Review

Applied Machine Learning Techniques to Diagnose Voice-Affecting Conditions and Disorders: Systematic Literature Review

Alper Idrisoglu[1], MSc; Ana Luiza Dallora[1], PhD; Peter Anderberg[1,2], PhD; Johan Sanmartin Berglund[1], MD, PhD

[1]Department of Health, Blekinge Institute of Technology, Karslkrona, Sweden

[2]School of Health Sciences, University of Skövde, Skövde, Sweden

Corresponding Author:
Alper Idrisoglu, MSc
Department of Health
Blekinge Institute of Technology
Valhallavägen 1
Karslkrona, 37141
Sweden
Phone: 46 701462619
Email: alper.idrisoglu@bth.se

Abstract

Background: Normal voice production depends on the synchronized cooperation of multiple physiological systems, which makes the voice sensitive to changes. Any systematic, neurological, and aerodigestive distortion is prone to affect voice production through reduced cognitive, pulmonary, and muscular functionality. This sensitivity inspired using voice as a biomarker to examine disorders that affect the voice. Technological improvements and emerging machine learning (ML) technologies have enabled possibilities of extracting digital vocal features from the voice for automated diagnosis and monitoring systems.

Objective: This study aims to summarize a comprehensive view of research on voice-affecting disorders that uses ML techniques for diagnosis and monitoring through voice samples where systematic conditions, nonlaryngeal aerodigestive disorders, and neurological disorders are specifically of interest.

Methods: This systematic literature review (SLR) investigated the state of the art of voice-based diagnostic and monitoring systems with ML technologies, targeting voice-affecting disorders without direct relation to the voice box from the point of view of applied health technology. Through a comprehensive search string, studies published from 2012 to 2022 from the databases Scopus, PubMed, and Web of Science were scanned and collected for assessment. To minimize bias, retrieval of the relevant references in other studies in the field was ensured, and 2 authors assessed the collected studies. Low-quality studies were removed through a quality assessment and relevant data were extracted through summary tables for analysis. The articles were checked for similarities between author groups to prevent cumulative redundancy bias during the screening process, where only 1 article was included from the same author group.

Results: In the analysis of the 145 included studies, support vector machines were the most utilized ML technique (51/145, 35.2%), with the most studied disease being Parkinson disease (PD; reported in 87/145, 60%, studies). After 2017, 16 additional voice-affecting disorders were examined, in contrast to the 3 investigated previously. Furthermore, an upsurge in the use of artificial neural network–based architectures was observed after 2017. Almost half of the included studies were published in last 2 years (2021 and 2022). A broad interest from many countries was observed. Notably, nearly one-half (n=75) of the studies relied on 10 distinct data sets, and 11/145 (7.6%) used demographic data as an input for ML models.

Conclusions: This SLR revealed considerable interest across multiple countries in using ML techniques for diagnosing and monitoring voice-affecting disorders, with PD being the most studied disorder. However, the review identified several gaps, including limited and unbalanced data set usage in studies, and a focus on diagnostic test rather than disorder-specific monitoring. Despite the limitations of being constrained by only peer-reviewed publications written in English, the SLR provides valuable insights into the current state of research on ML-based voice-affecting disorder diagnosis and monitoring and highlighting areas to address in future research.

(J Med Internet Res 2023;25:e46105) doi: 10.2196/46105

KEYWORDS

diagnosis; digital biomarkers; machine learning; monitoring; voice-affecting disorder; voice features

Introduction

Voice-Affecting Disorders

Voice and speech production relies on complex and multiorgan cooperation. The basic mechanics of speech and voice creation is that the airflow obtained by releasing the pressure in the lungs reaches the vocal folds in the larynx and vibrates the vocal cords that result in voice, and by articulating this voice speech is created [1]. The harmony between complex biological systems involved in voice and speech production leads to normal voice formation. However, at the same time, the functional dependency of several biological structures makes the voice vulnerable to being affected by diverse conditions, which may result in a pathological or disordered voice named hoarseness (ie, dysphonia).

The anomalies and the absence of vocal quality in relation to pitch, height, resonance, and duration, which are unexpected for individuals, regardless of their gender and age are characteristics of a disordered voice [2-5]. There is no globally accepted nomenclature for voice disorders. In general, structural, inflammatory, traumatic, systemic, aerodigestive, psychiatric and psychological, neurological, and functional voice disorders are substantial categories of voice problems [6]. This can be diagnosed by a health care specialist through several examinations and tests. The current approach for the diagnosis of voice disorders relies on clinical examinations consisting of interviews, perceptual voice evaluation, patient-reported outcome measures, laryngoscopy, aerodynamic assessment, voice profile, acoustic analysis, and laryngeal electromyography [7], which is time-consuming for both the patients and the clinicians and generates a high economic burden on society [8]. The appraisal based on the assessment of biomarkers gathered through clinical examinations is a crucial step that leads to a diagnosis. Here, it is necessary to point out that the clinicians do not diagnose dysphonia; instead, the target of the clinical examination is to identify the condition that leads to dysphonia, which will be addressed in this study as a voice-affecting disorder.

Voice as a Digital Biomarker

Measurable, reliable, and repeatable assets that can be correlated with a clinical outcome are defined as biomarkers. The criteria and context of use describe the category of biomarkers such as diagnostic, monitoring, pharmacodynamic/response, predictive, prognostic, and digital biomarkers [9]. Traditional biological markers (ie, biomarkers) are used to detect molecular changes associated with diseases and have been integrated with clinical practices for decades [10]. Digital biomarkers refer to measures or features collected by digital devices [11,12] and are a developing landscape that shares the same objectives as traditional biomarkers in answering health-related questions [13].

As aforementioned, voice and speech can be influenced by several conditions and disorders, which contribute to decreased quality of life. Nevertheless, being so sensitive could open possibilities for earlier diagnosis of disorders that affect the voice through the use of voice as a biomarker [14]. As the collection of voice and speech is a noninvasive process that can be performed at a low cost [11], the voice as a digital biomarker could be a diagnostic and prognostic resource with the potential to be more economically viable, in addition to being a more ecological measure than many of the currently used clinical alternatives for the assessment of cognition and function [9,15].

Machine Learning for the Assessment of Voice Signals

Most health problems could benefit from an early diagnosis for better treatment and management of outcomes. However, the growing pressure on health care systems, due to the increased life expectancy and an aging population, may hinder this early detection. Patients are usually referred to specialist care only when apparent signs of disease are present, and thus are at an already moderate advanced state. Fortunately, the existence of digital biomarkers (eg, voice), along with the trend of digitalization in health care, opens the possibility of using technologies such as machine learning (ML) to address these issues. Research on biomarkers extracted from voice and speech with ML techniques for diagnosing, prognosticating, and monitoring disease [14] has shown satisfactory outcomes for disorders such as dementia, depression, mild cognitive impairment (MCI), autism spectrum disorder, Alzheimer disease (AD), and PD [14,16-18].

ML techniques are becoming prevalent in health care for aiding decision-making in treatment and diagnosis [19]. These techniques involve extracting features from voice data and using an ML algorithm to classify the severity of disorders or to determine whether a voice is pathological. The 2 most commonly used ML techniques in this context are supervised and unsupervised learning. In supervised learning, an ML technique is trained using labeled data sets (training set) and its accuracy is evaluated using unlabeled data sets (validation and test sets). The labeled data contain the actual diagnostic information that allows the ML technique to compare its output and adjust its parameters for improved accuracy. There are also end-to-end algorithms (ie, deep learning), an ML subfield that uses artificial neural networks to model and solve complex problems. These networks are composed of multiple layers of interconnected nodes that enable them to learn hierarchical representations of data. An example of deep learning in action is the use of convolutional neural networks to classify images. Convolutional neural networks use multiple layers of convolution and pooling operations to extract features from the input image and then classify it into 1 of several categories. This approach has been very successful in tasks such as object recognition, image segmentation, and speech recognition [20]. Unsupervised learning involves applying clustering methods on training data without labels to group data through 1 or several clustering algorithms [21].

Prior studies provide comprehensive information on feature extraction and its application [22-25]. Encouraging results on

ML classifiers with voice biomarkers bring them into the focus of researchers. In a meta-analysis on voice disorders, Syed et al [26] applied ML techniques by setting the boundaries around 3 publicly available databases, namely, Saarbrucken Voice Database, Massachusetts Eye and Ear Infirmary, and Arabic voice pathology database. The systematic literature review (SLR) presented herein includes all possible data sources. Several reviews have investigated voice-based disease diagnostics with ML algorithms separately, focusing on only PD or AD [27-29], whereas in this study, multiple voice-affecting disorders are included.

This SLR investigates state of the art of clinical applications of voice-based diagnosis that make use of ML algorithms. Adapting voice-based diagnosis and prognosis into clinical practices requires solid evidence and research to clinically validate the usability and reliability of voice biomarkers and the performance of ML classifiers [12,15]. This SLR does not consider disorders directly related to the voice production mechanisms. Examples of included and excluded conditions are PD and polyps on the vocal cords, respectively, where PD is a neurodegenerative disorder that often causes voice changes [30] that are not directly related to voice box and polyps on the vocal cords occur in the voice box (larynx) [31]. More specifically, the conditions of interest in this SLR are listed as systematic conditions affecting voice, nonlaryngeal aerodigestive disorders affecting voice, and neurological disorders affecting voice in the *Classification Manual for Voice Disorders* [6]. These have a higher chance of being related to chronic conditions, which would benefit from having a scalable and noninvasive method for screening a large population. The expected outcome is a contribution, not only by summarizing the work done in the field of applied health technology that is interested in the application of the technology and its outcomes but also by pointing out the gaps in the literature and possible future research directions that could address the problems mentioned earlier of the next generations and the health care system.

Methods

Overview and Purpose of This SLR

An SLR is a summary of the results from research papers focused on a common context or a question. The summary action includes the identification, collection, assessment, and synthesizing of high-quality research evidence within the scope of the research question by following a predefined protocol. The aim of an SLR is to provide perspective on recent research so that decision makers can benefit from up-to-date knowledge and address the gaps that can be used as a basis for new research. The predefined protocol describes the methodology to follow; defines the research question; and contains information about inclusion/exclusion criteria and quality assessment [30]. This section specifies the methodology applied in this SLR to answer the research question “How is the voice as a digital biomarker being used in clinical applications that employ ML techniques for diagnosing and monitoring voice-affecting disorders?” Additionally, the main question is split into the following subquestions (SQs):

- *SQ1:* What are the aims of pathologic voice evaluation?
- *SQ2:* Which ML techniques are being used for the diagnosis and monitoring of voice-affecting disorder through voice and which voice-affecting disorders are being investigated?
- *SQ3:* What are the time and geographical trends of publications in the scope of SLR?
- *SQ4:* What are the data characteristics of the sound samples for different disorders and types of studies?
- *SQ5:* Are the studies cross-sectional or longitudinal?
- *SQ6:* How is performance being evaluated in the studies?

All the information on the methodological approach that guided the execution of this SLR is based on the prespecified SLR protocol [31].

Search Strategy

A search string was constructed by applying the *p*opulation, *i*ntervention, *c*omparison, and *o*utcome (PICO) framework [32-34]. The most used terms, suggested by authors ALD and JSB, were used to find relevant papers; the retrieved papers and their references in the field were then used to discover new adequate keywords. By adding the new keywords to the search string, a comprehensive search string was created. A customized version of the search string in Textbox 1 was used in PubMed, Scopus, and Web of Science databases to find all relevant peer-reviewed primary journal articles published between 2012 and 2022. The application of the PICO structure excludes the *comparison* due to the nature of this SLR being a characterization:

- *Population:* Disorders that affect the voice, given by the *Classification Manual for Voice Disorders* [30], referring to the systematic conditions affecting voice, nonlaryngeal aerodigestive disorders affecting voice, and neurological disorders affecting voice.
- *Intervention:* Use of ML techniques for the diagnosis or monitoring of disorders through voice samples.
- *Outcome:* Reported quantities or results such as precision and accuracy.

The search string was adapted based on the advanced search requirements of each database. The filter options were tuned to retrieve articles from January 1, 2012 to December 31, 2022. The period was chosen after consulting with experts in the medical field with regard to the development of new technologies for health care. The development of the search string was primarily based on the MeSH (Medical Subject Headings) terms, with the help of a librarian, and categories of voice disorders in the classification manual [6].

Textbox 1. Search string used in PubMed, Scopus, and Web of Science databases (search date: March 13, 2023).

