for computer science and electrical engineering, and SpringerLink for science, technology, and medicine. Additionally, two comprehensive multidisciplinary databases were used, Web of Science and Scopus which encompass relevant peer-reviewed publications from diverse scientific disciplines. Therefore, since voice alteration can be due to several physiological and psychological disorders, keywords were chosen to focus the search on vocal disorders related to laryngeal diseases. Therefore, the keywords “larynx” and “laryngeal” were added to items “voice disorder”, “machine learning”, and “neural network”. The same keywords were utilized in all databases. Logic operators “and” and “or” have been used with parentheses and quotation marks for maximum relevance. The search fields were limited to the paper’s title, abstract, and keywords.

Eligibility criteria

To ensure relevance and quality, we restricted our document selection to articles published from 2015-2024 and indexed in widely recognized databases. The exclusion criteria taken into consideration were chosen to avoid biasing the study. Therefore, the excluded articles are reviews, non-biomedical articles, and articles written in a language other than English. After searching by keywords and filtering records using advanced search tools, the title and the abstract are screened to eliminate duplicate articles, those that do not use AI for classification, and studies targeting special populations such as children, singers, etc. Other publications were excluded because they did not use voice signals but rather other signals such as neck vibrations measured by accelerometers, electroglottographic signals captured by electrodes, and images extracted from endoscopic video signals.

Data extraction

A systematic data extraction procedure was implemented by two authors working independently on the 102 selected studies. This process comprised two sequential stages: initial abstract screening followed by a comprehensive full-text review, which included a detailed examination of figures and tables. Standardized data extraction captured information on the targeted diseases, database sources, and datasets used. It also captured the classification methodology framework, specifically distinguishing between binary and multiclass systems, as well as the machine and deep learning classification algorithms employed, and the accuracy outcomes for each study.

Results

The primary objective of this review was to identify articles published 2015-2024 that utilize ML for laryngeal disorder diagnosis through voice analysis. A comprehensive search was conducted across specified databases. From the 300 identified records, 141 were retained after removing duplicates. A total of 39 articles were then excluded based on the exclusion criteria in our two-stage screening, leaving 102 articles for data extraction (Fig. 1).

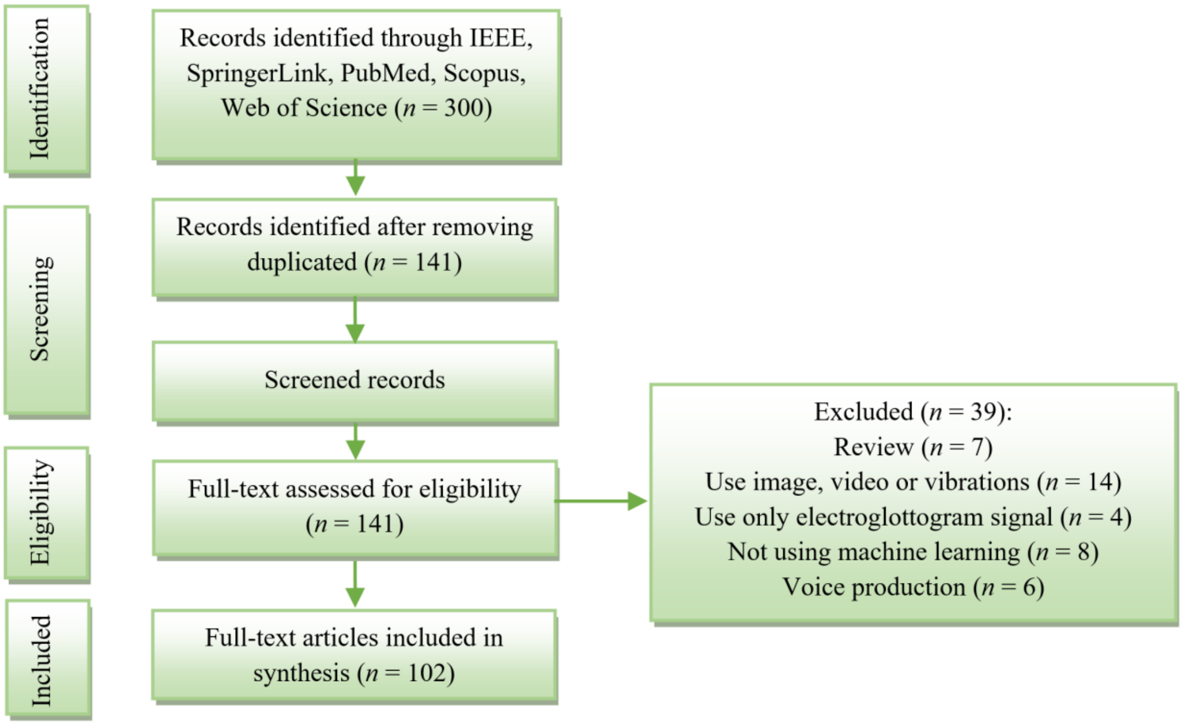


Fig. 1 Publications filtering flowchart

The review presents the key data from each study, including the targeted disease, the type of classification (detection or classification), the database used, the classification algorithm, the features, and the classification performance (Table 1). Subsequently, statistics regarding the databases, classifiers, and voice features are provided.

Table 1. List of included studies ordered by the year of publication and the best accuracy per year: class – binary (normal pathological) or multi (multiclass pathological), best accuracy is the one obtained in the current study

Ref.	Year	Diseases targeted	Database	Class	Classifiers	Features	Best accuracy
[107]	2024	laryngeal cancer, paralysis, mucosal diseases	Local (Korea)	Multi	CNN (ResNet)	MFCC, octave frequency, spectrum energy	100%
[94]	2024	voice pathologies (unspecified)	SVD, MEEI, Private	Binary	SVM, KNN, NB, DT	APQ, frequency, jitter, period, shimmer	99.97%

[87]	2024	cordectomy, dysphonia, partial resection, contact pachyderma, laryngitis, vox senilis, polyps, leukoplakia, Reinke's edema, paralysis	SVD	Multi	SVM	MFCC, RMSE, ZCR	99.50%
[31]	2024	dysphonia, laryngitis, Reinke's edema, vox senilis	SVD	Multi	CNN	waveforms (raw)	99.40%
[130]	2024	voice pathologies (unspecified)	MEEI, SVD, HUPA	Binary	CNN, MLP, LSTM	MFCC, GTCC, waveforms, Glottal flow features	98.22%
[17]	2024	voice pathologies (unspecified)	VOICED	Binary	SVM, KNN, DT, RF	Wav2vec features	98.00%
[66]	2024	laryngeal cancer, vocal cord paralysis, mucosal disease	Local (Korea)	Multi	SVM, ANN, GBM, CNN	MFCC	97.00%
[73]	2024	dysphonia	Local (Brazil)	Multi	XGBoost, KNN, DT, LR, MLP, RF, SVM	47 acoustic features (time, spectral, cepstral features, etc.)	96.20%
[29]	2024	Parkinson's disease	Local (UK)	Binary	KNN, SVM, wide NN	MFCC, GTCC	92.30%
[99]	2024	voice pathologies (unspecified)	Local (Iran)	Binary	DNN	chroma, jitter, MFCC, pitch, formants	91.00%
[57]	2024	dysphagia	Local	Binary	CNN	mel-spectrum	90.00%
[67]	2024	voice pathologies (unspecified)	HUPA	Binary	SVM, RF, XGBoost	fundamental frequency, MFCC	88.79%
[128]	2024	dysphonia, recurrent laryngeal nerve palsy	SVD	Binary	SVM, ANN	OpenSMILE features, MFCC	81.00%
[69]	2024	voice pathologies (unspecified)	SVD	Binary	CNN (ResNet18), SVM	handcrafted	80.90%

[60]	2024	voice pathologies (unspecified)	SVD, HUPA	Binary	SVM, RF, LSTM, CNN	MFCC, spectrogram, mel-spectrogram	76.00%
[89]	2023	dysphonia, laryngitis	VOICED	Multi	CNN	mel-spectrogram	99.50%
[124]	2023	Parkinson's disease	PC-GITA (Spain)	Binary	DT, KNN, NB, SVM	MFCC	99.00%
[103]	2023	reflux laryngitis, hyperkinetic/ hypokinetic dysphonia	VOICED	Multi	MLP, 1D-CNN	STFT spectrogram, MFCC	97.10%
[13]	2023	dysphonia, laryngeal motion disorder, laryngitis, Reinke's edema, vox senilis	SVD	Multi	RF, SVM	energy, entropy, ZCR	96.88%
[79]	2023	voice pathologies (unspecified)	SVD	Binary	CNN, LSTM	multi-layer hybrid network features	96.05%
[65]	2023	dysphagia	Local	Binary	CNN	STFT spectrogram, MFCC	95.00%
[120]	2023	voice pathologies (unspecified)	MEEI	Binary	LDA	GFCC, LPC, MFCC	92.73%
[126]	2023	dysphonia, paralysis, nodules, polyps	Local (China)	Multi	CNN	MFCC	92.00%
[25]	2023	voice pathologies (unspecified)	SVD	Binary	RF, XGBoost	spectral features	90.60%
[129]	2023	dysphonia	MVPD (Malaysia)	Binary	OSELM	MFCC	90.00%
[18]	2023	spasmodic dysphonia	Local (Italy)	Multi	KNN	48 acoustical features	89.00%
[9]	2023	vocal nodule, Reinke's edema, neurological dysphonia	local (Brazil)	Binary	SVM, KNN	MFCC, cepstral distances and peaks	82.00%
[115]	2023	leukoplakia	SVD	Binary	DNN	STFT spectrogram	81.25%
[59]	2023	spasmodic dysphonia, vocal fold paralysis	SVD	Binary	SVM	OpenSMILE features	72.10%