((“Voice” OR “Linguistic features” OR “acoustic parameters” OR “Vocal features” OR “Vocal” OR “Vocal Cords” OR “Vocal biomarker” OR “Voice biomarkers” OR “ Speech” OR “Vowel” OR “Sound Spectrography” OR “Cepstrum Vectors”) AND (“Deep phenotyping” OR “selection” OR “extrac- tion” OR “Detection” OR “Monitoring” OR “Classification” OR “Evaluation” OR “Analysis” OR “Estimation” OR “Projection” OR “Improving” OR “Investigation” OR “Prognosis” OR “Predict*”) AND (“Sensitivity” OR “Accuracy” OR “Specificity” OR “Performance” OR “Cross-validation” OR “precision”) AND (“Voice technology” OR “Machine learning” OR “Artificial Intelligence” OR “Gaussian mixture models” OR “Support vector machines” OR “Artificial neural network” OR “Data Mining” OR “Decision Support System” OR “Clinical Support System” OR “Deep Neural Network” OR “Kernel extreme learning machine” OR “Deep Learning”) AND (“voice disorder” OR “systemic conditions” OR “aerogestive disorders” OR “neurologic disorders” OR “central nervous system disturbance” OR “Endocrine” OR “Hypothyroidism” OR “Hyperthyroidism” OR “Sexual Hormone Imbalances” OR “Hyperpituitarism” OR “Immunologic” OR “Allergic” OR “HIV” OR “Chronic Fatigue Syndrome” OR “Systemic Lupus Erythematosus” OR “Sjogren's Syndrome” OR “Scleroderma” OR “Wegener's Disease” OR “Musculo-Skeletal Conditions Affecting Voice” OR “Overuse Injury and Repetitive Strain Injury” OR “Fibromyalgia” OR “Ehler Danlos Syndrome” OR “Dehydration” OR “Respiratory Diseases Affecting Voice” OR “Asthma” OR “Chronic Obstructive Pulmonary Disease” OR “Digastric” OR “Gastroesophageal Reflux Disease” OR “Infectious Diseases of the Aerodigestive Tract” OR “Laryngotracheobronchitis” OR “Pertussis” OR “Diphtheria” OR “Pneumonia” OR “Infectious Sinusitis” OR “Tuberculosis” OR “Upper Respiratory Infection” OR “Acute Epiglottitis” OR “Syphilis” OR “Sarcoidosis” OR “Scleroma” OR “Leprosy” OR “Actinomycosis” OR “Mycotic Infections” OR “Blastomycosis” OR “Histoplasmosis” OR “Candidiasis” OR “Coccidioidomycosis” OR “Peripheral Nervous System Pathology” OR “Superior Laryngeal Nerve Pathology” OR “Unilateral Recurrent Laryngeal Nerve Paralysis” OR “Recurrent Laryngeal Nerve Paresis” OR “Bilateral Recurrent Laryngeal Nerve Paralysis--Peripheral” OR “Myasthenia Gravis” OR “Peripheral Neuropathy” OR “Enhanced Physiologic Tremor” OR “Movement Disorders” OR “Adductor Spasmodic Dysphonia” OR “Abductor Spasmodic Dysphonia” OR “Abductor Spasmodic Dysphonia” OR “Dystonic Tremor” OR “Essential Tremor” OR “Meige's Syndrome” OR “Tardive Stereotypies” OR “Tourette's Syndrome” OR “Amyotrophic Lateral Sclerosis” OR “Wallenberg Syndrome” OR “Lateral Medullary Syndrome” OR “Infarct” OR “Parkinson Disease” OR “Multiple Systems Atrophy” OR “Shy-Drager Syndrome” OR “Striatonigral Degeneration” OR “Sporadic Olivoponto- cerebellar Atrophy” OR “Progressive Supranuclear Palsy” OR “Multiple Sclerosis” OR “Cerebellar Disorders” OR “Huntington's Chorea” OR “Bilateral Recurrent Laryngeal Nerve Paralysis--Central” OR “Myoclonus” OR “Neuromuscular” OR “cardiovascular” OR “coronary artery” OR “heart attack” OR “Voice disorders” OR “Neurological disorders” OR “multiple sclerosis” OR “Myasthenia gravis ” OR “ALS” OR “Amyotrophic lateral sclerosis” OR “Parkinson's disease” OR “Multiple sclerosis” OR “Dementia” OR “Alzheimer's disease” OR “Essential tremor” OR “Major depressive disorder” OR “pathological voice” OR “voice pathology” OR “neurodegenerative” OR “Cognitive impairment” OR “Nodule” OR “Polyp” OR “Neoplasm” OR “dysphonia” OR “Hoarseness” OR “Huntington disease”))

Study Selection

The search string was used to perform an automated search on each database. The Zotero (Corporation for Digital Scholarship) bibliography software was used to collect all relevant articles from all 3 databases and to remove duplicates [35]. First, authors AI and ALD applied the inclusion and exclusion criteria in Textbox 2 to assess the titles and abstracts of the retrieved papers. The first step was to assess randomly selected 50 papers to ensure the consistency of the criteria. Then, another batch containing 50 articles was assessed. Authors AI and ALD compared the results. Upon agreeing on the consistency of the criteria, they proceeded to assess the remainder of the papers. The degree of agreement was checked statistically by comparing the results between the first and second authors with an overall agreement of 96% using the Cohen κ index. During the evaluation, the papers were categorized into 3 groups: included, excluded, and “maybe” cases that could not be assessed by the content of the title and abstract alone. At this stage, author JSB acted as the advisor and expert in the field. Furthermore, after the evaluation of all papers, the results from both authors were cross-checked, and 30 conflicts were noticed. To minimize the risk of bias, all articles marked as included, “maybe,” and conflicts were grouped for full-text reading.

All articles in the group of full-text readings underwent a quality assessment procedure to assure high-quality evidence (Textbox 3), based on guidelines proposed by Kitchenham and Charters [36]. The quality threshold was set to 11 points, which means that articles below the score of 11 points would be rejected. The threshold of 11 points was stipulated through group discussions with authors. The questionnaire was designed in 3 sections, consisting of 5 questions each, general questions, data analysis, and results. Based on the given questions, author AI performed the quality assessment by grading the studies with scores 0, 0.5, and 1 for the sections 1, 2, and 3, respectively.

Textbox 2. Inclusion and exclusion criteria for the assessment of the articles.

Inclusion criteria

- Journal study
- Primary study written in English
- Research published not earlier than 2012
- Research that uses voice as the input data
- Research that employs at least one machine learning algorithm
- Research that aims to diagnose or monitor at least one voice-affecting disorder not related to the systematic conditions affecting voice, nonlaryngeal aerodigestive disorders affecting voice, and neurological disorders affecting voice

Exclusion criteria

- A nonpeer-reviewed study
- Research written in languages other than English
- Research published before 2012 or after 2022
- Research that does not use voice as a direct input, which means research employing various nonverbal forms of data input, such as written transcriptions, digital images, videos, electroencephalogram, and signals generated during vocalization
- Research that classifies voice-affecting disease without a machine learning approach
- Research that classifies voice disorders related to conditions other than systematic conditions affecting voice, nonlaryngeal aerodigestive disorders affecting voice, and neurological disorders affecting voice

Textbox 3. Quality assessment questionnaire.

General questions

- Are the aims clearly stated?
- Is the targeted population described?
- Has it discussed the contribution of the study?
- Are gender and age considered?
- Is/are the technique(s) being implemented clearly described?

Data analysis

- Is the origin of data given?
- Is the type of data clearly described?
- Do the data consist of voice recordings?
- Is the data validation method given?
- Is there a discussion on whether the data size can be generalized for the targeted population?

Result

- Is/are the result(s) clearly discussed?
- Are all aims or questions answered?
- Was the outcome related to the target population?
- Are the limitations discussed?
- Did results compare with previous rapports?

Data Extraction

Data extraction was carried out by author AI. Table 1 shows the list of attributes, definitions, and purpose of use for data extraction.

Table 1. Collected data attributes.

Attribute	Definition
ISSN	International Standard Serial Number recorded
Title	Full title of the research
Journal	Publication venue record
Authors	All authors' names
Publication date	The publication date of the paper
Publication type	The type of publication
Origin of publication	The geographical location of the first author's institution
Targeted disorder	Investigated disorder
Database	Source of the data
Origin of data	The geographical location of data sources
Data characteristics	Type of voice recordings
Additional data	Used additional data except for voice recordings
Data sets	The number of participants
Sample size	The number of recordings
Aim of the study	Purpose of the study
Age range	The considered age range of the participants
Gender	The number of participants (by gender)
Quantitative result(s)	Presented outcome measures
Feature sets	Excluded features from voice
The proposed features	The best feature set, if exists
Applied ML[a] technique(s)	All applied ML techniques
Outcome evaluation	How the pathological voice is evaluated
Type of validation(s)	How the data set is divided
Type of study	If the study is longitudinal or cross-sectional
The proposed ML algorithm(s)	ML technique with the best outcome

[a]ML: machine learning.

Data Analysis

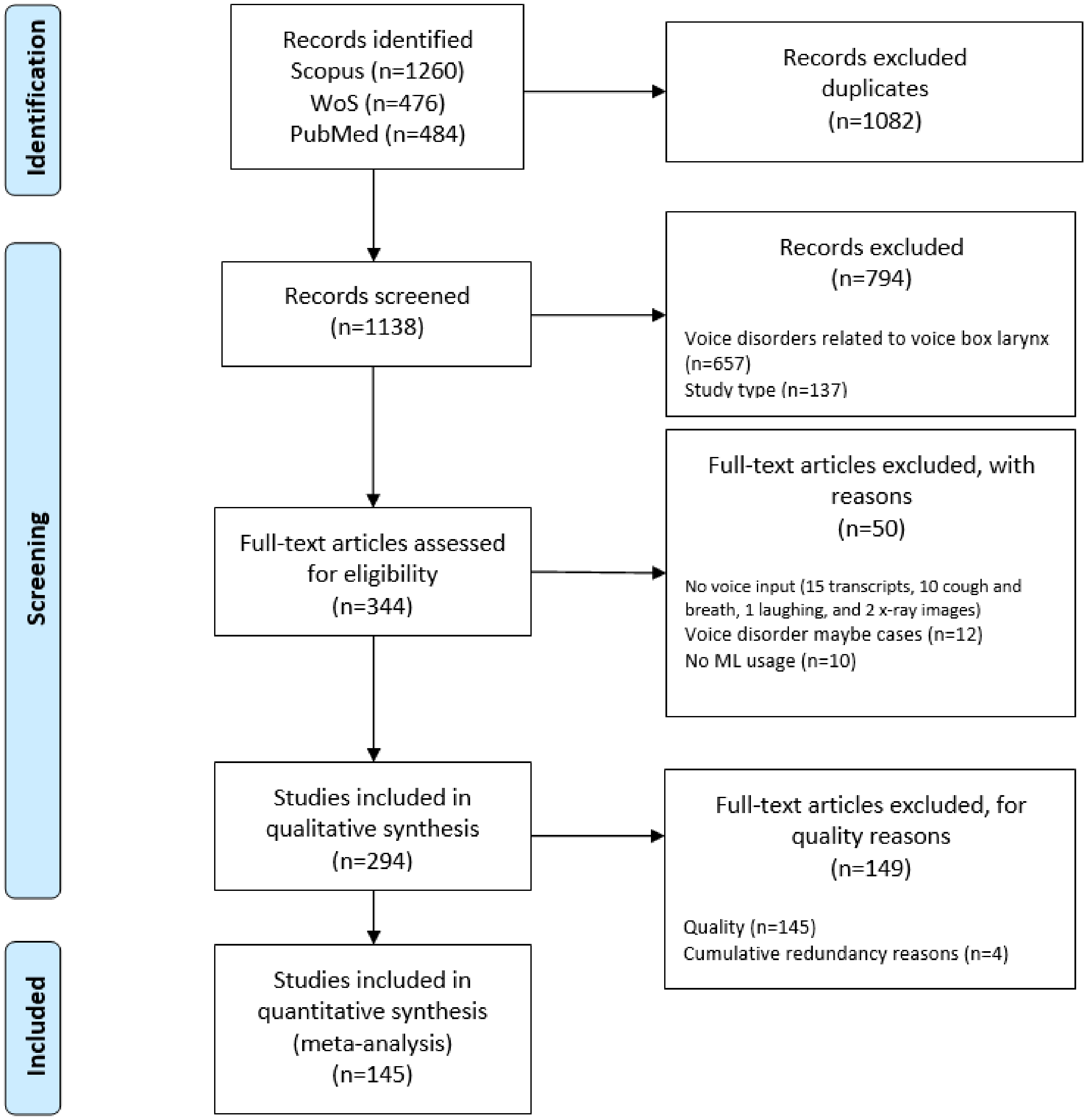
To analyze the etiology of changes over time and capture the heterogeneity, the studies were grouped into subgroup summary tables entitled with the name of disorders (see Multimedia Appendix 1). Numerical and statistical measures were used to represent the results. No assumption was made about the missing information. Microsoft Excel was used for data analysis. All the studies that successfully adhered to the inclusion and exclusion criteria and passed the quality assessment were eligible for data analysis. The results were presented in text, summary tables, and charts under a section for each research question. The robustness of the results was checked by conducting a sensitivity analysis through observations of the effect of some randomly removed data from summary tables [37,38]. The cumulative redundancy bias was checked by observing the similarity between author groups.

Results

Study Selection

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; also see Multimedia Appendices 2 and 3) flowchart for this study is shown in Figure 1 [39]. The automated search retrieved a total of 2220 articles from all 3 databases (Scopus, n=1260; Web of Science, n=476; and PubMed, n=484). After the removal of the duplicates, 1138 articles were assessed in the title and abstract screening. In total, 344 papers were included in the full-text reading group. During the full-text reading, 50 articles were found to be out of scope for the following reasons: related to voice box (n=12), voice was not an input (n=28; 15 transcripts, 10 coughs and breath, 1 laughing, and 2 x-ray images), and no ML technique (n=10) applied. A total of 294 articles were assessed for quality evaluation, which eliminated 145 articles and thus the final set included 149 articles that were used for data extraction.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart. ML: machine learning; WoS: Web of Science.



The included papers were assessed for cumulative redundancy bias during data extraction. The assessment showed that 8 of the included papers were published by 4 different author groups, 2 papers from Tuncer et al [40,41], 2 papers from Gunduz et al [42,43], 2 papers from Lamba et al [44,45] in the PD group, and 2 papers from Tena et al [46,47] in the amyotrophic lateral sclerosis (ALS) group. To reduce the risk of bias, only 4 of those 8 papers, one with the highest accuracy from each author, were included in the synthesis [40,42,44,46]. In total, 145 studies proceeded for analysis. Sensitivity analysis did not show any effect on trend analysis, but did show a minor effect on statistical analysis. Multimedia Appendix 1 shows the list of included studies in this SLR.

Aims of the Studies

To answer SQ1, this section sums up the aim and the assessment strategy used in studies. Many terminologies have been used to describe the aim of the studies. Generally, they can be arranged into 2 groups: diagnosis and monitoring. In the diagnosis group, 138 studies were identified. A total of 125 studies in the diagnostic group investigated ML methods to detect a pathological voice, where the participants were grouped as the healthy control (HC) group, with being "healthy" defined as people without a diagnosed disorder, and a group with known pathology [17,40,42,44,48-168]. The main idea was to deploy an ML technique for distinguishing those 2 groups from each other with high accuracy. Additionally, 13 studies [169-181] in the diagnostic group investigated ML techniques for separating several pathologies and clustered participants into several pathological groups. With the help of the ML technique, they tried to classify each group, where the primary purpose was to investigate a system that can classify multiple disorders. A total of 7 studies [182-188] were identified in the monitoring group. The pattern was trying to predict an established clinical severity assessment with the help of an ML algorithm where only participants with diagnosed disorders were involved.

Employed ML Techniques and Voice-Affecting Disorders

Table 2 shows the results related to SQ2. A total of 19 different disorders were identified where the focus was on monitoring or diagnosis through voice or speech with ML involvement. As many as 87/145 (60%) of the studies targeted PD; 18 studies targeted dementia or AD, 8 cognitive impairment (CI)/MCI, 4 ALS, 2 cardiovascular disorders, 7 COVID-19, 2 essential tremor, 2 multiple sclerosis, 1 neurodegenerative cognitive complaint (NCC), 1 functional dysphagia/oropharyngeal dysphagia, 4 depression, 1 influenza disease, 1 neurological disease (ND), 2 stroke, 1 fatigue, 1 autism, 1 traumatic brain injury, 1 asthma, and 1 chronic obstructive pulmonary disease. NCC and ND may potentially be classified within one of either PD, AD, or CI/MCI due to their similar symptoms, but the specific underlying disorder was not provided in the studies. Therefore, these 2 disorders were grouped separately.

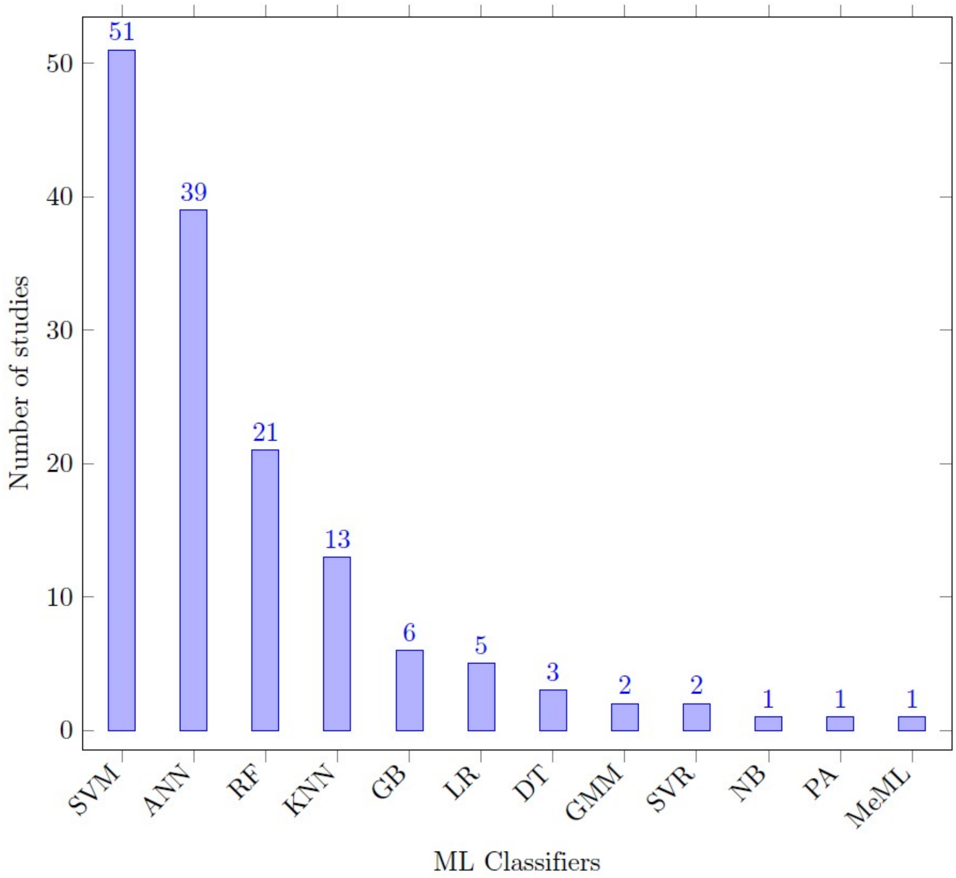
Table 2. Targeted disorders and ML[a] techniques.

Disorder	NR[b]	ML technique (NR of usage)	References
Parkinson disease	87	SVM[c] (34), ANN[d] (23), RF[e] (9), KNN[f] (6), GB[g] (5), GMM[h] (2), NB[i] (1), DT[j] (3), SVR[k] (2), LR[l] (1), PA[m] (1)	[17,40,42,44,48-102,117-135,168,170,171,180,181,183-185,188]
Dementia, Alzheimer disease	18	SVM (8), KNN (2), LR (2), RF (3), ANN (3)	[103-110,136-139,172-176,179]
Cognitive impairment/mild cognitive impairment	8	SVM (2), LR (2), RF (2), ANN (2)	[112-116,150,177,178]
COVID-19	7	ANN (3), SVM (1), RF (1), KNN (1), GB (1)	[143-147,156,157]
Amyotrophic lateral sclerosis	4	SVM (1), RF (2), MeML[n] (1)	[46,158-160]
Depression	4	ANN (3), SVM (1)	[140-142,161]
Cardiovascular disorders	2	KNN (2)	[111,186]
Essential tremor	2	SVM (2)	[162,163]
Multiple sclerosis	2	ANN (1), RF (1)	[149,164]
Stroke	2	ANN (2)	[148,153]
Asthma	1	RF	[182]
Autism	1	ANN	[151]
Fatigue	1	SVM	[187]
Chronic obstructive pulmonary disease	1	RF	[154]
Neurodegenerative cognitive complaint	1	SVM	[165]
Functional dysphagia, oropharyngeal dysphagia	1	RF	[166]
Influenza disease	1	KNN	[167]
Neurological disease	1	KNN	[169]
Traumatic brain injury	1	ANN	[152]

[a]ML: machine learning.
[b]NR: number of studies.
[c]SVM: support vector machine.
[d]ANN: artificial neural network.
[e]RF: random forest.
[f]KNN: K-nearest neighbor.
[g]GB: gradient boosting.
[h]GMM: Gaussian mixture model.
[i]NB: naïve Bayes.
[j]DT: decision tree.
[k]SVR: support vector regression.
[l]LR: logistic regression.
[m]PA: passive aggressive.
[n]MeML: mixed effect machine learning.

The usage of the 12 ML techniques is shown in Figure 2, where the support vector machine algorithm was the most used (51/145, 35.2%) and artificial neural networks were the second most utilized technique (39/145, 26.9%) among all ML techniques. Several studies have tested and compared different algorithms. Figure 2 shows the ML technique with the best results from each study. The support vector machine notation contains all different kernel combinations, and all utilized neural network architectures are grouped under artificial neural network. As many as 11 of the 12 ML techniques shown in Figure 2 have tested on PD, 5/12 on AD, 4/12 on CI/MCI, 5/12 on COVID-19, 3/12 on ALS, 1/12 on cardiovascular disorders, 1/12 on essential tremor, 2/12 on multiple sclerosis, 1/12 on stroke, 1/12 on asthma, 1/12 on autism, 1/12 on fatigue, 1/12 on chronic obstructive pulmonary disease, 1/12 on NCC, 1/12 on functional dysphagia/oropharyngeal dysphagia, 2/12 on depression, 1/12 on influenza disease, 1/12 on ND, and 1/12 on traumatic brain injury (Table 2).

Figure 2. The usage of machine learning algorithms. ANN: artificial neural network; DT: decision tree; GB: gradient boosting; GMM: Gaussian mixture model; KNN: K-nearest neighbor; LR: logic regression; MeML: mixed effect machine learning; ML: machine learning; NB: naïve Bayes; PA: passive active; RF: random forest; SVM: support vector machine; SVR: support vector regression.



Time and Geographical Trend of the Publications

Figure 3 shows the published studies by year and the investigated disorders. The results indicate that there is an upward trend in the studies involving the application of ML for voice-affecting disorder. Up to 2016, the focus of the research was solely on PD and AD. In the last 5 years, the research on voice-based diagnosis and monitoring with ML has not only increased but also diversified in terms of the investigated voice-affecting disorders with the addition of CI/MCI, ALS, cardiovascular disorder, essential tremor, COVID-19, multiple sclerosis, NCC, functional dysphagia/oropharyngeal dysphagia, depression, influenza disease, ND, stroke, asthma, autism, fatigue, chronic obstructive pulmonary disease, and traumatic brain injury. In addition, the highest publication rate occurred in 2022 (more than doubled compared with previous years); 51/145 studies included in this SLR have been published in 2022, which corresponds to 35.1% of all listed articles in Multimedia Appendix 1.

Figure 4 displays the contribution from countries for a specific disorder. Some countries tend to focus more on 1 disorder, while others investigated several voice-affecting disorders using ML techniques. In addition, PD seems to be the most investigated disorder for the majority of countries. The geographical trend described in this section reflects the country in which the study was performed and not the geographical source of the sample.

Figure 3. Usage of ML techniques and investigated disorders by year. AD: Alzheimer disease; ALS: amyotrophic lateral sclerosis; ANN: artificial neural network; CD: cardiovascular disease; CI: cognitive impairment; DT: decision tree; ET: essential tremor; GB: gradient boosting; GMM: Gaussian mixture model; KNN: K-nearest neighbor; LR: logic regression; MCI: mild cognitive impairment; MeML: mixed effect machine learning; ML: machine learning; MS: multiple sclerosis; NB: naïve Bayes; PA: passive active; PD: Parkinson disease; RF: random forest; SVM: support vector machine; SVR: support vector regression.

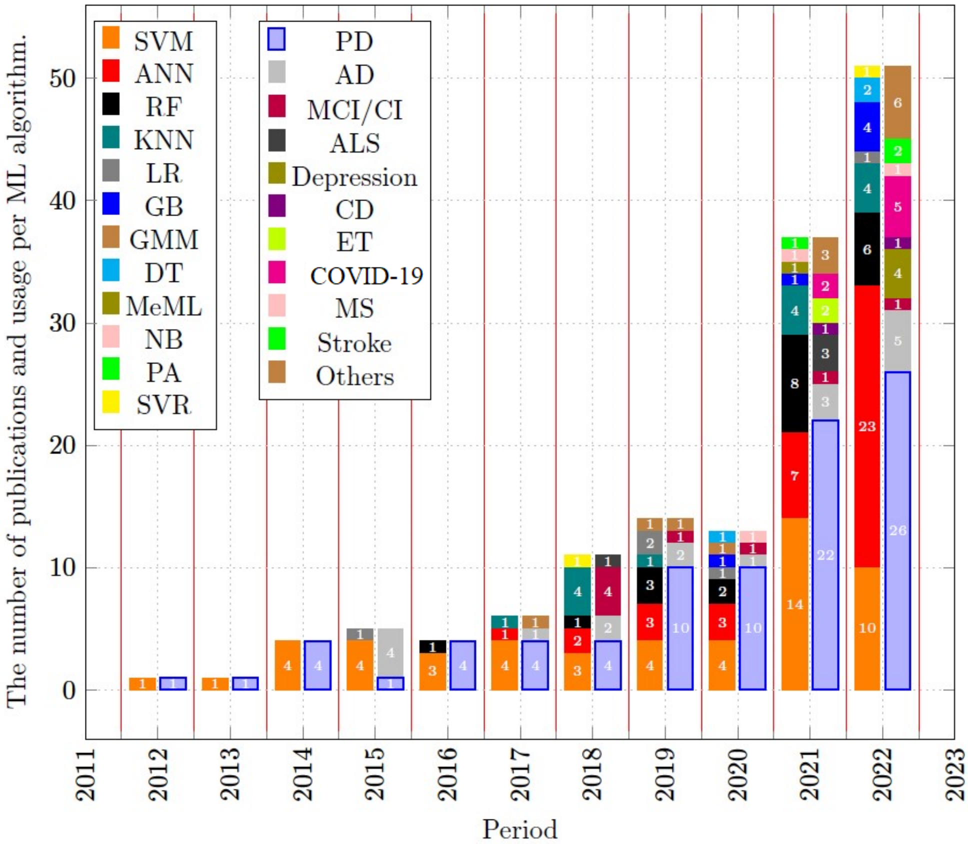
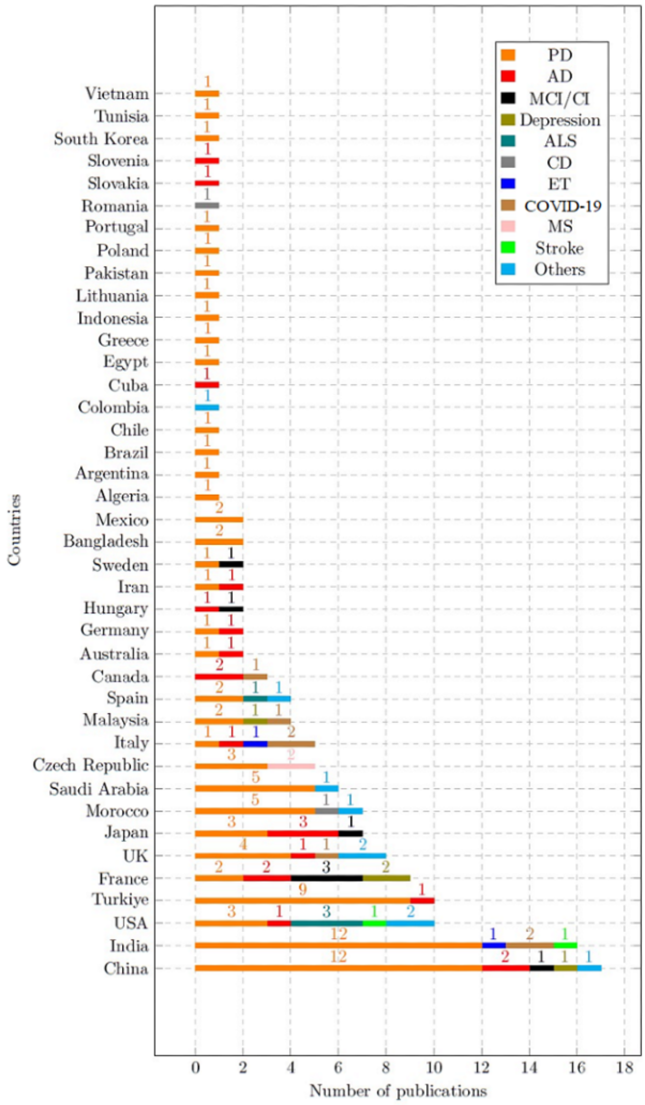