[78]	2023	hyperkinetic dysphonia	VOICED	Multi	CNN	MFCC, Q-factor wavelets transform	67.91%
[117]	2023	hyperfunctional dysphonia vocal fold paresis	SVD	Multi	SVM	MFCC	62.77%
[104]	2022	voice pathologies (unspecified)	SVD, AVPD, MEEI HUPA,	Binary	SVM, SGD	SWT-based framework	99.97%
[82]	2022	mild degree of vocal deviation	Local (Brazil)	Binary	ANN	wavelet transform coefficients	99.75%
[131]	2022	neuromuscular pathologies, structural pathologies	MEEI, SVD, HUPA	Multi	SVM, MLP, RF	GTSL	99.60%
[39]	2022	laryngitis, leukoplakia, carcinoma, cysts, dysphonia, nodules, paralysis, polyps, Reinke's edema	SVD	Multi	LSTM	MFCC	98.02%
[95]	2022	voice pathologies (unspecified)	SVD	Binary	SVM, CNN	MFCC	97.80%
[22]	2022	voice pathologies (unspecified)	VOICED	Binary	DNN	MFCC	97.80%
[71]	2022	dysphonia	Paraiba (Brazil)	Binary	RF, NB, SVM, RL, KNN	acoustical (34 features)	91.00%
[38]	2022	voice pathology severity (unspecified)	Local	Multi	1D-CNN	waveforms (raw)	88.30%
[4]	2022	dysphonia	SVD	Binary	NB	MFCC	81.48%
[61]	2022	voice pathologies (unspecified)	HUPA	Binary	SVM, RF, DNN, LSTM, CNN, Adaboost	MFCC, spectrogram, mel-spectrogram	81.00%
[127]	2022	recurrent laryngeal nerve palsy, dysphonia	SVD	Binary	SVM	MFCC, perceptual LPCC	76.25%
[116]	2022	hyperkinetic/ hypokinetic dysphonia, laryngitis	SVD	Binary	SVM	MFCC	75.00%

[20]	2022	cordectomy, dysphonia, frontolateral laryngectomy, pachydermia larynges	SVD	Multi	CNN, LSTM	spectrogram	71.15%
[111]	2021	cysts, dysphonia, laryngitis, nonfluency syndrome	SVD	Binary	SVM, NB, DT	MFCC, pitch, ZCR, spectral entropy, energy, roll-off	100%
[12]	2021	dysphonia, cysts, vocal fold polyps, sulcus vocalis	SVD, local	Binary	CNN, LSTM	MFCC, ZCR, energy entropy, shimmer	99.58%
[74]	2021	aphasia	Local (China)	Binary	CNN, KNN, LDA, RF, SVM	MFCC, formant frequency, energy	99.23%
[23]	2021	voice pathologies (unspecified)	VOICED	Binary	KNN	LPCC	98.23%
[122]	2021	dysphonia, reflux laryngitis	VOICED	Multi	3D CNN	Bump/Morlet/ Morse wavelets	97.70%
[83]	2021	voice pathologies (unspecified)	SVD	Binary	CNN, LSTM	mel-spectrogram	95.60%
[3]	2021	cysts, polyps, paralysis	SVD	Multi	OSELM (ELM)	MFCC	91.17%
[112]	2021	dysphonia, polyps, partial resection, vox senilis, laryngeal motion disorder, Reinke's edema, carcinoma, leukoplakia	SVD	Multi	CNN, LSTM	MFCC, pitch, ZCR, spectral entropy, energy, roll-off	86.87%
[70]	2021	dysphonia, laryngitis, vocal fold polyps, leukoplakia, vocal fold cancer, nodules, Reinke's edema, granuloma, contact ulcers	SVD	Binary	CNN, FNN	LPCC, MFCC, normalized Skewness and Kurtosis	82.69%
[30]	2021	dysphonia, laryngitis, recursion parsing	SVD, private	Multi	CNN (ResNet)	MFCC	82.20%

[58]	2021	vocal atrophy, unilateral vocal paralysis, organic vocal fold lesions, dysphonia	Local (Taiwan)	Multi	CNN	MFCC	67.00%
[41]	2020	cysts, paralysis, polyps	SVD, HUPA, MEEI, AVPD	Binary	SVM, ANN, SGD	glottal volume velocity waveform	99.98%
[15]	2020	voice pathologies (unspecified)	MEEI	Binary	GMM, NB	H-KLD, MFCC	99.55%
[51]	2020	Alzheimer's disease, dysphonia, laryngitis, paralysis, Parkinson's disease, recursion parsing	SVD, local (Tunisia)	Binary	SVM	discrete wavelet transform features	99.26%
[101]	2020	paralysis, polyps	Local	binary	RF, XGBoost, NB, SVM, KNN.	MFCC, octave-based spectral contrast,	95.65%
[80]	2020	voice pathologies (unspecified)	SVD	Binary	CNN (ResNet34)	spectrogram	95.41%
[1]	2020	erythroplakia, keratosis, leukoplakia	MEEI, SVD	Binary	SVM, ELM, XGBoost	MFCC, LPCC, log area ratios, reflection coefficient	86.11%
[64]	2020	laryngeal cancer	Local	Binary	SVM, XGBoost, LightGBM, ANN, CNN	MFCC	85.00%
[44]	2020	laryngeal cancer	local (India)	Binary	SVM, RF	MFCC	80.00%
[62]	2020	voice pathologies (unspecified)	HUPA, SVD	Binary	SVM	MFCC, spectrogram	76.19%
[36]	2019	voice pathologies (unspecified)	MEEI	Binary	DNN, GMM, SVM	MFCC	99.32%
[46]	2019	voice pathologies (unspecified)	MEEI	Binary	SVM, CNN	MFCC	95.90%
[47]	2019	voice pathologies (unspecified)	MEEI	Binary	SVM, CNN,	MFCC, mel-spectrogram	93.50%,
[63]	2019	vocal nodules, cysts, paralysis, polyps	MEEI, SVD	Multi	KNN, SVM	DTCWT	93.32%

[91]	2019	dysphonia, paralysis, polyps, polypoid chorditis, papillomatosis	Local (USA)	Binary	CNN	spectrogram	90.00%
[118]	2019	voice pathologies (unspecified)	MEEI SVD VOICED	Binary	Boosted Trees	jitter, fundamental frequency, shimmer, HNR	84.50%
[49]	2019	dysphonia, laryngitis, paralysis	SVD	Multi	CNN, LSTM	spectrogram	80.00%
[48]	2018	laryngitis	SVD	Binary	LSTM, ANN	jitter, shimmer, autocorrelation	100%
[7]	2018	cysts, paralysis, polyps	MEEI, SVD, AVPD	Multi	SVM	peaks, lags, entropy	99.79%
[121]	2018	dysphonia	SVD	Binary	ANN	MDVP Features	93.33%
[109]	2018	voice pathologies (unspecified)	MEEI	Binary	MLPNN	Haralick texture features	89.71%
[119]	2018	voice pathologies (unspecified)	SVD	Binary	SVM, DT, KNN, NB, Logistic Tree	jitter, HNR, formant, MFCC, fundamental frequency, shimmer	85.77%
[53]	2018	voice pathologies (unspecified)	AVPD, MEEI, SVD, HUPA	Binary	XGBoost, isolation forest, CNN	energy, jitter, HNR, MFCC, pitch, shimmer	73.30%
[45]	2018	cysts, polyps, neoplasms, paralysis	Local	Binary	CNN	MFCC, jitter	71.20%
[113]	2018	dysphonia, laryngitis, paralysis	SVD	Binary	SVM	jitter, shimmer, HNR, MFCC	71.00%
[90]	2018	neoplasm, phonotrauma, Vocal palsy	local (Taiwan)	Multi	SVM, RF, KNN, GBoost	MFCC	68.48%
[114]	2017	dysphonia	SVD	Binary	ANN	jitter, shimmer, HNR	100%
[8]	2017	cysts, paralysis, polyps	AVPD, MEEI, SVD	Binary	SVM	multi-dimensional voice program features	99.68%

[106]	2017	voice pathologies (unspecified)	MEEI	Binary	ANN	19 acoustical features (jitter, shimmer, HNR, etc.)	99.40%
[84]	2017	cysts, paralysis, polyps	MEEI, SVD, AVPD	Multi	SVM	IDP	99.40%
[10]	2017	voice pathologies (unspecified)	MEEI	Binary	SVM	MFCC	98.00%
[28]	2017	dysphonia, paralysis	SVD	Binary	NB	MFCC	90.00%
[55]	2017	dysphonia, laryngitis, paralysis	SVD	Multi	ANN	entropy, jitter, MFCC, Kurtosis, shimmer	87.50%
[52]	2017	voice pathologies (unspecified)	SVD	Binary	DNN, LSTM	waveform (raw)	71.36%
[56]	2016	voice pathologies (unspecified)	SVD	Binary	k-means, RF	jitter, shimmer, energy, MFCC	100%
[6]	2016	voice pathologies (unspecified)	AVPD	Binary	SVM	voice contour, intensity	100%
[85]	2016	voice pathologies (unspecified)	MEEI	Binary	SVM	vocal tract area variance, average	99.02%
[26]	2016	physiological pathologies, neuromuscular pathologies	MEEI	Binary	GMM, SVM, LDA	MFCC, spectral	98.70%
[34]	2016	laryngeal cancer	SVD	Binary	ANN	temporal/ frequency descriptors	98.00%
[16]	2016	laryngeal cancer	SVD	Binary	ANN	temporal/ frequency descriptors	96.90%
[11]	2016	dysphonia	SVD	Binary	SVM, GMM	MFCC	96.50%
[37]	2016	voice pathologies (unspecified)	MEEI, local	Binary	ANN	modified Mellin transform features	96.48%
[50]	2016	laryngitis, paralysis, dysphonia, voice tremor, Reinke's edema	SVD	Binary	SVM	pitch, formants	86.00%

[76]	2015	voice pathologies (unspecified)	MEEI, HUPA, PARCZ	Binary	SVM, RF	local and high-level feature combinations	99.90%
[40]	2015	voice pathologies (unspecified)	MEEI	Binary	Genetic algorithm, LDA, SVM	phase space parameters, Lyapunov spectrum	98.40%
[2]	2015	voice pathologies (unspecified)	MEEI	Binary	SVM	discrete wavelet coefficients	97.01%
[33]	2015	voice pathologies (unspecified)	SVD	Binary	GMM	MFCC	95.00%
[108]	2015	cysts, dysphonia, laryngitis Reinke's edema	SVD	Binary	SVM	MFCC	86.44%

Number of publications

To gauge the interest in a research topic, it is sufficient to look at the number of publications dedicated to it. When a subject is thoroughly addressed, the articles that deal with it decrease. However, the use of ML in the detection of voice disorders continues to attract the attention of researchers, and Fig. 2 shows sustained growth since 2015. The number of publications per year has sustained growth. The increasing nature of the curve indicates that researchers are increasingly interested in detecting and classifying vocal disorders using ML and deep learning.