Figure 4. Investigated disorders by country. AD: Alzheimer disease; ALS: amyotrophic lateral sclerosis; CD: cardiovascular disorder; CI: cognitive impairment; ET: essential tremor; MCI: mild cognitive impairment; MS: multiple sclerosis; PD: Parkinson disease.



Data Characteristics

This section describes the characteristics of voice and nonvoice data used as input into the ML models. A total of 11/145 (7.6%) studies integrated nonvocal data in conjunction with vocal features to form the input feature sets for the ML models; 6 of these studies [54,66,96,105,149,163] incorporated demographic data, including gender, age, BMI, comorbidities, weight, height, and disease duration. Meanwhile, 2 of the studies [141,153] used video inputs, while 3 studies [83,158,159] incorporated external sensor signals such as electromyography and motion trackers.

Table 3 compiles the disorders and frequency of recorded data characteristics with the density of extracted vocal features and data source. Results indicate that vowel phonations are one of the most adopted recording types among almost all listed disorders. A total of 68 studies chose to base their analysis on vowel recordings, 33 combining different recordings, 20 free speeches, 12 scripted speeches, 9 picture descriptions, and 3 studies used syllable recordings. Cognitive disorders (eg, AD and MCI) that tend to use voice features extracted from speech and other disorders (eg, PD) lean toward features extracted from vowel phonation.

In all studies, raw data that consist of recordings underwent signal processing to extract features that were used as input data into ML techniques. Identified signal processing implementations were baseline acoustic (BLA), Mel-frequency cepstral coefficients, tunable Q-factor wavelet transform, wavelet transform, and spectrogram, which are frequency-transformed versions of the input signal that generates features in the form of digits and images. Other utilized features were linguistic and vocal features that generate statistical outputs (eg, silence rate, pause rate, duration, and ineligibility). In addition, combining several vocal features is more popular than only using BLA features, where almost 69/145 (47.6%) of the studies combined several features as input. But still, BLA features (36/145, 24.8%) are one of the most separately used feature sets. In this SLR BLA corresponds to all or a portion of acoustic, time, and frequency domain features calculated from raw recordings (eg, pitch, zero cross rate, jitter, shimmer, and formant frequencies).

Table 3. Characteristics of the input data and data source.

Disorder	Recording	Feature	Data source
Parkinson disease	• Vowel: 59 [17,40,42,44,48,50,54–59,61–64,68,69,74–85,87–91,93–97,99,100,102,119–122,124,126–128,130–132,135,170,183,184,192] • Combined: 20 [49,51,53,60,65,66,71,73,98,101,118,123,125,133,134,171,180,181,185,188] • Scripted speech: 3 [52,72,117] • Free speech: 3 [70,92,129] • Syllable: 2 [67,86]	• BLA[a]: 26 [17,40,50,53,56,58-60,64-66,71,73,77,79,82,84,87,88,92,98,118,132,133,184,188] • Combined: 45 [42,44,48,51,52,54,57,62,63,67–70,72,74,76,79,81,85,86,89–91,93,94,96,97,99,100,119,120,122,124,126–131,134,135,170,171,184,192] • MFCC[b]: 5 [49,61,80,94,101] • Spectrogram: 7 [75,117,123,125,180,181,185] • RP[c]: 1 [95] • TQWT[d]: 2 [55,121] • WT[e]: 1 [83]	• CFS[f]: 31 [48,51,52,65,67,70,72-74,78-81,83,85,86,88,92,94,96,97,101,102,117,130,131,170,171,180,185] • UCI[g]: 33 [40,42,44,50,53,55–57,60,61,63,64,68,69,71,82,84,91,93,99,100,119,120,122,124,126,127,129,133,135,183,188,192] • Multiple: 14 [49,54,58,59,66,75,87,90,98,118,123,125,132,181] • mPower: 4 [17,62,77,128] • NCVS[h]: 1 [88] • PARCZ[i]: 1 [89] • NG[j]: 3 [95,121,134]
Dementia, Alzheimer disease	• Free speech: 8 [103,105,108,172,173,175] • Picture description: 7 [109,110,136-139,174] • Scripted speech: 2 [104,176] • Combined: 1 [179]	• Combined: 11 [103,104,107,109,110,136,138,139,173-175] • Vocal: 2 [105,172] • BLA: 2 [176,179] • Spectogram: 2 [108,137] • Speech statistic: 1 [106]	• CFS: 10 [103-108,172,173,175,176] • ADBC[k]: 7 [109,110,136-139,174] • Multiple: 1 [179]
Cognitive impairment/mild cognitive impairment	• Free speech: 3 [112,114,115] • Picture description: 2 [116,178] • Scripted speech: 2 [113,177] • Combined [150]	• BLA: 3 [113,114,177] • Combined: 2 [112,116] • Linguistic: 1 [115] • Vocal: 1 [178] • Spectogram [150]	• CFS: 8 [112-116,150,177,178]
COVID-19	• Vowel: 2 [156,157] • Combined: 5 [143-147]	• Combined: 3 [144,156,157] • Spectogram: 3 [143,145,146] • MFCC [147]	• Coswara: 4 [143,144,156,157] • Multiple: 1 [146] • CFS: 1 [145] • CHRSD[l]: 1 [147]
Amyotrophic lateral sclerosis	• Vowel: 2 [46,160] • Scripted speech: 1 [159] • Combined: 1 [158]	• BLA: 2 [46,160] • Combined: 2 [158,159]	• CFS: 4 [46,158-160]
Depression	• Free speech: 2 [141,161] • Scripted speech [140] • Combined [142]	• Combined: 2 [142,161] • Spectrogram: 2 [140,141]	• Multiple: 2 [141,161] • CFS: 2 [140,142]
Cardiovascular disorders	• Vowel: 1 [111] • Combined: 1 [186]	• BLA: 1 [111] • MFCC: 1 [173	• CFS: 2 [111,186]
Essential tremor	• Vowel: 2 [162,163]	• BLA: 1 [162] • Spectrogram: 1 [163]	• CFS: 2 [162,163]
Multiple sclerosis	• Syllable [164] • Scripted speech [149]	• Spectrogram [164] • BLA [149]	• CFS: 2 [149,164]
Stroke	• Combined: 2 [148,153]	• MFCC [148] • Spectrogram [153]	• CFS: 2 [148,153]
Autism	• Free speech [151]	• Spectrogram [151]	• CFS: 1 [151]
Asthma	• Free speech [182]	• BLA [182]	• CFS: 1 [182]
Fatigue	• Scripted speech [187]	• Combined [187]	• CFS: 1 [187]

Disorder	Recording	Feature	Data source
Chronic obstructive pulmonary disease	• Scripted speech [154]	• BLA [154]	• CFS: 1 [154]
Neurodegenerative cognitive complaint	• Free speech [165]	• Combined [165]	• CFS: 1 [165]
Functional dysphagia/oropharyngeal dysphagia	• Combined [166]	• Combined [166]	• CFS: 1 [166]
Influenza disease	• Vowel [167]	• WT [167]	• CFS: 1 [167]
Neurological disease	• Vowel [169]	• Combined [169]	• CFS: 1 [169]
Traumatic brain injury	• Free speech [152]	• Spectrogram [152]	• TBIBank[m] Coelho corpus: 1 [152]

[a]BLA: baseline acoustic.
[b]MFCC: Mel-frequency cepstral coefficients.
[c]RP: recurrence plot.
[d]TQWT: tunable Q-factor wavelet transform.
[e]WT: wavelet.
[f]CFS: collected for study.
[g]UCI: University of California, Irvine.
[h]NCVS: National Center for Voice and Speech.
[i]PARCZ: Czech Parkinsonian Speech Database.
[j]NG: not given.
[k]ADBC: Alzheimer Dementia Bank blog corpus.
[l]CHRSD: Corona Hack Respiratory Sound data set.
[m]TBIBank: Traumatic Brain injury bank.

A total of 70 studies collected data for a specific study and 75 studies gathered data from an available data set. Sakar et al (2013) [73] and Sakar et al (2019) [48] are 2 different data sets donated to the UCI (University of California, Irvine), which have been used in 15 different included studies in this SLR; 5 studies [53,57,60,71,100] used the UCI data set containing 20 participants with PD and 20 HC participants, and 15 studies [40,42,44,48,55,61,64,68,99,119,121,122,124,126,127] used the UCI data set having 188 participants with PD and 64 HC participants from the same source. UCI and Coswara provide data sets that can be accessed and downloaded without any additional application [73,189]. All other data sources identified in this SLR require an application or are not publicly available. Data sets used in studies are unbalanced. Even if there is equality between the number of participants in terms of disordered and HC groups, a closer inspection of data sets reveals gender inequality. For example, Sakar et al (2013) [73] included 20 participants with PD and 20 HC participants; however, a closer inspection showed that the PD group comprised 6 females and 14 males, and the HC group consisted of 10 females and males, respectively. Another issue is the low number of participants in studies, where only 8/145 studies [17,62,77,78,81,92,113,170] based their outcome on more than 100 participants for both pathological and HC groups at the same time.

Observation Time

SQ5 aims to find out whether studies rely on longitudinal data and observation over time or observations at the same time that study was done. As the authors predefine the participants and measure the exposures and outcomes at the same time in all included studies, all studies in this SLR follow the cross-sectional study design [190].

Performance Evaluation

Measures presented to assess the efficiency of the ML techniques used show diversity in the included articles. Accuracy is one of the most used measures to present the outcome of almost all studies. Sensitivity, specificity, precision, Matthew's correlation coefficient, area under the curve, F_1-score, recall, mean absolute error, R^2, positive predictive value, and negative predictive value were other used measures in combination with accuracy without any standard order. Under the *performance* column in Multimedia Appendix 1, all combinations can be seen; 7 articles, 5 from the PD group [63,89,97,170,171], 1 from the AD group [172], and 1 from the ALS group [114], have presented results discriminated by gender and only 1 study [92] paid attention to language differences.

Two groups of studies [48,73] from UCI data sets were found to be suitable for meta-analysis due to the homogeneity between studies. The first group consisted of 15 studies using a data set containing voice recordings from 188 participants with PD and 64 HC participants [40,42,44,48,55,61,64,68,99,119,121,122,124,126,127]. The second group consisted of 5 studies [53,57,60,71,100] using voice recordings from 20 participants with PD and 20 HC participants (Table 4). Studies employing the first data set achieved 0.925 average accuracy within an accuracy range of 0.790-0.997. Studies employing the second data set achieved 0.869 average accuracy within an accuracy range of 0.670-0.990.

Table 4. List of comparable studies.

Data set[a]	Classifier	Feature	Performance	Reference
CFS[b] (donator)	SVM[c]	MFCC[d] and TQWT[e]	Accuracy: 0.8600	[48]
UCI[f]	GB[g]	BLA[h] and spectrum	Accuracy: 0.9388	[44]
UCI	KNN[i]	TQWT	Accuracy: 0.9800	[55]
UCI	ANN[j]	BLA	Accuracy: 0.9921	[40]
UCI	SVM	BLA, MFCC, WT[k], and TQWT	Accuracy: 0.9160	[42]
UCI	ANN	MFCC	Accuracy: 0.9674	[61]
UCI	NB[l]	BLA	Accuracy: 0.7897	[64]
UCI	SVM	BLA, MFCC, TQWT, and WT	Accuracy: 0.9470	[68]
UCI	SVM	BLA, MFCC, WT, and TQWT	Accuracy: 0.9350	[99]
UCI	SVM	BLA, MFCC, and TQWT	Accuracy: 0.8660	[119]
UCI	KNN	TQWT	Accuracy: 0.9890	[121]
UCI	RF[m]	BLA and MFCC	Accuracy: 0.8884	[122]
UCI	ANN	BLA, MFCC, and TQWT	Accuracy: 0.9200	[124]
UCI	SVM	BLA, MFCC, TQWT, and WT	Accuracy: 0.9621	[126]
UCI	ANN	BLA, MFCC, and TQWT	Accuracy: 0.9974	[127]
UCI	SVM	BLA	Accuracy: 0.6701	[53]
UCI	RF	BLA and MFCC	Accuracy: 0.9433	[57]
UCI	ANN	BLA	Accuracy: 0.9903	[60]
UCI	ANN	BLA	Accuracy: 0.8647	[72]
UCI	SVM	BLA and MFCC	Accuracy: 0.8750	[100]

[a]Italicized data sets represent Parkinson disease data set 2 containing data on patients with Parkinson disease (n=20) and HC (n=20); all other data sets correspond to Parkinson disease data set 1 containing data on patients with Parkinson disease (n=188) and HC (n=64).

[b]CFS: collected for study.

[c]SVM: support vector machine.

[d]MFCC: Mel-frequency cepstral coefficients.

[e]TQWT: tunable Q-factor wavelet transform.

[f]UCI: University of California, Irvine.

[g]GB: gradient boosting.

[h]BLA: baseline acoustic.

[i]KNN: K-nearest neighbor.

[j]ANN: artificial neural network.

[k]WT: wavelet.

[l]NB: naïve Bayes.

[m]RF: random forest.

Discussion

Principal Findings

In this SLR, 10 years of research on ML techniques applied for diagnosing and monitoring voice-affecting disorders indicates an extended interest from many countries. It seems that researchers have focused mostly on the detection of 19 identified disorders with low number of individuals in data sets that lead to gaps identified as the main findings of this SLR. These are summarized below:

- Most studies aimed to perform a diagnostic test through the detection or classification of disorders, and only a few studies aimed to monitor a specific disorder.
- PD was the most investigated disorder among all 19 voice-affecting disorders.
- There was a broad interest from many counties.
- Data sets used in studies were unbalanced, and most studies collected their data without providing open access. Additionally, only 11/145 (7.6%) included studies considered using additional data in conjunction with voice features.
- All studies were cross-sectional.
- Accuracy was the most common metric for the overall performance evaluation.

The majority of the studies focused on the detection or classification of the 19 identified voice-affecting disorders through emerging ML techniques. However, it is important to also consider the need for continuous monitoring of these disorders to improve the quality of life for those affected. Another consequence of focusing solely on detection is that it may not provide enough information about the severity of the disorder, which is a vital measure for decision-making on treatment or determining correct dosage for medication. Therefore, to improve the applicability of findings in clinical practices, it may be beneficial to navigate the focus of research toward methods for monitoring the progression, which involve severity measures of voice-affecting disorders.

Verdolini et al [6] give an intuition that the 19 disorders identified in this SLR correspond only to a small number of voice-affecting disorders that have been studied in research. This small correspondence makes it troublesome to highlight the digital biomarkers that are specifically related to a single disorder, which is essential for distinguishing underlying conditions that lead to altered voice quality. To address this issue, it is worth extending the research to other voice-affecting disorders that have been underrepresented in previous studies. This would not only extend the number of disorders being studied but also allow for the identification of differences and similarities in terms of digital biomarkers or other features across a wider range of disorders. Exploring the differences and similarities between disorders, syndromes, and symptoms is also beneficial because some disorders can function as symptoms of other underlying conditions affecting voice production, that is, while depression can be considered a disorder in and of itself, it can also manifest as a symptom of PD [6].

Based on the origin of the publication and the origin of the data sets, a wide interest from many countries was observed. However, many countries conduct research on the same data sets, which can lead to both positive and negative results regarding the clinical applicability of outcomes. Concentration on a group of data sets may increase the performance of the ML technique for the represented input data attributes. By contrast, it may also introduce limitations for the nonrepresented or underrepresented data. For example, the UCI data set in Sakar et al [73] contains several voice recordings in Turkish; using this data set may give satisfying results for recordings in the same language, but using it on English recordings could be problematic. However, interest from many countries shows enormous potential for collecting more available data sets and generalized ML techniques.

A balanced data set means the numbers of samples are relatively equal between classes, giving equivalent contributions from all classes during training, which eventually improves the performance of the ML technique on new data. By contrast, imbalanced data can lead to bias. The results of our SLR show that using balanced data has not been considered in studies. As the voice is used as a medium to detect a disorder, it is important to consider the effect of linguistic diversity, gender, age, and other sociodemographic differences on the generalizability of a system. Training and testing an ML technique on balanced data offer higher reliability for use in clinical practices. Balancing data based on different characteristics may be another option for higher reliability (eg, only male or only female). The studies included in the analysis provided demographic information about their respective data sets. However, only a limited number of studies incorporated this information into the vocal feature set that use additional nonvoice data for training the ML models. Integrating the demographic data into the automated process of data set preparation could prove beneficial, as opposed to the manual preparation of data sets based on disparate attributes. Additionally, combining multiple sensory inputs along with vocal features may further enhance the performance of the ML algorithms. However, this practice appears to be infrequently observed in recent studies.

Results of this SLR showed that 70/145 (48.3%) studies collected data specific to the research without making them publicly available. It is observable that PD is one of the most investigated disorders. That might be a result of publicly available data obtained from the UCI Parkinson data set repository. It is worthwhile to extend publicly available data sources with varied voice-affecting disorders and features to preserve research reliability and homogeneity in the scope. Another aspect that would influence clinical applicability is the small number of participants being considered in the research. Increasing the number of participants might increase the reliability of ML techniques.

In SLRs, “longitudinal study” refers to a recurrent sample taken from the same participant over time, which is a way of following the progression and trend of a disorder that helps to identify the patterns and causal relationships. Conducting a longitudinal study may even help to reduce the confounding variables [191].

The absence of longitudinal studies makes it difficult to conduct an epidemiological trend analysis of time effects on digital features extracted for ML techniques to diagnose and monitor voice-affecting disorders. All included studies in this SLR considered cross-sectional analysis, which does not represent the possible divergence tied to the progression of a specific disorder and individual. Therefore, longitudinal studies are essential to discover the voice changes over time [191,192].

The majority of the studies chose to represent the performance in accuracy, specificity, and sensitivity metrics, which were tied to the overall classification performance in research on voice-affecting disorder diagnosis and monitoring. In this SLR, only 8 results indicated that gender and language diversity may affect the performance in terms of accuracy [63,89,92,97,114,170-172]. As none of the studies address the effect of unbalanced data on performance evaluation, it is noteworthy that the divergence in accuracy in those outcomes could be the effect of unbalanced data. However, different accuracy results may be due to many other aspects. Regardless of the employed ML technique, used features; number of features; the proportion between training, validation, test sets; and feature extraction techniques may also be a factor in deviating accuracy results.

Limitations

The decision to include studies published only in English is a risk of missing important evidence in other languages, which at the same time is an unavoidable limitation for the generalization of this SLR. Another factor that can be considered as a limitation is only including peer-reviewed studies, which do not consider conference papers. Relying on the conference papers can be problematic due to the limitations including the potential for incomplete or preliminary results [193]. Additionally, including low-quality studies may introduce a risk of bias, as these studies may have suffered from selective reporting bias. In the SLR presented herein, this risk was mitigated by checking beyond what is presented in the paper, that is, when the methodological information was referenced elsewhere, the authors checked and considered the referenced material when conducting the quality assessment. Additionally, during the screening phase, when the abstract did not contain the full information to fulfill the inclusion criteria, these pieces were marked as “maybe” cases that were checked further before being fully read.

Future Work

Underrepresented monitoring purposes, research on a low number of voice-affecting disorders, unbalanced data, limited public voice data, lack of longitudinal research, and performance evaluation without paying attention to diversities were 6 gaps addressed in this SLR, which may be considered in future research. We suggest the following:

- One research direction may be to include disorders that were underrepresented in the state of the art. It is essential to take the gaps into consideration, such as working with balanced and extended data sets, to generate more reliable results.
- Conducting cross-sectional and longitudinal studies to identify specific digital features that are associated with voice-affecting disorders can be beneficial for determining the severity of the disorder and monitoring it over time.
- Studying the effects of demographic characteristics, such as gender, age, linguistic factors, and other relevant additional data on the classification models may also provide insights for building more accurate ML techniques for specific disorders.

Conclusions

Through the methodology of an SLR, we identified 145 studies on the use of voice for diagnosing or prognosing disorders, by the means of ML algorithms. These studies were summarized in terms of many aspects, including disorders and conditions that affect the voice, characteristics of the input data, ML techniques used for voice-based diagnosis, and research interests from countries. The findings of this SLR indicated that most of the studies are concerned with the detection and classification of investigated disorders and conditions based on cross-sectional studies. This study also found gaps in the literature, such as the usage of unbalanced data sets, lack of longitudinal studies, research not addressing nonvoice data in the voice studies, and most voice-affecting disorders in the interest of this study being underrepresented in research. Research in the field of voice-based diagnostics with the utilization of ML is making the practical application of this technology in health care more achievable. The use of voice as a digital biomarker could open the possibilities to large population screening of many disorders in a low-cost, noninvasive, and scalable way. To implement such a system in a clinical setting, the exploration of unknown aspects is an essential process to proceed with. To do that, it is necessary to extend the research on all possible voice-affecting disorders and identify the nuances between all different voice-affecting disorders and their effect on vocal features. Currently, research in this field primarily focuses on detection using a limited number of participants. However, for more generalizable results in the future, research may not only consider increasing the participant numbers but also maintaining a balance among them and identifying the measures that can be used for monitoring purposes.

There is a broad research interest from many countries, which creates a potential for observing the effects of cultural and language differences on ML algorithms. However, contribution to data collection and increasing the size of available data with diverse characteristics are crucial steps that each country might consider.

Acknowledgments

The authors thank the Excellence Center at Linköping – Lund in Information Technology (ELLIIT) for funding and supporting this project.

Data Availability

A summary table of all included studies and extracted data is available as Multimedia Appendix 1.

Authors' Contributions

AI is the primary contributor to the study and manuscript, with involvement in all aspects. ALD assisted in the design of the study, simultaneous study selection, and revisions of the manuscript. PA contributed to the final revisions of the manuscript. JSB provided medical expertise in the field and assisted with revisions of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Summary of included studies.
[PDF File (Adobe PDF File), 225 KB-Multimedia Appendix 1]

Multimedia Appendix 2

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist.
[DOCX File , 33 KB-Multimedia Appendix 2]

Multimedia Appendix 3

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 for Abstract checklist.
[DOCX File , 27 KB-Multimedia Appendix 3]

References

1. Zhang Z. Mechanics of human voice production and control. J Acoust Soc Am 2016 Oct;140(4):2614 [FREE Full text] [doi: 10.1121/1.4964509] [Medline: 27794319]
2. Merrill RM, Roy N, Lowe J. Voice-related symptoms and their effects on quality of life. Ann Otol Rhinol Laryngol 2013 Jun;122(6):404-411 [doi: 10.1177/000348941312200610] [Medline: 23837394]
3. Payten CL, Chiapello G, Weir KA, Madill CJ. Frameworks, Terminology and Definitions Used for the Classification of Voice Disorders: A Scoping Review. J Voice 2022 Mar 19:S0892-1997(22)00039-X [FREE Full text] [doi: 10.1016/j.jvoice.2022.02.009] [Medline: 35317970]
4. Payten C, Chiapello G, Weir K, Madill C. Terminology and frameworks used for the classification of voice disorders: a scoping review protocol. JBI Evid Synth 2021;19:454-462 [doi: 10.11124/jbies-20-00066]
5. Andrea M, Dias Ó, Andrea M, Figueira ML. Functional Voice Disorders: The Importance of the Psychologist in Clinical Voice Assessment. J Voice 2017 Jul;31(4):507.e13-507.e22 [doi: 10.1016/j.jvoice.2016.10.013] [Medline: 27876300]
6. Verdolini K, Rosen C, Branski R, editors. Classification Manual for Voice Disorders. New York, NY: Psychology Press; 2005.
7. Umeno H, Hyodo M, Haji T, Hara H, Imaizumi M, Ishige M, et al. A summary of the Clinical Practice Guideline for the Diagnosis and Management of Voice Disorders, 2018 in Japan. Auris Nasus Larynx 2020 Feb;47(1):7-17 [doi: 10.1016/j.anl.2019.09.004] [Medline: 31587820]
8. Cohen SM, Kim J, Roy N, Asche C, Courey M. Direct health care costs of laryngeal diseases and disorders. Laryngoscope 2012 Jul;122(7):1582-1588 [doi: 10.1002/lary.23189] [Medline: 22544473]
9. Califf RM. Biomarker definitions and their applications. Exp Biol Med (Maywood) 2018 Feb;243(3):213-221 [FREE Full text] [doi: 10.1177/1535370217750088] [Medline: 29405771]
10. Mayeux R. Biomarkers: potential uses and limitations. NeuroRx 2004 Apr;1(2):182-188 [FREE Full text] [doi: 10.1602/neurorx.1.2.182] [Medline: 15717018]
11. Babrak L, Menetski J, Rebhan M, Nisato G, Zinggeler M, Brasier N, et al. Traditional and Digital Biomarkers: Two Worlds Apart? Digit Biomark 2019;3(2):92-102 [FREE Full text] [doi: 10.1159/000502000] [Medline: 32095769]
12. Coravos A, Khozin S, Mandl KD. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. NPJ Digit Med 2019;2(1):14 [FREE Full text] [doi: 10.1038/s41746-019-0090-4] [Medline: 30868107]
13. Dorsey ER, Papapetropoulos S, Xiong M, Kieburtz K. The First Frontier: Digital Biomarkers for Neurodegenerative Disorders. Digit Biomark 2017;1(1):6-13 [FREE Full text] [doi: 10.1159/000477383] [Medline: 32095743]
14. Robin J, Harrison JE, Kaufman LD, Rudzicz F, Simpson W, Yancheva M. Evaluation of Speech-Based Digital Biomarkers: Review and Recommendations. Digit Biomark 2020;4(3):99-108 [FREE Full text] [doi: 10.1159/000510820] [Medline: 33251474]