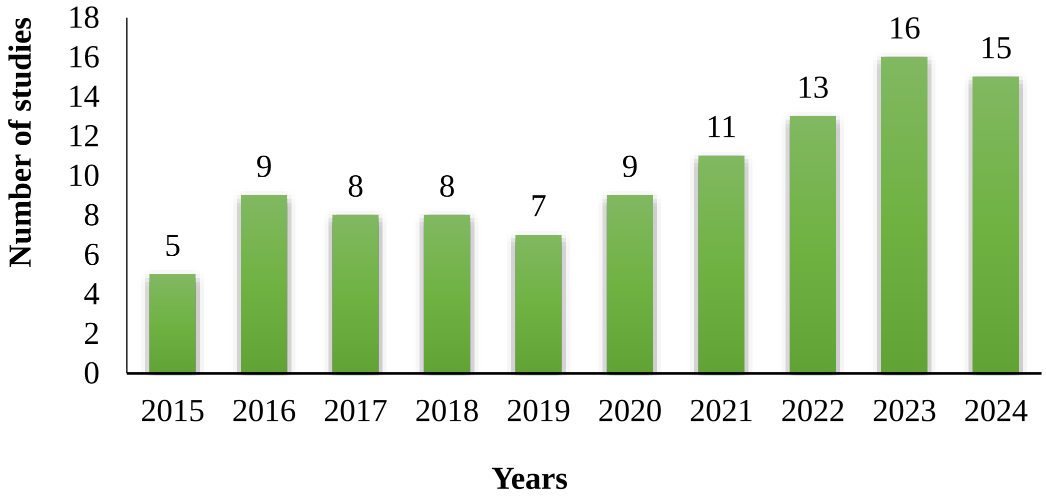


Fig. 2 Number of included publications from 2015 to 2024

Databases

The researchers used approximately 35 databases in their studies. As shown in Table 1, most of them used a single database, while several studies employed two, three [8, 76, 84, 131], or even four databases simultaneously [41, 104]. Fig. 3 illustrates how frequently each database occurs.

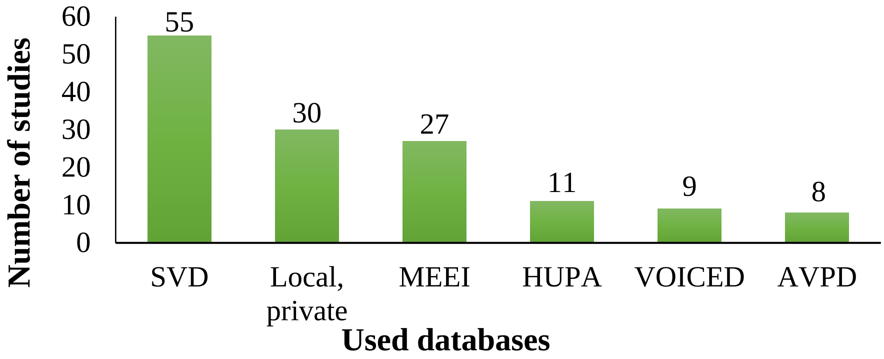


Fig. 3 Number of studies for various databases

SVD [125] is the most commonly used database, employed 55 times (53.90% of studies). Its best score was achieved by RF, k-means [56], ANN [114], and LSTM [48], displaying 100% accuracy. SVD is a free-access data collection that contains more than 2 000 German-speaking individuals, including 1 290 patients and 687 healthy persons. Each record in the database contains three vowels (|a|, |i|, and |u|) in high, neutral, and low tones in addition to the recording of a German passage “Guten morgen, wie geht es ihnen?” (in English means “Good morning, how are you?”). For each recording, the speech signal and the electroglottogram signal are saved in separate files; a text file contains the necessary comments. The signal is 16 bit and sampled at 50 kHz.

The MEEI [75] is the second most frequently used database. It was utilized 27 times and permitted to reach 99.96% accuracy [76]. The MEEI is one of the most popular databases in voice analysis. It contains more than 53 normal and 657 pathological voice samples and a wide variety of organic, neurological, psychogenic, and traumatic voice disorders (e.g., adductor spasmodic dysphonia, conversion dysphonia, erythema, hyperfunction, etc.) [76]. Vocal recordings of the sustained vowel |a| are 1-3 seconds in length, and readings of the first sentence of the Rainbow passage are 12 seconds length. The database was created by the MEEI voice and speech lab and commercialized by Kay Elemetrics [75]. The sampling frequency was 10 kHz, 25 kHz, and 50 kHz, while the signal resolution was 16 bits. In this collection, clinical information, age, and gender were included in all recordings.

The HUPA [42], developed in Alcalá de Henares, Madrid, Spain, is the third most frequently used database. This freely accessible database comprises 440 recordings of sustained vowel |a| phonations from 366 adult Spanish speakers. The speech recordings include 239 samples from 197 healthy subjects and 201 samples from 169 patients with pathological conditions. All speech files were digitized with 16 bit resolution and sampled at 25 kHz [130]. The pathological samples represent a range of organic pathologies such as nodules, polyps, and Reinke’s edema. It was employed in 11 studies, and the highest accuracy reached was 99.98% [41].

The VOICED PhysioNet database has been publicly available since 2018 and contains a collection of 208 voice recordings [19]. This collection includes 150 recordings from individuals diagnosed with various voice disorders and 58 recordings from healthy individuals. Volunteers aged 18-70 contributed to this voice recording dataset. All recordings were acquired in a professionally controlled environment. Records have 32 bit resolution sampled at 8 kHz. The Vox4Health management system was used to capture the recordings on Samsung Galaxy S4 smartphones. The voice recordings from patients with voice disorders are grouped into three distinct categories: reflux laryngitis (38 recordings), hypokinetic dysphonia (40 recordings), and hyperkinetic dysphonia (72 recordings). The hyperkinetic dysphonia group encompasses several specific vocal fold conditions: vocal fold nodules, Reinke's edema, chorditis, rigid vocal fold, polyps, and prolapse. The hypokinetic dysphonia category covers other conditions such as conversion dysphonia, extraglottic air leak, dysphonia of the chordal groove, presbyphonia, vocal fold paralysis, glottic insufficiency, laryngitis, and adduction deficit.

The AVPD [75] is also used in 8 articles. In this database, samples of words and voices were recorded at various sessions at King Abdul Aziz University Hospital in Riyadh, Saudi Arabia. The recordings were made using KayPENTAX professional equipment. The database protocol has been developed to overcome specific MEEI deficiencies [62]. It contains 366 samples (175 healthy and 187 pathological). Normal and pathological vocal folds were determined after clinical laryngeal stroboscopic examination. For pathological cases, the perceptual severity of voice disorders was rated on a 3-point scale, where 1 represents mild, 2 represents moderate, and three represents the most severe. A consensus among three expert medical doctors established the severity rating for each sample. The recording has different types of texts:

- three sustained vowels with the onset and offset information;
- isolated words, including Arabic digits and other common words;
- continuous, simple, and short speech.

Most speakers provided three repetitions of each vowel. Single-word and repetitive speech recordings were limited to discourage patient fatigue. Both normal and disease voice signals in AVPD are sampled at 50 kHz using 16 bit resolution.

In addition to these five databases, other databases are used by researchers, and are either private or of limited scope (used only once). These databases are grouped into a single class "Private/local". Using data collected in Table 1, the distribution of targeted impairments is given in Fig. 4, where only addressed diseases are represented. In terms of pathology coverage, studies conducted on the SVD database are addressing 12 pathologies (cancer, contact pachyderma, cordectomy, cysts, dysphonia, vocal folds partial resection, laryngitis, leukoplakia, paralysis, Reinke's edema, and vox senilis) in addition to studies that address unspecified pathologies using binary classification (pathological/healthy). Those conducted on VOICED data address only two specific pathologies (four studies for dysphonia and three studies for laryngitis) [48, 79, 90, 103, 122]. Thus, it is the least frequently used database among the five major ones, in terms of disease variety.

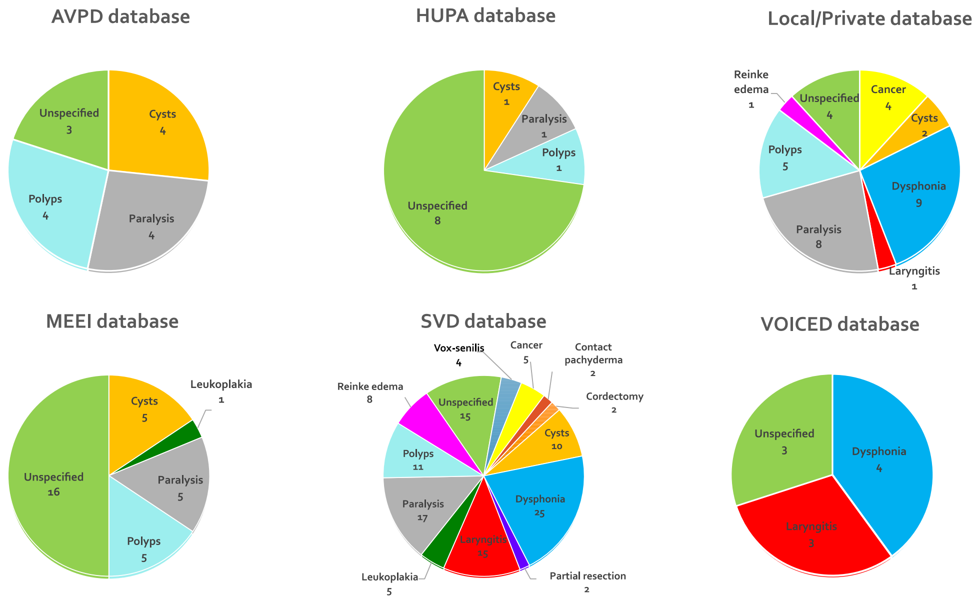


Fig. 4 Distribution of addressed pathologies in various databases: numbers indicate the number of studies addressing the pathology

Classifiers

The papers included in this review employed a variety of algorithms for classification as indicated in Fig. 5. While the primary objective was typically the binary detection of voice disorders (i.e., classifying a voice as either normal or pathological), several studies extended this to multiclass classification Table 1. These more granular approaches aimed to not only identify the presence of a voice disorder but also to determine the specific type of disease and, in some cases, its severity [18, 20, 96, 98]. In their work, [38] created an original dataset and applied a CNN architecture for pathological voice severity classification, achieving accuracy of 88.30%. The approach processed raw waveform data from sustained voice recordings sampled at 48 kHz with 16 bit resolution, with mono signals partitioned into analysis of 200 ms frames.