15. Fagherazzi G, Fischer A, Ismael M, Despotovic V. Voice for Health: The Use of Vocal Biomarkers from Research to Clinical Practice. Digit Biomark 2021;5(1):78-88 [FREE Full text] [doi: 10.1159/000515346] [Medline: 34056518]
16. Lin H, Karjadi C, Ang T, Prajakta J, McManus C, Alhanai T, et al. Identification of digital voice biomarkers for cognitive health. Explor Med 2020;1:406-417 [FREE Full text] [doi: 10.37349/emed.2020.00028] [Medline: 33665648]
17. Tracy JM, Özkanca Y, Atkins DC, Hosseini Ghomi R. Investigating voice as a biomarker: Deep phenotyping methods for early detection of Parkinson's disease. J Biomed Inform 2020 Apr;104:103362 [FREE Full text] [doi: 10.1016/j.jbi.2019.103362] [Medline: 31866434]
18. López-de-Ipiña K, Alonso JB, Travieso CM, Solé-Casals J, Egiraun H, Faundez-Zanuy M, et al. On the selection of non-invasive methods based on speech analysis oriented to automatic Alzheimer disease diagnosis. Sensors (Basel) 2013 May 21;13(5):6730-6745 [FREE Full text] [doi: 10.3390/s130506730] [Medline: 23698268]
19. Van Stan J, Mehta DD, Hillman RE. Recent Innovations in Voice Assessment Expected to Impact the Clinical Management of Voice Disorders. Perspect ASHA SIGs 2017 Jan;2(3):4-13 [doi: 10.1044/persp2.SIG3.4]
20. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med 2019 Jan;25(1):24-29 [doi: 10.1038/s41591-018-0316-z] [Medline: 30617335]
21. Shailaja K, Seetharamulu B, Jabbar M. Machine Learning in Healthcare: A Review. 2018 Presented at: Second International Conference on Electronics, Communication and Aerospace Technology (ICECA); March 29-31, 2018; Coimbatore, Tamil Nadu, India p. 910-914 [doi: 10.1109/ICECA.2018.8474918]
22. Garg A, Sharma P. Survey on acoustic modeling and feature extraction for speech recognition. 2016 Presented at: 3rd International Conference on Computing for Sustainable Global Development (INDIACom); March 16-18, 2016; New Delhi, India p. 2291-2295
23. Department OEE. G. Balekundri Institute of Technology, Belagavi-India, Kanabur V, Harakannanavar SS, Torse D. An Extensive Review of Feature Extraction Techniques, Challenges and Trends in Automatic Speech Recognition. Int J Image Graph Signal Process. 2019;11:12 [doi: 10.5815/ijigsp.2019.05.01]
24. Abdallah B, Abdallah A, Ratte S. Detecting depression in Alzheimer and MCI using artificial neural networks (ANN). 2021 Apr Presented at: International Conference on Data Science, E-learning and Information Systems 2021; April 5-7, 2021; Ma'an, Jordan p. 250-253 [doi: 10.1145/3460620.3460765]
25. Kurzekar PK, Deshmukh RR, Waghmare VB, Shrishrimal PP. A Comparative Study of Feature Extraction Techniques for Speech Recognition System. International Journal of Innovative Research in Science, Engineering and Technology 2014 Dec;3(12):18006-18016 [FREE Full text] [doi: 10.15680/IJIRSET.2014.0312034]
26. Syed SA, Rashid M, Hussain S. Meta-analysis of voice disorders databases and applied machine learning techniques. Math Biosci Eng 2020 Nov 11;17(6):7958-7979 [FREE Full text] [doi: 10.3934/mbe.2020404] [Medline: 33378928]
27. Pahuja G, Nagabhushan TN. A Comparative Study of Existing Machine Learning Approaches for Parkinson's Disease Detection. IETE Journal of Research 2018 Oct 22;67(1):4-14 [doi: 10.1080/03772063.2018.1531730]
28. Petti U, Baker S, Korhonen A. A systematic literature review of automatic Alzheimer's disease detection from speech and language. J Am Med Inform Assoc 2020 Nov 01;27(11):1784-1797 [FREE Full text] [doi: 10.1093/jamia/ocaa174] [Medline: 32929494]
29. Saravanan S, Ramkumar K, Adalarasu K, Sivanandam V, Kumar SR, Stalin S, et al. A Systematic Review of Artificial Intelligence (AI) Based Approaches for the Diagnosis of Parkinson’s Disease. Arch Computat Methods Eng 2022 Jan 20;29(6):3639-3653 [doi: 10.1007/s11831-022-09710-1]
30. Bettany-Saltikov J. How to do a Systematic Literature Review in Nursing: A Step-by-Step Guide. London, UK: McGraw-Hill Education; 2016.
31. Idrisoglu A. Protocol of the Systematic Literature Review. GitHub. 2023 Jan 28. URL: https://github.com/AIITPlanet/Protocol [accessed 2023-07-11]
32. Considine J, Shaban RZ, Fry M, Curtis K. Evidence based emergency nursing: Designing a research question and searching the literature. Int Emerg Nurs 2017 May;32:78-82 [doi: 10.1016/j.ienj.2017.02.001] [Medline: 28233626]
33. Butler A, Hall H, Copnell B. A Guide to Writing a Qualitative Systematic Review Protocol to Enhance Evidence-Based Practice in Nursing and Health Care. Worldviews Evid Based Nurs 2016 Jun 20;13(3):241-249 [doi: 10.1111/wvn.12134] [Medline: 26790142]
34. Schiavenato M, Chu F. PICO: What it is and what it is not. Nurse Educ Pract 2021 Oct;56:103194 [doi: 10.1016/j.nepr.2021.103194] [Medline: 34534728]
35. Mueen Ahmed KK, Dhubaib BEA. Zotero: A bibliographic assistant to researcher. Journal of Pharmacology and Pharmacotherapeutics 2022 Apr 11;2(4):304-305 [doi: 10.4103/0976-500x.85940]
36. Kitchenham B, Charters S. Guidelines for Performing Systematic Literature Reviews in Software Engineering (Technical Report EBSE-2007-01). University of Auckland. 2007. URL: https://www.cs.auckland.ac.nz/~norsaremah/2007%20Guidelines%20for%20performing%20SLR%20in%20SE%20v2.3.pdf [accessed 2023-07-11]
37. Pianosi F, Beven K, Freer J, Hall JW, Rougier J, Stephenson DB, et al. Sensitivity analysis of environmental models: A systematic review with practical workflow. Environmental Modelling & Software 2016 May;79:214-232 [doi: 10.1016/j.envsoft.2016.02.008]

38. Christopher Frey H, Patil SR. Identification and Review of Sensitivity Analysis Methods. Risk Analysis 2002 Jun;22(3):553-578 [doi: 10.1111/0272-4332.00039]
39. PRISMA 2020 Checklist. PRISMA. 2020. URL: http://www.prisma-statement.org/PRISMAStatement/Checklist [accessed 2023-03-28]
40. Tuncer T, Dogan S. A novel octopus based Parkinson’s disease and gender recognition method using vowels. Applied Acoustics 2019 Dec;155:75-83 [doi: 10.1016/j.apacoust.2019.05.019]
41. Tuncer T, Dogan S, Acharya UR. Automated detection of Parkinson's disease using minimum average maximum tree and singular value decomposition method with vowels. Biocybernetics and Biomedical Engineering 2020 Jan;40(1):211-220 [doi: 10.1016/j.bbe.2019.05.006]
42. Gunduz H. Deep Learning-Based Parkinson’s Disease Classification Using Vocal Feature Sets. IEEE Access 2019;7:115540-115551 [doi: 10.1109/access.2019.2936564]
43. Gunduz H. An efficient dimensionality reduction method using filter-based feature selection and variational autoencoders on Parkinson's disease classification. Biomedical Signal Processing and Control 2021 Apr;66:102452 [doi: 10.1016/j.bspc.2021.102452]
44. Lamba R, Gulati T, Jain A. A Hybrid Feature Selection Approach for Parkinson’s Detection Based on Mutual Information Gain and Recursive Feature Elimination. Arab J Sci Eng 2022 Jan 18;47(8):10263-10276 [doi: 10.1007/s13369-021-06544-0]
45. Lamba R, Gulati T, Jain A. An Intelligent System for Parkinson's Diagnosis Using Hybrid Feature Selection Approach. Int J Softw Innov 2022;10(1):1-13 [doi: 10.4018/ijsi.292027]
46. Tena A, Claria F, Solsona F, Meister E, Povedano M. Detection of Bulbar Involvement in Patients With Amyotrophic Lateral Sclerosis by Machine Learning Voice Analysis: Diagnostic Decision Support Development Study. JMIR Med Inform 2021 Mar 10;9(3):e21331 [FREE Full text] [doi: 10.2196/21331] [Medline: 33688838]
47. Tena A, Clarià F, Solsona F, Povedano M. Detecting Bulbar Involvement in Patients with Amyotrophic Lateral Sclerosis Based on Phonatory and Time-Frequency Features. Sensors (Basel) 2022 Feb 02;22(3):1137 [FREE Full text] [doi: 10.3390/s22031137] [Medline: 35161881]
48. Sakar CO, Serbes G, Gunduz A, Tunc HC, Nizam H, Sakar BE, et al. A comparative analysis of speech signal processing algorithms for Parkinson’s disease classification and the use of the tunable Q-factor wavelet transform. Applied Soft Computing 2019 Jan;74:255-263 [doi: 10.1016/j.asoc.2018.10.022]
49. Moro-Velazquez L, Gomez-Garcia JA, Godino-Llorente JI, Villalba J, Rusz J, Shattuck-Hufnagel S, et al. A forced gaussians based methodology for the differential evaluation of Parkinson's Disease by means of speech processing. Biomedical Signal Processing and Control 2019 Feb;48:205-220 [doi: 10.1016/j.bspc.2018.10.020]
50. Meghraoui D, Boudraa B, Merazi-Meksen T, Gómez Vilda P. A novel pre-processing technique in pathologic voice detection: Application to Parkinson’s disease phonation. Biomedical Signal Processing and Control 2021 Jul;68:102604 [doi: 10.1016/j.bspc.2021.102604]
51. Quan C, Ren K, Luo Z. A Deep Learning Based Method for Parkinson’s Disease Detection Using Dynamic Features of Speech. IEEE Access 2021;9:10239-10252 [doi: 10.1109/ACCESS.2021.3051432]
52. Goyal J, Khandnor P, Aseri TC. A Hybrid Approach for Parkinson’s Disease diagnosis with Resonance and Time-Frequency based features from Speech signals. Expert Systems with Applications 2021 Nov;182:115283 [doi: 10.1016/j.eswa.2021.115283]
53. Cantürk I, Karabiber F. A Machine Learning System for the Diagnosis of Parkinson's Disease from Speech Signals and Its Application to Multiple Speech Signal Types. Arab J Sci Eng 2016;41:5059 [doi: 10.1007/s13369-016-2206-3]
54. Carrón J, Campos-Roca Y, Madruga M, Pérez CJ. A mobile-assisted voice condition analysis system for Parkinson's disease: assessment of usability conditions. Biomed Eng Online 2021 Nov 21;20(1):114 [FREE Full text] [doi: 10.1186/s12938-021-00951-y] [Medline: 34802448]
55. Yücelbaş C. A new approach: information gain algorithm-based k-nearest neighbors hybrid diagnostic system for Parkinson's disease. Phys Eng Sci Med 2021 Jun;44(2):511-524 [doi: 10.1007/s13246-021-01001-6] [Medline: 33852120]
56. Cai Z, Gu J, Chen H. A New Hybrid Intelligent Framework for Predicting Parkinson’s Disease. IEEE Access 2017;5:17188-17200 [doi: 10.1109/ACCESS.2017.2741521]
57. Jahnavi BS, Supraja BS, Lalitha S. A vital neurodegenerative disorder detection using speech cues. IFS 2020 May 29;38(5):6337-6345 [doi: 10.3233/JIFS-179714]
58. Zhang L, Qu Y, Jin B, Jing L, Gao Z, Liang Z. An Intelligent Mobile-Enabled System for Diagnosing Parkinson Disease: Development and Validation of a Speech Impairment Detection System. JMIR Med Inform 2020 Sep 16;8(9):e18689 [FREE Full text] [doi: 10.2196/18689] [Medline: 32936086]
59. Cai Z, Gu J, Wen C, Zhao D, Huang C, Huang H, et al. An Intelligent Parkinson's Disease Diagnostic System Based on a Chaotic Bacterial Foraging Optimization Enhanced Fuzzy KNN Approach. Comput Math Methods Med 2018;2018:2396952 [FREE Full text] [doi: 10.1155/2018/2396952] [Medline: 30034509]
60. Rizvi DR, Nissar I, Masood S, Ahmed M, Ahmad F. An LSTM based deep learning model for voice-based detection of Parkinson's disease. Int J Adv Sci Technol 2020;29(5s):337-343 [FREE Full text]
61. Olivares R, Munoz R, Soto R, Crawford B, Cárdenas D, Ponce A, et al. An Optimized Brain-Based Algorithm for Classifying Parkinson’s Disease. Applied Sciences 2020 Mar 06;10(5):1827 [doi: 10.3390/app10051827]

62. Tougui I, Jilbab A, Mhamdi JE. Analysis of Smartphone Recordings in Time, Frequency, and Cepstral Domains to Classify Parkinson’s Disease. Healthc Inform Res 2020 Oct;26(4):274-283 [FREE Full text] [doi: 10.4258/hir.2020.26.4.274] [Medline: 33190461]
63. Solana-Lavalle G, Rosas-Romero R. Analysis of voice as an assisting tool for detection of Parkinson's disease and its subsequent clinical interpretation. Biomedical Signal Processing and Control 2021 Apr;66:102415 [doi: 10.1016/j.bspc.2021.102415]
64. Pramanik M, Pradhan R, Nandy P, Qaisar SM, Bhoi AK. Assessment of Acoustic Features and Machine Learning for Parkinson's Detection. J Healthc Eng 2021;2021:9957132 [FREE Full text] [doi: 10.1155/2021/9957132] [Medline: 34471507]
65. Ali L, Zhu C, Zhang Z, Liu Y. Automated Detection of Parkinson's Disease Based on Multiple Types of Sustained Phonations Using Linear Discriminant Analysis and Genetically Optimized Neural Network. IEEE J Transl Eng Health Med 2019;7:2000410 [FREE Full text] [doi: 10.1109/JTEHM.2019.2940900] [Medline: 32166050]
66. Braga D, Madureira AM, Coelho L, Ajith R. Automatic detection of Parkinson’s disease based on acoustic analysis of speech. Engineering Applications of Artificial Intelligence 2019 Jan;77:148-158 [doi: 10.1016/j.engappai.2018.09.018]
67. Novotny M, Rusz J, Cmejla R, Ruzicka E. Automatic Evaluation of Articulatory Disorders in Parkinson’s Disease. IEEE/ACM Trans. Audio Speech Lang. Process 2014 Sep;22(9):1366-1378 [doi: 10.1109/TASLP.2014.2329734]
68. Solana-Lavalle G, Galán-Hernández JC, Rosas-Romero R. Automatic Parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features. Biocybernetics and Biomedical Engineering 2020 Jan;40(1):505-516 [doi: 10.1016/j.bbe.2020.01.003]
69. Zhang HH, Yang L, Liu Y, Wang P, Yin J, Li Y, et al. Classification of Parkinson's disease utilizing multi-edit nearest-neighbor and ensemble learning algorithms with speech samples. Biomed Eng Online 2016 Nov 16;15(1):122 [FREE Full text] [doi: 10.1186/s12938-016-0242-6] [Medline: 27852279]
70. Khan T, Westin J, Dougherty M. Classification of speech intelligibility in Parkinson's disease. Biocybernetics and Biomedical Engineering 2014;34(1):35-45 [doi: 10.1016/j.bbe.2013.10.003]
71. Berus L, Klancnik S, Brezocnik M, Ficko M. Classifying Parkinson's Disease Based on Acoustic Measures Using Artificial Neural Networks. Sensors (Basel) 2018 Dec 20;19(1):16 [FREE Full text] [doi: 10.3390/s19010016] [Medline: 30577548]
72. García AM, Arias-Vergara TC, C Vasquez-Correa J, Nöth E, Schuster M, Welch AE, et al. Cognitive Determinants of Dysarthria in Parkinson's Disease: An Automated Machine Learning Approach. Mov Disord 2021 Dec;36(12):2862-2873 [doi: 10.1002/mds.28751] [Medline: 34390508]
73. Sakar BE, Isenkul M, Sakar CO, Sertbas A, Gurgen F, Delil S, et al. Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE J Biomed Health Inform 2013 Jul;17(4):828-834 [doi: 10.1109/JBHI.2013.2245674] [Medline: 25055311]
74. Viswanathan R, Arjunan SP, Bingham A, Jelfs B, Kempster P, Raghav S, et al. Complexity Measures of Voice Recordings as a Discriminative Tool for Parkinson's Disease. Biosensors (Basel) 2019 Dec 20;10(1):10 [FREE Full text] [doi: 10.3390/bios10010001] [Medline: 31861890]
75. Hireš M, Gazda M, Drotár P, Pah ND, Motin MA, Kumar DK. Convolutional neural network ensemble for Parkinson's disease detection from voice recordings. Comput Biol Med 2022 Feb;141:105021 [doi: 10.1016/j.compbiomed.2021.105021] [Medline: 34799077]
76. Majda-Zdancewicz E, Potulska-Chromik A, Jakubowski J, Nojszewska M, Kostera-Pruszczyk A. Deep learning vs feature engineering in the assessment of voice signals for diagnosis in Parkinson's disease. Bull Pol Acad Sci Tech Sci 2021;69(3):e137347 [FREE Full text] [doi: 10.24425/bpasts.2021.137347]
77. Ozkanca Y, Öztürk MG, Ekmekci M, Atkins D, Demiroglu C, Ghomi RH. Depression Screening from Voice Samples of Patients Affected by Parkinson's Disease. Digit Biomark 2019 Jun 12;3(2):72-82 [FREE Full text] [doi: 10.1159/000500354] [Medline: 31872172]
78. Rahman W, Lee S, Islam MS, Antony VN, Ratnu H, Ali MR, et al. Detecting Parkinson Disease Using a Web-Based Speech Task: Observational Study. J Med Internet Res 2021 Oct 19;23(10):e26305 [FREE Full text] [doi: 10.2196/26305] [Medline: 34665148]
79. Almeida JS, Rebouças Filho P, Carneiro T, Wei W, Damaševičius R, Maskeliūnas R, et al. Detecting Parkinson’s disease with sustained phonation and speech signals using machine learning techniques. Pattern Recognition Letters 2019 Jul;125:55-62 [doi: 10.1016/j.patrec.2019.04.005]
80. Benba A, Jilbab A, Hammouch A. Detecting multiple system atrophy, Parkinson and other neurological disorders using voice analysis. Int J Speech Technol 2017 Mar 4;20(2):281-288 [doi: 10.1007/s10772-017-9404-6]
81. Arora S, Baghai-Ravary L, Tsanas A. Developing a large scale population screening tool for the assessment of Parkinson's disease using telephone-quality voice. J Acoust Soc Am 2019 May;145(5):2871 [FREE Full text] [doi: 10.1121/1.5100272] [Medline: 31153319]
82. Yang S, Zheng F, Luo X, Cai S, Wu Y, Liu K, et al. Effective dysphonia detection using feature dimension reduction and kernel density estimation for patients with Parkinson's disease. PLoS One 2014;9(2):e88825 [FREE Full text] [doi: 10.1371/journal.pone.0088825] [Medline: 24586406]

83. Oung QW, Muthusamy H, Basah SN, Lee H, Vijean V. Empirical Wavelet Transform Based Features for Classification of Parkinson's Disease Severity. J Med Syst 2017 Dec 29;42(2):29 [doi: 10.1007/s10916-017-0877-2] [Medline: 29288342]
84. Haq AU, Li JP, Memon MH, khan J, Malik A, Ahmad T, et al. Feature Selection Based on L1-Norm Support Vector Machine and Effective Recognition System for Parkinson’s Disease Using Voice Recordings. IEEE Access 2019;7:37718-37734 [doi: 10.1109/ACCESS.2019.2906350]
85. Rusz J, Novotny M, Hlavnicka J, Tykalova T, Ruzicka E. High-accuracy voice-based classification between patients with Parkinson's disease and other neurological diseases may be an easy task with inappropriate experimental design. IEEE Trans Neural Syst Rehabil Eng 2017 Aug;25(8):1319-1321 [doi: 10.1109/TNSRE.2016.2621885] [Medline: 28113773]
86. Klempír O, Krupicka R. Machine learning using speech utterances for Parkinson disease detection. Lekar a Technika 2018 Jan;48(2):66-71 [FREE Full text]
87. Alhussein M. Monitoring Parkinson’s Disease in Smart Cities. IEEE Access 2017;5:19835-19841 [doi: 10.1109/ACCESS.2017.2748561]
88. Tsanas A, Little MA, McSharry PE, Spielman J, Ramig LO. Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. IEEE Trans Biomed Eng 2012 May;59(5):1264-1271 [doi: 10.1109/TBME.2012.2183367] [Medline: 22249592]
89. Gómez-Vilda P, Mekyska J, Ferrández JM, Palacios-Alonso D, Gómez-Rodellar A, Rodellar-Biarge V, et al. Parkinson Disease Detection from Speech Articulation Neuromechanics. Front Neuroinform 2017;11:56 [FREE Full text] [doi: 10.3389/fninf.2017.00056] [Medline: 28970792]
90. Zhang T, Zhang Y, Sun H, Shan H. Parkinson disease detection using energy direction features based on EMD from voice signal. Biocybernetics and Biomedical Engineering 2021 Jan;41(1):127-141 [doi: 10.1016/j.bbe.2020.12.009]
91. Karan B, Sahu SS, Mahto K. Parkinson disease prediction using intrinsic mode function based features from speech signal. Biocybernetics and Biomedical Engineering 2020 Jan;40(1):249-264 [doi: 10.1016/j.bbe.2019.05.005]
92. Laganas C, Iakovakis D, Hadjidimitriou SK, Charisis V, Dias SB, Bostantzopoulou S, et al. Parkinson's Disease Detection Based on Running Speech Data From Phone Calls. IEEE Trans Biomed Eng 2022 May;69(5):1573-1584 [doi: 10.1109/TBME.2021.3116935] [Medline: 34596531]
93. Bchir O. Parkinson’s Disease Classification using Gaussian Mixture Models with Relevance Feature Weights on Vocal Feature Sets. Int J Adv Comput Sci Appl 2020;11(4):413-419 [FREE Full text] [doi: 10.14569/IJACSA.2020.0110456]
94. Rahman A, Rizvi SS, Khan A, Abbasi AA, Khan SU, Chung TS. Parkinson’s Disease Diagnosis in Cepstral Domain Using MFCC and Dimensionality Reduction with SVM Classifier. Mobile Information Systems 2021;2021:1-10 [FREE Full text] [doi: 10.1155/2021/8822069]
95. Fujita T, Luo Z, Quan C, Mori K, Cao S. Performance Evaluation of RNN with Hyperbolic Secant in Gate Structure through Application of Parkinson’s Disease Detection. Applied Sciences 2021 May 11;11(10):4361 [doi: 10.3390/app11104361]
96. Vital TPR, Nayak J, Naik B, Jayaram D. Probabilistic Neural Network-based Model for Identification of Parkinson’s Disease by using Voice Profile and Personal Data. Arab J Sci Eng 2021 Jan 03;46(4):3383-3407 [doi: 10.1007/s13369-020-05080-7]
97. Azadi H, Akbarzadeh-T M, Kobravi HR, Shoeibi A. Robust Voice Feature Selection Using Interval Type-2 Fuzzy AHP for Automated Diagnosis of Parkinson's Disease. IEEE/ACM Trans. Audio Speech Lang. Process 2021;29:2792-2802 [doi: 10.1109/TASLP.2021.3097215]
98. Amato F, Borzi L, Olmo G, Artusi CA, Imbalzano G, Lopiano L. Speech Impairment in Parkinson’s Disease: Acoustic Analysis of Unvoiced Consonants in Italian Native Speakers. IEEE Access 2021;9:166370-166381 [doi: 10.1109/ACCESS.2021.3135626]
99. Hoq M, Uddin MN, Park SB. Vocal Feature Extraction-Based Artificial Intelligent Model for Parkinson's Disease Detection. Diagnostics (Basel) 2021 Jun 11;11(6):11 [FREE Full text] [doi: 10.3390/diagnostics11061076] [Medline: 34208330]
100. Benba A, Jilbab A, Hammouch A. Voice assessments for detecting patients with Parkinson’s diseases using PCA and NPCA. Int J Speech Technol 2016 Sep 3;19(4):743-754 [doi: 10.1007/s10772-016-9367-z]
101. Jeancolas L, Petrovska-Delacrétaz D, Mangone G, Benkelfat BE, Corvol JC, Vidailhet M, et al. X-Vectors: New Quantitative Biomarkers for Early Parkinson's Disease Detection From Speech. Front Neuroinform 2021;15:578369 [FREE Full text] [doi: 10.3389/fninf.2021.578369] [Medline: 33679361]
102. Saloni S, Sharma R, Gupta A. Human voice waveform analysis for categorization of healthy and Parkinson subjects. International Journal of Healthcare Information Systems and Informatics 2016;11:21-35 [doi: 10.4018/ijhisi.2016010102]
103. Nasreen S, Rohanian M, Hough J, Purver M. Alzheimer’s Dementia Recognition From Spontaneous Speech Using Disfluency and Interactional Features. Front. Comput. Sci 2021 Jun 18;3:3 [doi: 10.3389/fcomp.2021.640669]
104. Gonzalez-Moreira E, Torres-Boza D, Kairuz HA, Ferrer C, Garcia-Zamora M, Espinoza-Cuadros F, et al. Automatic Prosodic Analysis to Identify Mild Dementia. Biomed Res Int 2015;2015:916356 [FREE Full text] [doi: 10.1155/2015/916356] [Medline: 26558287]
105. Shimoda A, Li Y, Hayashi H, Kondo N. Dementia risks identified by vocal features via telephone conversations: A novel machine learning prediction model. PLoS One 2021;16(7):e0253988 [FREE Full text] [doi: 10.1371/journal.pone.0253988] [Medline: 34260593]