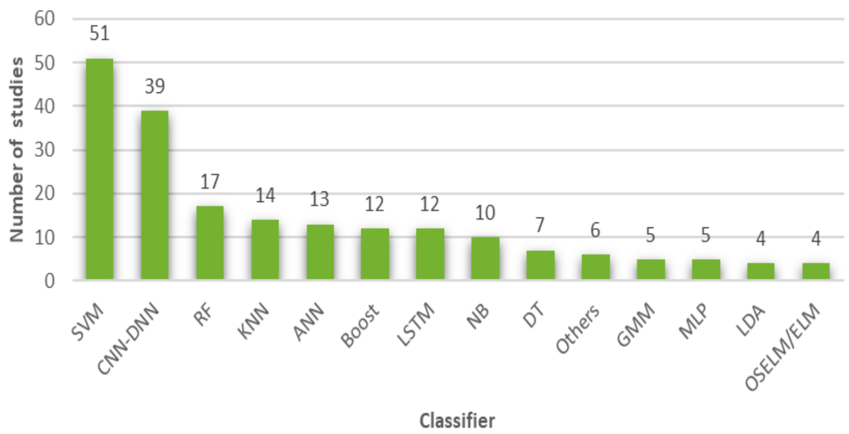
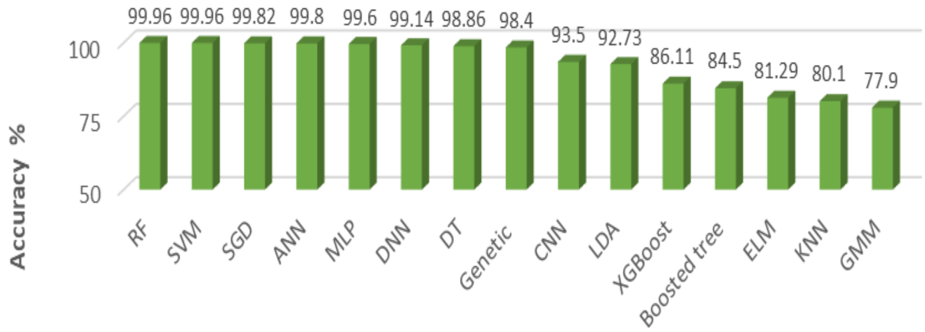


Fig. 5 Number of studies for the most common classifiers

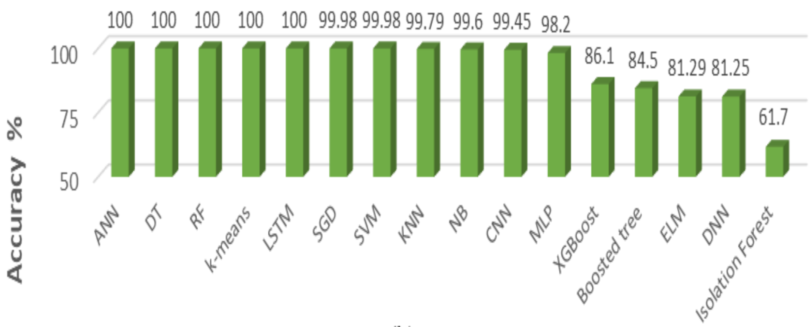
In Fig. 5, algorithms used only once, such as adaptive boosting [61], isolation forest [53], k-means [56], and genetic algorithm [40], are grouped in class “others”. It should be noted that many studies have used more than one algorithm [73]. The SVM is the most employed algorithm with 51 occurrences, followed by CNN-DNN and its variants with 39 occurrences.

However, if we consider all NN structures, including CNN, DNN, MLP, and LSTM, as a single group, we can see that this is the most commonly used category. In the same context, the boosting technique was used 12 times but with different algorithms such as XGBoost [25, 67, 73], adaptive boosting [61] and LightGBM [64]. Another result concerns the classifiers of 102 studies included in this review, 27 of them adopt a multiclass classification and 75 use a binary classification.

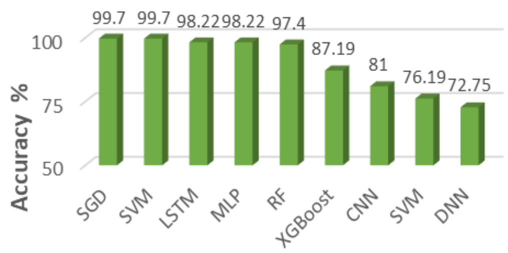
A quantitative analysis of classifiers' performance across different databases is established as illustrated in Fig. 6, where several classifiers achieve accuracy of 100%, such as k-means, RF [56], ANN [114], LSTM [48], and DT [111]. All these studies are conducted in SVD using binary classification and various features (MFCC, jitter, shimmer, pitch, spectral roll-off, etc.).



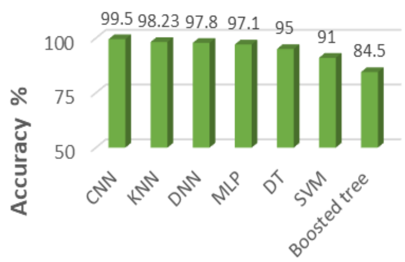
a) MEEI



b) SVD



c) HUPA



d) VOICED

Fig. 6 Classifier performances through different databases

In MEEI database, the best accuracy of 99.96% is achieved by RF and SVM [76], and the study employed time-domain features and binary classification to detect voice disorders. Within the same context, the CNN algorithm demonstrated its optimal performance using the VOICED database, achieving an accuracy of 99.50%. This finding originated from a study focused on the detection and classification (multiclassification) of dysphonia and laryngitis [90].

An examination of the trends for "machine learning algorithms" and "deep learning algorithms" over the last decade reveals a substantial increase in the adoption of deep learning relative to ML (Fig. 7). The "deep learning algorithms" category encompasses CNN, DNN, and LSTM algorithms, whereas traditional ML methods are classified under the "machine learning algorithms" category.

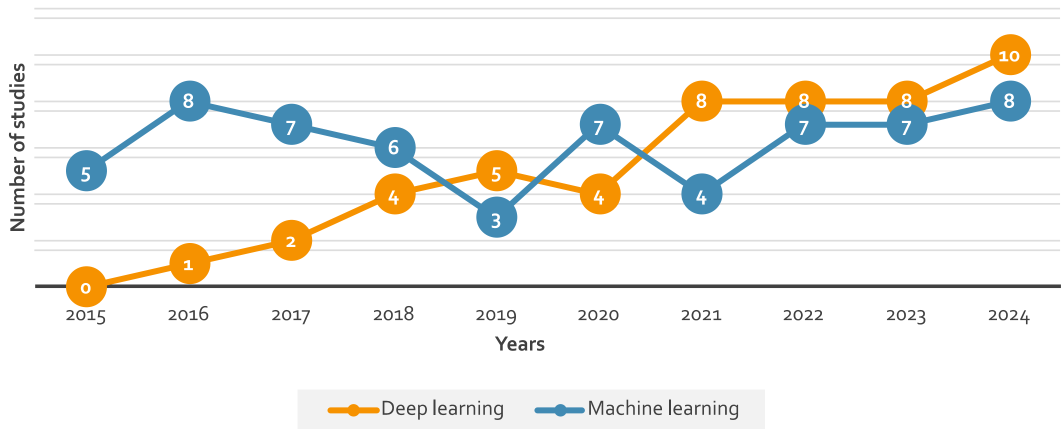


Fig. 7 Evolution of deep learning usage vs ML between 2015-2024

Voice features

A wide range of voice signal features is utilized in machine/deep learning laryngeal disease detection and classification. Several classification methods exist in the literature [32, 102]. But features can be classified by the domain from which they are extracted into 3 main categories:

- Time-domain features: characteristics of a speech signal that are directly extracted from its raw waveform, such as amplitude envelope, RMSE, ZCR, short-time energy, and pitch period [68]. The main advantages are that it requires less computation time and does not involve complex transformations [88]. This category is more and more used with deep learning classifiers [31, 38, 52].
- Frequency-domain features, also called physical features [105], such as octave frequency spectrum energy [107], formants [50, 119], and spectral roll-off [111, 112]. These kinds of features are calculated by converting a one-dimensional time-domain speech signal into a two-dimensional representation using the STFT [123].
- Cepstral features (quefrency-domain or perceptual features), such as MFCC, perceptual LPC, LPCC [21, 92, 105]. These features are computed using the inverse of the Fourier transform of the speech logarithmic spectrum.

In speech analysis and laryngeal disease detection, several researchers utilize a combination of the three categories to increase performance. For example, in [87], the authors used MFCC, RMSE, and ZCR, while entropy, jitter, Kurtosis, MFCC, and shimmer are employed in [56]. Fig. 8 displays the number of studies employing each category. The most frequently used class is "cepstral features" with 63 occurrences, 35 of these were used independently (e.g. [3, 4, 39, 66, 116, 117, 124, 126, 129]), and 28 were used in combination with the other two classes (e.g. [12, 61, 74, 111, 112, 128]). Frequency-domain features were used 47 times, making them the second most common, with time-domain features close behind at 36 uses.

Fig. 9 presents a statistic showing the number of studies including each voice feature. The most sought-after features by researchers are MFCCs. More than a half of the included studies use MFCC features (55 occurrences), while the second rank is occupied, ex aequo, by shimmer and jitter with 12 occurrences. Speech signal spectrogram and mel-spectrogram were exploited 10 and 8 times, respectively. Other features are less frequently utilized. For instance, the raw signal was used four times, and HNR was used three times.

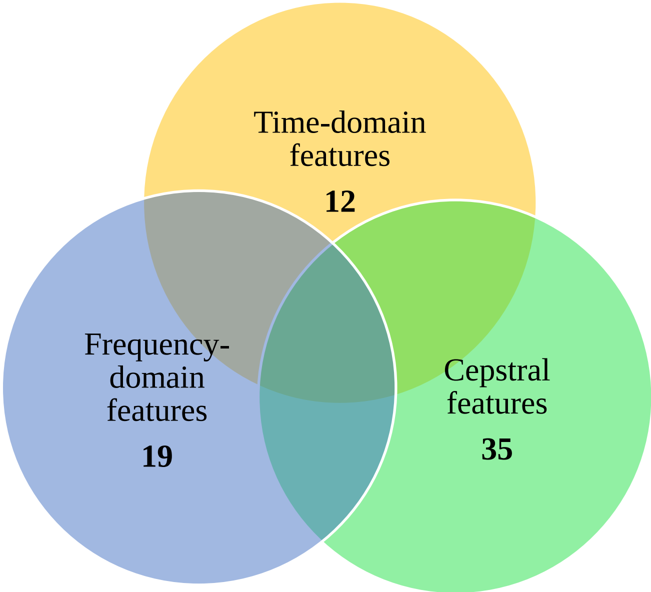


Fig. 8 Feature class utilization (temporal, spectral, cepstral)

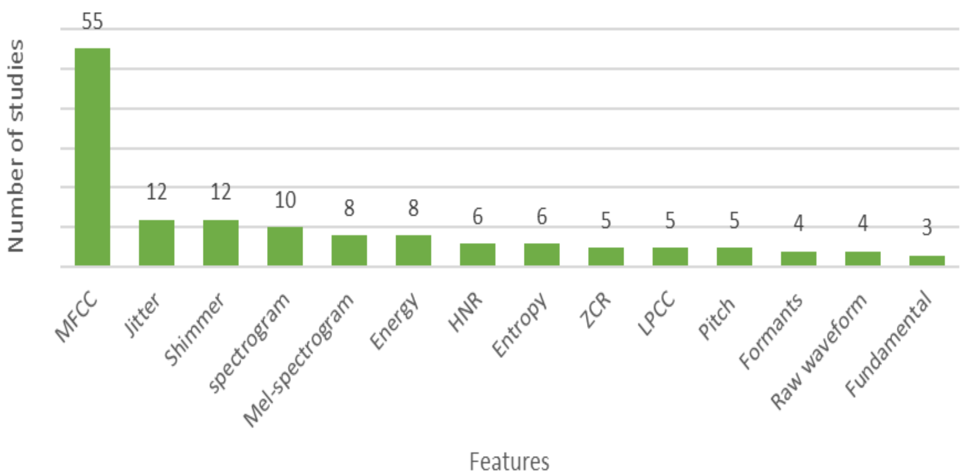


Fig. 9 Features utilization: number of studies for the most commonly used features

Discussion

Both qualitative and quantitative analyses of the review findings addressed key elements of voice disorder detection and classification systems related to laryngological diseases. It was challenging to focus the research on a specific algorithm or database. This review confirmed the statement that the SVD [91] and MEEI [131] databases are the most widely used in this field, supporting the findings established by [43] and confirmed by [108] and [109]. Other well-known databases, such as HUPA, VOICED, and AVPD, were also utilized by researchers. Regarding the number of diseases addressed, SVD offers the broadest scope, encompassing 12 diseases, and is consequently the most frequently used database for multiclass classification tasks. These databases present certain limitations that warrant attention. A significant concern is the uneven distribution of samples across different classes. The MEEI database exemplifies this issue, containing 657 pathological voice samples compared to merely 53 normal voice samples. This pronounced class imbalance can adversely affect the reliability and generalizability of classification outcomes [24, 35].

In this work, classification algorithms were also examined in terms of accuracy and frequency of use. SVM is the most frequently used algorithm (51 occurrences) and achieved excellent accuracies of 100% [6] and 99.98% [41]. Additionally, it was necessary to address the question: “Which approach is more prevalent, traditional machine learning or deep learning?”. Deep learning algorithms are gaining prominence, as the results indicate, given their capacity to learn from raw data via hidden layers without the need for transformation [71]. However, both categories continue to demonstrate highly satisfactory results, including SVM, SGD, LSTM [48], DT, ANN, and CNN [107] algorithms. While it’s true that classical ML algorithms are easy to implement, they generally require signal pre-processing, signal extraction calculations, and sometimes voice characteristic selection to improve model performance [69, 109]. Therefore, feature selection is a crucial technique during the preprocessing phase while creating ML models. This approach involves identifying the most relevant features in the data by removing those that are redundant, irrelevant, or inappropriate (e.g., minimum redundancy, maximum relevance, principal component analysis, and LDA). Many studies employed this technique to remove irrelevant features using the variance technique to reduce processing time and improve accuracy [71].

MFCC remain by far the most used. More than half of the papers adopted them. Generally, these coefficients are standard when calculated with a common framework and standard values (pre-emphasis gain, window length, mel-filterbank, etc.). Some authors used MFCCs and their derivatives (delta MFCCs, delta-delta MFCCs, etc.) to enhance voice signal dynamic representation and improve the performance of disease detection and classification [26, 105]. The other acoustic features, such as HNR, which assesses the ratio of harmonic sound energy to noise energy and indicates breathiness or hoarseness, pitch, jitter, shimmer, measuring fundamental frequency, frequency variation, and amplitude variation, respectively, were also utilized in many studies. Spectral features were employed in more than 20 articles, especially when high-dimensional data was needed, such as when using deep learning.

Numerous researchers have incorporated handcrafted features, comprising a mixture of spectral, cepstral, and other acoustic characteristics. For example, the authors in [120, 130] leveraged MFCC, GTCC, waveforms, and glottal flow features. In contrast, [111] employed 13 MFCC features, along with pitch, roll-off, ZCR, energy, entropy, spectral flux, spectral centroid, and energy to generate enhanced voice samples for subsequent processing. The frame length parameter in MFCC computation constitutes another non-standardized aspect of speech signal processing, with researchers selecting different windowing durations. In [130], the author used a 50 ms frame and a frame shift of 25 ms, while [62] used 25 ms frames with a 5 ms overlap, and [36] used 16 ms windowed signal using a shift of 8 ms. Sometimes, speech spectrograms are combined with electroglottographic signals [84]. This disparity complicates feature comparison in terms of efficiency and ease of calculation, especially when the formulas are not provided. In response to the inherent difficulties associated with feature extraction and selection, there's a growing trend among researchers toward end-to-end systems [86]. Predominantly rooted in deep learning methodologies, these systems exploit their advanced capabilities to directly process raw input signals, streamlining the overall analytical pipeline [86].

This work emphasizes the pivotal role of ML and deep learning in the detection and classification of laryngeal pathologies. Their value is multifaceted, extending to ensuring early detection (particularly crucial for conditions like laryngeal cancer), enabling the objective assessment of disease severity, and supporting advancements in telemedicine.

Conclusion

The field of voice analysis for the detection and classification of laryngeal pathologies is experiencing rapid growth. This paper provides a comprehensive analysis of the evolution of machine learning and deep learning methodologies within this domain over the past decade. It systematically presents a compilation of high-quality research published between 2015-2024 that addresses this subject. Through a detailed examination of these works, we focused on the critical components of a system designed for laryngeal disease detection and classification via voice analysis, including an analysis of key voice features and the prevalent databases employed, alongside an evaluation of the performance of various classification algorithms. Our study also corroborates the current shift towards deep learning approaches. Besides, statistical insights into voice characteristic usage offer a comprehensive overview of their application and efficacy. Moreover, a list has been provided detailing the diseases most frequently targeted by researchers, categorized by their binary or multi-class classifications.

This compilation can serve as a valuable resource for researchers when considered alongside other relevant reviews. This review aims to promote enhanced reproducibility in research and to present the current state of the art. Last but not least, it serves as a strong motivation for us to develop an embedded deep learning system for laryngeal pathology detection.

References

1. Aicha A. B. (2020). Conventional Machine Learning Techniques with Features Engineering for Preventive Larynx Cancer Detection, 5th International Conference on Advanced Technologies for Signal and Image Processing, 1-5.
2. Akbari A., M. K. Arjmandi (2015). Employing Linear Prediction Residual Signal of Wavelet Sub-bands in Automatic Detection of Laryngeal Pathology, Biomedical Signal Processing and Control, 18, 293-302.
3. Al-Dhief F. T., M. M. Baki, N. M. A. A. Latiff, N. N. N. A. Malik, et al. (2021). Voice Pathology Detection and Classification by Adopting Online Sequential Extreme Learning Machine, IEEE Access, 9, 77293-77306.
4. Al-Dhief F. T., N. M. A. Latiff, N. N. N. A. Malik, M. M. Baki, et al. (2022). Dysphonia Detection Based on Voice Signals Using Naive Bayes Classifier, 6th International Symposium on Telecommunication Technologies, 56-61.
5. Alharbi N. S., S. Alotaibi, A. I. Alnughaythir, F. Abohelaibah, et al. (2024). Prevalence and Risk Factors of Voice Disorders Among Teachers in Saudi Arabia, Cureus, 16, e56540.
6. Ali Z., M. Alsulaiman, I. Elamvazuthi, G. Muhammad, et al. (2016). Voice Pathology Detection Based on the Modified Voice Contour and SVM, Biologically Inspired Cognitive Architectures, 15, 10-18.

7. Al-Nasheri A., G. Muhammad, M. Alsulaiman, Z. Ali, et al. (2018). Voice Pathology Detection and Classification Using Auto-correlation and Entropy Features in Different Frequency Regions, IEEE Access, 6, 6961-6974.
8. Al-Nasheri A., G. Muhammad, M. Alsulaiman, Z. Ali, et al. (2017). An Investigation of Multidimensional Voice Program Parameters in Three Different Databases for Voice Pathology Detection and Classification, Journal of Voice, 31(1), 113-e9.
9. Alves M., G. Silva, B. C. Bispo, M. E. Dajer, et al. (2023). Voice Disorders Detection Through Multiband Cepstral Features of Sustained Vowel, Journal of Voice, 37(3), 322-331.
10. Amami R., A. Smiti (2017). An Incremental Method Combining Density Clustering and Support Vector Machines for Voice Pathology Detection, Computers & Electrical Engineering, 57, 257-265.
11. Amara F., M. Fezari, H. Bourouba (2016). An Improved GMM-SVM System Based on Distance Metric for Voice Pathology Detection, Applied Mathematics and Information Sciences, 10, 1061-1070.
12. Ankışhan H., S. Ç. İnam (2021). Voice Pathology Detection by Using the Deep Network Architecture, Applied Soft Computing, 106, 107310.
13. Barbon S., R. C. Guido, G. J. Aguiar, E. J. Santana, et al. (2023). Multiple Voice Disorders in the Same Individual: Investigating Handcrafted Features, Multi-label Classification Algorithms, and Base-learners, Speech Communication, 152, 102952.
14. Barlow J., Z. Sragi, G. Rivera-Rivera, A. Al-Awady, et al. (2024). The Use of Deep Learning Software in the Detection of Voice Disorders: A Systematic Review, Otolaryngology-Head and Neck Surgery, 170(6), 1531-1543.
15. Barreira R. R. A., L. L. Ling (2020). Kullback-Leibler Divergence and Sample Skewness for Pathological Voice Quality Assessment, Biomedical Signal Processing and Control, 57, 101697.
16. Ben Aicha A., K. Ezzine (2016). Cancer Larynx Detection Using Glottal Flow Parameters and Statistical Tools, International Symposium on Signal, Image, Video and Communications, 65-70.
17. Cai J., Y. Song, J. Wu, X. Chen (2024). Voice Disorder Classification Using Wav2vec 2.0 Feature Extraction, Journal of Voice, S0892-1997(24), 00293-5.
18. Calà F., L. Frassineti, C. Manfredi, P. Dejonckere, et al. (2023). Machine Learning Assessment of Spasmodic Dysphonia Based on Acoustical and Perceptual Parameters, Bioengineering, 10(4), 426.
19. Cesari U., G. De Pietro, E. Marciano, C. Niri, et al. (2018). A New Database of Healthy and Pathological Voices, Computers & Electrical Engineering, 68, 310-321.
20. Chaiani M., S. A. Selouani, M. Boudraa, M. S. Yakoub (2022). Voice Disorder Classification Using Speech Enhancement and Deep Learning Models, Biocybernetics and Biomedical Engineering, 42, 463-480.
21. Chechi R. (2013). Performance Analysis of MFCC and LPCC Techniques in Automatic Speech Recognition, International Journal of Engineering Research & Technology, 2, 3142-3146.
22. Chen L., J. Chen (2022). Deep Neural Network for Automatic Classification of Pathological Voice Signals, Journal of Voice, 36(2), 288-e15.
23. Chen L., C. Wang, J. Chen, Z. Xiang, et al. (2021). Voice Disorder Identification by Using Hilbert-Huang Transform (HHT) and K Nearest Neighbor (KNN), Journal of Voice, 35(6), 932-e1.
24. Chicco D., G. Jurman (2020). The Advantages of the Matthews Correlation Coefficient (MCC) over F1 Score and Accuracy in Binary Classification Evaluation, BMC Genomics, 21(1), 6.

25. Coelho S., H. L. Shashirekha (2023). Identification of Voice Disorders: A Comparative Study of Machine Learning Algorithms, International Conference on Speech and Computer, 565-578.
26. Cordeiro H., J. Fonseca, I. Guimarães, C. Meneses (2017). Hierarchical Classification and System Combination for Automatically Identifying Physiological and Neuromuscular Laryngeal Pathologies, Journal of Voice, 31(3), 384-e9.
27. Crowson M. G., J. Ranisau, A. Eskander, A. Babier, et al. (2020). A Contemporary Review of Machine Learning in Otolaryngology-Head and Neck Surgery, Laryngoscope, 130, 45-51.
28. Dahmani M., M. Guerti (2017). Vocal Folds Pathologies Classification Using Naïve Bayes Networks, 6th International Conference on Systems and Control, 426-432.
29. Di Cesare M. G., D. Perpetuini, D. Cardone, A. Merla (2024). Machine Learning-assisted Speech Analysis for Early Detection of Parkinson's Disease: A Study on Speaker Diarization and Classification Techniques, Sensors, 24, 1499.
30. Ding H., Z. Gu, P. Dai, Z. Zhou, et al. (2021). Deep Connected Attention (DCA) ResNet for Robust Voice Pathology Detection and Classification, Biomedical Signal Processing and Control, 70, 102973.
31. Disken G. (2024). Multilabel Voice Disorder Classification Using Raw Waveforms, Turkish Journal of Electrical Engineering and Computer Sciences, 32(4), 590-604.
32. El Ayadi M., M. S. Kamel, F. Karray (2011). Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases, Pattern Recognition, 44, 572-587.
33. Eskidere Ö., A. Gürhanlı (2015). Voice Disorder Classification Based on Multitaper Mel Frequency Cepstral Coefficients Features, Computational and Mathematical Methods in Medicine, 2015(1), 956249.
34. Ezzine K., A. B. Hamida, Z. B. Messaoud, M. Frikha (2016). Towards a Computer Tool for Automatic Detection of Laryngeal Cancer, International Conference on Advanced Technologies for Signal and Image Processing, 387-392. 34.
35. Fan Z., Y. Wu, C. Zhou, X. Zhang, et al. (2021). Class-imbalanced Voice Pathology Detection and Classification Using Fuzzy Cluster Oversampling Method, Applied Sciences, 11, 3450.
36. Fang S. H., Y. Tsao, M. J. Hsiao, J. Y. Chen, et al. (2019). Detection of Pathological Voice Using Cepstrum Vectors: A Deep Learning Approach, Journal of Voice, 33(5), 634-641.
37. Francis C. R., V. V. Nair, S. Radhika (2016). A Scale Invariant Technique for Detection of Voice Disorders Using Modified Mellin Transform, International Conference on Emerging Technological Trends, 1-6.
38. Fujimura S., T. Kojima, Y. Okanoue, K. Shoji, et al. (2022). Classification of Voice Disorders Using a One-Dimensional Convolutional Neural Network, Journal of Voice, 36(1), 15-20.
39. Geng L., Y. Liang, H. Shan, Z. Xiao, et al. (2022). Pathological Voice Detection and Classification Based on Multimodal Transmission Network, Journal of Voice, 39(3), 591-601.
40. Ghasemzadeh H. (2015). Detection of Vocal Disorders Based on Phase Space Parameters and Lyapunov Spectrum, Biomedical Signal Processing and Control, 22, 135-145.
41. Gidaye G., J. Nirmal, K. Ezzine, M. Frikha (2020). Wavelet Sub-band Features for Voice Disorder Detection and Classification, Multimedia Tools and Applications, 79(39), 28499-28523.
42. Godino-Llorente J. I., P. Gómez-Vilda, F. Cruz-Roldán, M. Blanco-Velasco, et al. (2010). Pathological Likelihood Index as a Measurement of the Degree of Voice Normality and Perceived Hoarseness, Journal of Voice, 24(6), 667-677.

43. Gomez P., A. Schutzenberger, M. Semmler, M. Dollinger (2019). Laryngeal Pressure Estimation with a Recurrent Neural Network, IEEE Journal of Translational Engineering in Health and Medicine, 7, 1-11.
44. Gour G. B., V. Udayashankara, D. K. Badakh, Y. A. Kulkarni (2020). Voice-disorder Identification of Laryngeal Cancer Patients, International Journal of Advanced Computer Science and Applications, 11(11).
45. Grzywalski T., A. Maciaszek, A. Biniakowski, J. Orwat, et al. (2018). Parameterization of Sequence of MFCCs for DNN-based Voice Disorder Detection, IEEE International Conference on Big Data (Big Data), 5247-5251.
46. Guan H., A. Lerch (2019). Evaluation of Feature Learning Methods for Voice Disorder Detection, International Journal of Semantic Computing, 13(04), 453-470.
47. Guan H., A. Lerch (2019). Learning Strategies for Voice Disorder Detection, 13th International Conference on Semantic Computing, 295-301.
48. Guedes V., A. Junior, J. Fernandes, F. Teixeira, et al. (2018). Long Short Term Memory on Chronic Laryngitis Classification, Procedia Computer Science, 138, 250-257.
49. Guedes V., F. Teixeira, A. Oliveira, J. Fernandes, et al. (2019). Transfer Learning with AudioSet to Voice Pathologies Identification in Continuous Speech, Procedia Computer Science, 164, 662-669.
50. Hammami I., L. Salhi, S. Labidi (2016). Pathological Voices Detection Using Support Vector Machine, 2nd International Conference on Advanced Technologies for Signal and Image Processing, 662-666.
51. Hammami I., L. Salhi, S. Labidi (2020). Voice Pathologies Classification and Detection Using EMD-DWT Analysis Based on Higher Order Statistic Features, IRBM, 41(3), 161-171.
52. Harar P., J. B. Alonso-Hernandezy, J. Mekyska, Z. Galaz, et al. (2017). Voice Pathology Detection Using Deep Learning: A Preliminary Study, International Conference and Workshop on Bioinspired Intelligence, 1-4.
53. Harar P., Z. Galaz, J. B. Alonso-Hernandez, J. Mekyska, et al. (2018). Towards Robust Voice Pathology Detection: Investigation of Supervised Deep Learning, Gradient Boosting, and Anomaly Detection Approaches Across Four Databases, Neural Computing and Applications, 32(20), 15747-15757.
54. Hegde S., S. Shetty, S. Rai, T. Dodderi (2019). A Survey on Machine Learning Approaches for Automatic Detection of Voice Disorders, Journal of Voice, 33(6), 947-e11.
55. Hemmerling D. (2017). Voice Pathology Distinction Using Autoassociative Neural Networks, 25th European Signal Processing Conference, 1844-1847.
56. Hemmerling D., A. Skalski, J. Gajda (2016). Voice Data Mining for Laryngeal Pathology Assessment, Computers in Biology and Medicine, 69, 270-276.
57. Heo S., K. E. Uhm, D. Yuk, B. M. Kwon, et al. (2024). Deep Learning Approach for Dysphagia Detection by Syllable-based Speech Analysis with Daily Conversations, Scientific Reports, 14(1), 20270.
58. Hu H. C., S. Y. Chang, C. H. Wang, K. J. Li, et al. (2021). Deep Learning Application for Vocal Fold Disease Prediction Through Voice Recognition: Preliminary Development Study, Journal of Medical Internet Research, 23(6), e25247.
59. Huckvale M., Z. Liu, C. Buciuleac (2023). Automated Voice Pathology Discrimination from Audio Recordings Benefits from Phonetic Analysis of Continuous Speech, Biomedical Signal Processing and Control, 86, 105201.
60. Javanmardi F., S. R. Kadiri, P. Alku (2024). A Comparison of Data Augmentation Methods in Voice Pathology Detection, Computer Speech & Language, 83, 101552.

61. Javanmardi F., S. R. Kadiri, M. Kodali, P. Alku (2022). Comparing 1-dimensional and 2-dimensional Spectral Feature Representations in Voice Pathology Detection Using Machine Learning and Deep Learning Classifiers, Interspeech, 2173-2177.
62. Kadiri S. R., P. Alku (2020). Analysis and Detection of Pathological Voice Using Glottal Source Features, IEEE Journal of Selected Topics in Signal Processing, 14, 367-379.
63. Kassim F. N. C., V. Vijean, H. Muthusamy, R. Abdullah, et al. (2019). Voice Pathology Analysis Using DT-CWPT and ReliefF Algorithm, Journal of Physics: Conference Series, 1372(1), 012029.
64. Kim H., J. Jeon, Y. J. Han, Y. Joo, et al. (2020). Convolutional Neural Network Classifies Pathological Voice Change in Laryngeal Cancer with High Accuracy, Journal of Clinical Medicine, 9, 3415.
65. Kim H., H. Y. Park, D. Park, S. Im, et al. (2023). Non-invasive Way to Diagnose Dysphagia by Training Deep Learning Model with Voice Spectrograms, Biomedical Signal Processing and Control, 86, 105259.
66. Kim H. B., J. Song, S. Park, Y. O. Lee (2024). Classification of Laryngeal Diseases Including Laryngeal Cancer, Benign Mucosal Disease, and Vocal Cord Paralysis by Artificial Intelligence Using Voice Analysis, Scientific Reports, 14(1), 9297.
67. Korba M. C. A., H. Doghmane, K. Khelil, K. Messaoudi (2024). Improved laryngeal Pathology Detection Based on Bottleneck Convolutional Networks and MFCC, IEEE Access, 12, 124801-124815.
68. Krishnamurthi R., D. Gopinathan, A. Kumar (2022). Using Wavelet Transformation for Acoustic Signal Processing in Heavy Vehicle Detection and Classification, Autonomous and Connected Heavy Vehicle Technology, 199-209.
69. Lee J. B., H. G. Lee (2024). Quantitative Analysis of Automatic Voice Disorder Detection Studies for Hybrid Feature and Classifier Selection, Biomedical Signal Processing and Control, 91, 106014.
70. Lee J. Y. (2021). Experimental Evaluation of Deep Learning Methods for an Intelligent Pathological Voice Detection System Using the Saarbruecken Voice Database, Applied Sciences, 11(15), 7149.
71. Leite D. R. A., R. M. de Moraes, L. W. Lopes (2022). Different Performances of Machine Learning Models to Classify Dysphonic and Non-Dysphonic Voices, Journal of Voice, 39(3), 577-590.
72. Liberati A., D. G. Altman, J. Tetzlaff, C. Mulrow, et al. (2009) The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies that Evaluate Health Care Interventions: Explanation and Elaboration, British Medical Journal, 339.
73. Lima-Filho L. M. A., L. W. Lopes, T. D. M. e Silva Filho (2024). Integrated Vocal Deviation Index (IVDI): A Machine Learning Model to Classifier of the General Grade of Vocal Deviation, Journal of Voice, S0892-1997 (24) 00384-9.
74. Mahmoud S. S., A. Kumar, Y. Li, Y. Tang, et al. (2021). Performance Evaluation of Machine Learning Frameworks for Aphasia Assessment, Sensors, 21, 2582.
75. Massachusetts Eye and Ear Infirmary (MEEI), Voice Disorders Database, Version 1.03 (CD-ROM), Kay Elemetrics Corporation, Lincoln Park, New York, USA, 1994.
76. Mekyska J., E. Janousova, P. Gomez-Vilda, Z. Smekal, et al. (2015). Robust and Complex Approach of Pathological Speech Signal Analysis, Neurocomputing, 167, 94-111.
77. Mesallam T. A., M. Farahat, K. H. Malki, M. Alsulaiman, et al. (2017). Development of the Arabic Voice Pathology Database and Its Evaluation by Using Speech Features and Machine Learning Algorithms, Journal of Healthcare Engineering, 2017(1), 8783751.
78. Mittapalle K. R., M. K. Yagnavajjula, P. Alku (2023). Classification of Functional Dysphonia Using the Tunable Q Wavelet Transform, Speech Communication, 155, 102989.

79. Mohammed H. M. A., A. N. Omeroglu, E. A. Oral (2023). MMHFNet: Multi-modal and Multi-layer Hybrid Fusion Network for Voice Pathology Detection, Expert Systems with Applications, 223,119790.
80. Mohammed M., K. Abdulkareem, S. Mostafa, M. Maashi, et al. (2020). Voice Pathology Detection and Classification Using Convolutional Neural Network Model, Applied Sciences, 10(11), 3723.
81. Morawska J., E. Niebudek-Bogusz (2017). Risk Factors and Prevalence of Voice Disorders in Different Occupational Groups – A Review of Literature, Otorynolaryngologia, 16(3), 94-102.
82. Morikawa M., D. H. Spatti, M. E. Dajer (2022). Wavelet Packet Transform and Multilayer Perceptron to Identify Voices with a Mild Degree of Vocal Deviation, Revista de Investigación e Innovación en Ciencias de la Salud, 4(1), 16-25.
83. Muhammad G., M. Alhussein (2021). Convergence of Artificial Intelligence and Internet of Things in Smart Healthcare: A Case Study of Voice Pathology Detection, IEEE Access, 9, 89198-89209.
84. Muhammad G., M. Alsulaiman, Z. Ali, T. Mesallam, et al. (2017). Voice Pathology Detection Using Interlaced Derivative Pattern on Glottal Source Excitation, Biomedical Signal Processing and Control, 31, 156-164.
85. Muhammad G., G. Altuwaijri, M. Alsulaiman, Z. Ali, et al. (2016). Automatic Voice Pathology Detection and Classification Using Vocal Tract Area Irregularity, Biocybernetics and Biomedical Engineering, 36, 309-317.
86. Narendra N. P., P. Alku (2020). Glottal Source Information for Pathological Voice Detection, IEEE Access, 8, 67745-67755.
87. Özbay E., F. A. Özbay, N. Khodadadi, F. S. Gharehchopogh, et al. (2024). Multifeature Fusion Method with Metaheuristic Optimization for Automated Voice Pathology Detection, Journal of Voice, doi.org/10.1016/j.jvoice.2024.08.018.
88. Paulraj M. P., S. Yaacob, M. Hariharan (2009). Diagnosis of Vocal Fold Pathology Using Time-domain Features and Systole Activated Neural Network, 5th International Colloquium on Signal Processing & Its Applications, 29-32.
89. Peng X., H. Xu, J. Liu, J Wang, et al. (2023). Voice Disorder Classification Using Convolutional Neural Network Based on Deep Transfer Learning, Scientific Reports, 13(1), 7264.
90. Pham M., J. Lin, Y. Zhang (2018). Diagnosing Voice Disorder with Machine Learning, International Conference on Big Data (Big Data), 5263-5266.
91. Powell M. E, M. Rodriguez Cancio, D. Young, W. Nock, et al. (2019). Decoding Phonation with Artificial Intelligence (DeP AI): Proof of Concept, Laryngoscope Investigative Otolaryngology, 4(3), 328-334.
92. Prithvi P., T. K. Kumar (2016). Comparative Analysis of MFCC, LFCC, RASTA-PLP, International Journal of Scientific Engineering and Research, 4(5), 1-4.
93. Pützer M., J. Koreman (1997). A German Database of Patterns of Pathological Vocal Fold Vibration, Phonus, 3, 143-153.
94. Rehman M. (2024). Voice Disorder Detection Using Machine Learning Algorithms: An Application in Speech and Language Pathology, Engineering Applications of Artificial Intelligence, 133, 108047.
95. Reid J. (2022). Development of a Machine-learning Based Voice Disorder Screening Tool, American Journal of Otolaryngology, 43, 103327.
96. Rejaibi E., A. Komaty, F. Meriaudeau, S. Agrebi, et al. (2022). MFCC-based Recurrent Neural Network for Automatic Clinical Depression Recognition and Assessment from Speech, Biomedical Signal Processing and Control, 71, 103107.

97. Ribas D., M. A. Pastor, A. Miguel, D. Martínez, et al. (2023). Automatic Voice Disorder Detection Using Self-supervised Representations, IEEE Access, 11, 14915-14927.
98. Sabir B., F. Rouda, Y. Khazri, B. Touri, et al. (2017). Improved Algorithm for Pathological and Normal Voices Identification, International Journal of Electrical and Computer Engineering, 7, 238-243.
99. Sayadi M., M. Langarizadeh, F. Torabinezhad, G. Bayazian (2024). Voice as an Indicator for Laryngeal Disorders Using Data Mining Approach, Frontiers in Health Informatics 13(1).
100. Sayadi M., V. Varadarajan, M. Langarizadeh, G. Bayazian, et al. (2022). A Systematic Review on Machine Learning Techniques for Early Detection of Mental, Neurological and Laryngeal Disorders Using Patient's Speech, Electronics, 11, 4235.
101. Seedat N., V. Aharonson, Y. Hamzany (2020). Automated and Interpretable M-health Discrimination of Vocal Cord Pathology Enabled by Machine Learning, IEEE Asia-Pacific Conference on Computer Science and Data Engineering, 1-6.
102. Sezgin M., B. Gunsel, G. K. Kurt (2012). Perceptual Audio Features for Emotion Detection, EURASIP Journal on Audio, Speech, and Music Processing, 2012(1), 16.
103. Shaikh A. A. S., M. S. Bhargavi, G. R. Naik (2023). Unraveling the Complexities of Pathological Voice Through Saliency Analysis, Computers in Biology and Medicine, 166, 107566.
104. Shrivas A., S. Deshpande, G. Gidaye, J. Nirmal, et al. (2022). Employing Energy and Statistical Features for Automatic Diagnosis of Voice Disorders, Diagnostics, 12(11), 2758.
105. Singh Y. B., S. Goel (2022). A Systematic Literature Review of Speech Emotion Recognition Approaches, Neurocomputing, 492, 245-263.
106. Sodre B., M. Rosa, A. Dassie-Leite (2017). Evaluating the Use of Neural Networks and Acoustic Measurements to Identify Laryngeal Pathologies, 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 4590-4593.
107. Song J., H. Kim, Y. O. Lee (2024). Laryngeal Disease Classification Using Voice Data: Octave-band vs. Mel-frequency Filters, Heliyon, 10(24).
108. Souissi N., A. Cherif (2015). Dimensionality Reduction for Voice Disorders Identification System Based on Mel Frequency Cepstral Coefficients and Support Vector Machine, 7th International Conference on Modelling, Identification and Control, 1-6.
109. Souza T. A. D., V. J. Vieira, M. A. D. Souza, S. E. Correia, et al. (2018). Feature Selection Based on Binary Particle Swarm Optimisation and Neural Networks for Pathological Voice Detection, International Journal of Bio-Inspired Computation, 11(2), 91-101.
110. Syed S. A., M. Rashid, S. Hussain (2020). Meta-analysis of Voice Disorders Databases and Applied Machine Learning Techniques, Mathematical Biosciences and Engineering, 17(6), 7958-7979.
111. Syed S. A., M. Rashid, S. Hussain, A. Imtiaz, et al. (2021). Inter Classifier Comparison to Detect Voice Pathologies, Mathematical Biosciences and Engineering, 18, 2258-2273.
112. Syed S. A., M. Rashid, S. Hussain, H. Zahid (2021). Comparative Analysis of CNN and RNN for Voice Pathology Detection, BioMed Research International, 2021(1), 6635964.
113. Teixeira F., J. Fernandes, V. Guedes, A. Junior, et al. (2018). Classification of Control/Pathologic Subjects with Support Vector Machines, Procedia Computer Science, 138, 272-279.
114. Teixeira J. P., P. O. Fernandes, N. Alves (2017). Vocal Acoustic Analysis – Classification of Dysphonic Voices with Artificial Neural Networks, Procedia Computer Science, 121, 19-26.
115. Thennal D. K., V. V. Nair, R. Indudharan, D. P. Gopinath (2023). Performance Enhancement of Deep Neural Network Based Automatic Voice Disorder Detection

System with Data Augmentation – Detection of Leukoplakia: A Case Study, Biomedical Engineering: Applications, Basis and Communications, 35(01), 2250041.
116. Tirronen S., S. R. Kadiri, P. Alku (2022). The Effect of the MFCC Frame Length in Automatic Voice Pathology Detection, Journal of Voice, 38(5), 975-982.
117. Tirronen S., S. R. Kadiri, P. Alku (2023). Hierarchical Multi-class Classification of Voice Disorders Using Self-supervised Models and Glottal Features, Open Journal of Signal Processing, 4, 80-88.
118. Verde L., G. De Pietro, M. Alrashoud, A. Ghoneim, et al. (2019). Leveraging Artificial Intelligence to Improve Voice Disorder Identification Through the Use of a Reliable Mobile App, IEEE Access, 7, 124048-124054.
119. Verde L., G. De Pietro, G. Sannino (2018). Voice Disorder Identification by Using Machine Learning Techniques, IEEE Access, 6, 16246-16255.
120. Vieira V. J. D., S. C. Costa, S. E. N. Correia (2023). Non-stationarity-based Adaptive Segmentation Applied to Voice Disorder Discrimination, IEEE Access, 11, 54750-54759.
121. Vinod H., R. K. Sharma, R. Shandilya (2018). Dysphonic Voice Detection Using MDVP Parameters, International Students' Conference on Electrical, Electronics and Computer Science, 1-4.
122. Wahengbam K., M. P. Singh, K. Nongmeikapam, A. D. Singh (2021). A Group Decision Optimization Analogy-based Deep Learning Architecture for Multiclass Pathology Classification in a Voice Signal, IEEE Sensors Journal, 21, 8100-8116.
123. Wang A., W. Zhang, X. Wei (2019). A Review on Weed Detection Using Ground-based Machine Vision and Image Processing Techniques, Computers and Electronics in Agriculture, 158, 226-240.
124. Warule P., S. P. Mishra, S. Deb (2023). Time-frequency Analysis of Speech Signal Using Chirplet Transform for Automatic Diagnosis of Parkinson's Disease, Biomedical Engineering Letters, 13(4), 613-623.
125. William J. B., P. Manfred (2007). Saarbrücken Voice Database, Institute of Phonetics, University of Saarland. Available online: https://stimmdb.coli.uni-saarland.de/ (Access date 14 November 2025).
126. Xie X, H. Cai, C. Li, Y. Wu, et al. (2023). A Voice Disease Detection Method Based on MFCCs and Shallow CNN, Journal of Voice, doi.org/10.1016/j.jvoice.2023.09.024.
127. Yagnavajjula M. K., P. Alku, K. S. Rao, P. Mitra (2022). Detection of Neurogenic Voice Disorders Using the Fisher Vector Representation of Cepstral Features, Journal of Voice, 39(3), 757-763.
128. Yagnavajjula M. K., K. R. Mittapalle, P. Alku, P. Mitra (2024). Automatic Classification of Neurological Voice Disorders Using Wavelet Scattering Features, Speech Communication, 157, 103040.
129. Za'im N. A. N., F. T. Al-Dhief, M. Azman, M. R. M. Alsemawi, et al. (2023). The Accuracy of an Online Sequential Extreme Learning Machine in Detecting Voice Pathology Using the Malaysian Voice Pathology Database, Journal of Otolaryngology-Head & Neck Surgery, 52(1), s40463-023.
130. Zhao D., Z. Qiu, Y. Jiang, X. Zhu, et al. (2024). A Depthwise Separable CNN-based Interpretable Feature Extraction Network for Automatic Pathological Voice Detection, Biomedical Signal Processing and Control, 88, 105624.
131. Zhou C., Y. Wu, Z. Fan, X. Zhang, et al. (2022). Gammatone Spectral Latitude Features Extraction for Pathological Voice Detection and Classification, Applied Acoustics, 185, 108417.

Hassan Ezzahori, M.Sc.
E-mail: hasan.ez-zahori@usms.ma

Hassan Ezzahori had received a M.Sc. Degree from Mohamed V University, Rabat, Morocco, in 2019. He is a member of the Engineering and Applied Physics Team at the Superior School of Technology, Sultan Moulay Slimane University, Beni Mellal, Morocco. His research topics are machine learning, deep learning, and embedded systems in biomedical applications.

Abdelkrim Hammimou, Ph.D. Student
E-mail: abdelkrim.hammimmou@usms.ma

Abdelkrim Hammimou was born in Settat, Morocco in 1990. He received his M.Sc. Degree in Biomedical Engineering from Hassan I University, Settat, Morocco in 2016. He is currently pursuing a Ph.D. Degree as a member of the Engineering and Applied Physics Team, at the Superior School of Technology, Sultan Moulay Slimane University, Beni Mellal, and Morocco. His current research topics include biomedical engineering, signal and image processing, and artificial intelligence applied to the biomedical field.

Prof. Abdelghani Boudaoud, Ph.D.
E-mail: abdelghani.boudaoud@usms.ma

Abdelghani Boudaoud was born in Azilal, Morocco in 1973. He received his Ph.D. Degree in Electronics and Telecommunications from Hassan I University, Settat, Morocco in 2019. In 2020, he joined as a Professor at the Mechatronics Department at Higher School of Technology, Sultan Moulay Slimane University, Beni Mellal, Morocco. His current research interests include hardware design of algorithms applied to information processing and control of industrial systems.

 doi: 10.7546/ijba.2025.29.4.001071

Prof. Mounaim Aqil, Ph.D.
E-mail: mounaim.aqil@usms.ma

Mounaim Aqil is currently a Professor in the Mechatronics Department at Higher School of Technology, Sultan Moulay Slimane University, Béni-Mellal, Morocco. In 2018, he received his Ph.D. Degree in Electrical Engineering from Mohammed V University, Rabat, Morocco. His current research interests include signal processing, embedded electronic systems and IoT. He is a member of the Engineering and Applied Physics Team.