106. Tanaka H, Adachi H, Ukita N, Ikeda M, Kazui H, Kudo T, et al. Detecting Dementia Through Interactive Computer Avatars. IEEE J Transl Eng Health Med 2017;5:2200111 [FREE Full text] [doi: 10.1109/JTEHM.2017.2752152] [Medline: 29018636]
107. Khodabakhsh A, Yesil F, Guner E, Demiroglu C. Evaluation of linguistic and prosodic features for detection of Alzheimer's disease in Turkish conversational speech. J AUDIO SPEECH MUSIC PROC 2015 Mar 25;2015(1):9 [doi: 10.1186/s13636-015-0052-y]
108. Nasrolahzadeh M, Mohammadpoory Z, Haddadnia J. Higher-order spectral analysis of spontaneous speech signals in Alzheimer's disease. Cogn Neurodyn 2018 Dec;12(6):583-596 [FREE Full text] [doi: 10.1007/s11571-018-9499-8] [Medline: 30483366]
109. Fraser KC, Meltzer JA, Rudzicz F. Linguistic Features Identify Alzheimer's Disease in Narrative Speech. J Alzheimers Dis 2016;49(2):407-422 [doi: 10.3233/JAD-150520] [Medline: 26484921]
110. Guo Z, Ling Z, Li Y. Detecting Alzheimer's Disease from Continuous Speech Using Language Models. J Alzheimers Dis 2019;70(4):1163-1174 [doi: 10.3233/JAD-190452] [Medline: 31322577]
111. Bourouhou A, Jilbab A, Nacir C, Hammouch A. Classification of cardiovascular disease using dysphonia measurement in speech. Diagnostyka 2021;22(1):31-37 [FREE Full text] [doi: 10.29354/diag/132586]
112. Toth L, Hoffmann I, Gosztolya G, Vincze V, Szatloczki G, Banreti Z, et al. A Speech Recognition-based Solution for the Automatic Detection of Mild Cognitive Impairment from Spontaneous Speech. Curr Alzheimer Res 2018;15(2):130-138 [FREE Full text] [doi: 10.2174/1567205014666171121114930] [Medline: 29165085]
113. Nagumo R, Zhang Y, Ogawa Y, Hosokawa M, Abe K, Ukeda T, et al. Automatic Detection of Cognitive Impairments through Acoustic Analysis of Speech. CAR 2020 Mar 20;17(1):60-68 [FREE Full text] [doi: 10.2174/1567205017666200213094513] [Medline: 32053074]
114. König A, Linz N, Zeghari R, Klinge X, Tröger J, Alexandersson J, et al. Detecting Apathy in Older Adults with Cognitive Disorders Using Automatic Speech Analysis. J Alzheimers Dis 2019;69(4):1183-1193 [doi: 10.3233/JAD-181033] [Medline: 31127764]
115. König A, Linz N, Tröger J, Wolters M, Alexandersson J, Robert P. Fully Automatic Speech-Based Analysis of the Semantic Verbal Fluency Task. Dement Geriatr Cogn Disord 2018;45(3-4):198-209 [doi: 10.1159/000487852] [Medline: 29886493]
116. Wang T, Hong Y, Wang Q, Su R, Ng ML, Xu J, et al. Identification of Mild Cognitive Impairment Among Chinese Based on Multiple Spoken Tasks. JAD 2021 Jun 29;82(1):185-204 [doi: 10.3233/jad-201387]
117. Maskeliūnas R, Damaševičius R, Kulikajevas A, Padervinskis E, Pribuišis K, Uloza V. A Hybrid U-Lossian Deep Learning Network for Screening and Evaluating Parkinson's Disease. Applied Sciences 2022 Nov 15;12(22):11601 [doi: 10.3390/app122211601]
118. Lamba R, Gulati T, Jain A, Rani P. A Speech-Based Hybrid Decision Support System for Early Detection of Parkinson's Disease. Arab J Sci Eng 2022 Sep 12;48(2):2247-2260 [doi: 10.1007/s13369-022-07249-8]
119. Dao SVT, Yu Z, Tran LV, Phan PNK, Huynh TTM, Le TM. An Analysis of Vocal Features for Parkinson's Disease Classification Using Evolutionary Algorithms. Diagnostics (Basel) 2022 Aug 16;12(8):12 [FREE Full text] [doi: 10.3390/diagnostics12081980] [Medline: 36010330]
120. Barukab O, Ahmad A, Khan T, Thayyil Kunhumuhammed MR. Analysis of Parkinson's Disease Using an Imbalanced-Speech Dataset by Employing Decision Tree Ensemble Methods. Diagnostics (Basel) 2022 Nov 30;12(12):12 [FREE Full text] [doi: 10.3390/diagnostics12123000] [Medline: 36553007]
121. Kaya D. Automated gender ‐ Parkinson's disease detection at the same time via a hybrid deep model using human voice. Concurrency and Computation 2022 Aug 23;34(26):34 [doi: 10.1002/cpe.7289]
122. Hawi S, Alhozami J, AlQahtani R, AlSafran D, Alqarni M, Sahmarany L. Automatic Parkinson's disease detection based on the combination of long-term acoustic features and Mel frequency cepstral coefficients (MFCC). Biomedical Signal Processing and Control 2022 Sep;78:104013 [doi: 10.1016/j.bspc.2022.104013]
123. Quan C, Ren K, Luo Z, Chen Z, Ling Y. End-to-end deep learning approach for Parkinson's disease detection from speech signals. Biocybernetics and Biomedical Engineering 2022 Apr;42(2):556-574 [doi: 10.1016/j.bbe.2022.04.002]
124. El-Habbak O, Abdelalim A, Mohamed N, Abd-Elaty H, Hammouda M, Mohamed Y. Enhancing Parkinson's disease diagnosis accuracy through speech signal algorithm modeling. CMC-Computers Materials & Continua 2022;70:2953-2969 [doi: 10.32604/cmc.2022.020109]
125. Xie JC, Gan Y, Liang P, Lan R, Gao H. Exploring robust computer-aided diagnosis of Parkinson's disease based on various voice signals. Front. Phys 2022 Nov 3;10:10 [doi: 10.3389/fphy.2022.1048833]
126. Abdul Gafoor S, Theagarajan P. Intelligent approach of score-based artificial fish swarm algorithm (SAFSA) for Parkinson's disease diagnosis. IJICC 2022 Jan 17;15(4):540-561 [doi: 10.1108/IJICC-10-2021-0226]
127. Senturk Z. Layer recurrent neural network-based diagnosis of Parkinson's disease using voice features. Biomed Tech (Berl) 2022 Aug 26;67(4):249-266 [doi: 10.1515/bmt-2022-0022] [Medline: 35659859]
128. Tougui I, Jilbab A, Mhamdi JE. Machine Learning Smart System for Parkinson Disease Classification Using the Voice as a Biomarker. Healthc Inform Res 2022 Jul;28(3):210-221 [FREE Full text] [doi: 10.4258/hir.2022.28.3.210] [Medline: 35982595]

129. Almasoud A, Eisa T, Al-Wesabi F, Elsafi A, Al DM, Yaseen I. Parkinson's Detection Using RNN-Graph-LSTM with Optimization Based on Speech Signals. CMC-Computers, Materials & Continua 2022;72(1):871-886 [FREE Full text] [doi: 10.32604/cmc.2022.024596]
130. Motin MA, Pah ND, Raghav S, Kumar DK. Parkinson’s Disease Detection Using Smartphone Recorded Phonemes in Real World Conditions. IEEE Access 2022;10:97600-97609 [doi: 10.1109/ACCESS.2022.3203973]
131. Yu Q, Zou X, Quan F, Dong Z, Yin H, Liu J, et al. Parkinson's disease patients with freezing of gait have more severe voice impairment than non-freezers during "ON state". J Neural Transm (Vienna) 2022 Mar;129(3):277-286 [doi: 10.1007/s00702-021-02458-1] [Medline: 34989833]
132. Pah ND, Motin MA, Kumar DK. Phonemes based detection of parkinson's disease for telehealth applications. Sci Rep 2022 Jun 11;12(1):9687 [FREE Full text] [doi: 10.1038/s41598-022-13865-z] [Medline: 35690657]
133. Liu W, Liu J, Peng T, Wang G, Balas VE, Geman O, et al. Prediction of Parkinson’s disease based on artificial neural networks using speech datasets. J Ambient Intell Human Comput 2022 Apr 12;1(1):e1 [doi: 10.1007/s12652-022-03825-w]
134. Khaskhoussy R, Ayed Y. Speech processing for early Parkinson’s disease diagnosis: machine learning and deep learning-based approach. Soc. Netw. Anal. Min 2022 Jul 04;12(1):12 [doi: 10.1007/s13278-022-00905-9]
135. Pramanik M, Pradhan R, Nandy P, Bhoi AK, Barsocchi P. The ForEx++ based decision tree ensemble approach for robust detection of Parkinson’s disease. J Ambient Intell Human Comput 2022 Feb 10;1(1):e1 [doi: 10.1007/s12652-022-03719-x]
136. Bertini F, Allevi D, Lutero G, Calzà L, Montesi D. An automatic Alzheimer’s disease classifier based on spontaneous spoken English. Computer Speech & Language 2022 Mar;72:101298 [doi: 10.1016/j.csl.2021.101298]
137. Agbavor F, Liang H. Artificial Intelligence-Enabled End-To-End Detection and Assessment of Alzheimer's Disease Using Voice. Brain Sci 2022 Dec 23;13(1):13 [FREE Full text] [doi: 10.3390/brainsci13010028] [Medline: 36672010]
138. Pérez-Toro PA, Rodríguez-Salas D, Arias-Vergara T, Klumpp P, Schuster M, Nöth E, et al. Interpreting acoustic features for the assessment of Alzheimer’s disease using ForestNet. Smart Health 2022 Dec;26:100347 [doi: 10.1016/j.smhl.2022.100347]
139. Hason L, Krishnan S. Spontaneous speech feature analysis for alzheimer's disease screening using a random forest classifier. Front Digit Health 2022;4:901419 [FREE Full text] [doi: 10.3389/fdgth.2022.901419] [Medline: 36465088]
140. Lin Y, Liyanage BN, Sun Y, Lu T, Zhu Z, Liao Y, et al. A deep learning-based model for detecting depression in senior population. Front Psychiatry 2022;13:1016676 [FREE Full text] [doi: 10.3389/fpsyt.2022.1016676] [Medline: 36419976]
141. Othmani A, Zeghina AO, Muzammel M. A Model of Normality Inspired Deep Learning Framework for Depression Relapse Prediction Using Audiovisual Data. Comput Methods Programs Biomed 2022 Nov;226:107132 [doi: 10.1016/j.cmpb.2022.107132] [Medline: 36183638]
142. Hashim NNWN, Basri NA, Ezzi MAEA, Hashim NMHN. Comparison of classifiers using robust features for depression detection on Bahasa Malaysia speech. IAES Int J Artif Intell.. IAES Int J Artif Intell 2022;11:238-253 [doi: 10.11591/ijai.v11.i1.pp238-253]
143. Sharma G, Umapathy K, Krishnan S. Audio texture analysis of COVID-19 cough, breath, and speech sounds. Biomed Signal Process Control 2022 Jul;76:103703 [FREE Full text] [doi: 10.1016/j.bspc.2022.103703] [Medline: 35464186]
144. Dar JA, Srivastava KK, Ahmed Lone S. Design and development of hybrid optimization enabled deep learning model for COVID-19 detection with comparative analysis with DCNN, BIAT-GRU, XGBoost. Comput Biol Med 2022 Oct 03;150:106123 [FREE Full text] [doi: 10.1016/j.compbiomed.2022.106123] [Medline: 36228465]
145. Dang T, Han J, Xia T, Spathis D, Bondareva E, Siegele-Brown C, et al. Exploring Longitudinal Cough, Breath, and Voice Data for COVID-19 Progression Prediction via Sequential Deep Learning: Model Development and Validation. J Med Internet Res 2022 Jun 21;24(6):e37004 [FREE Full text] [doi: 10.2196/37004] [Medline: 35653606]
146. Dash TK, Chakraborty C, Mahapatra S, Panda G. Gradient Boosting Machine and Efficient Combination of Features for Speech-Based Detection of COVID-19. IEEE J. Biomed. Health Inform 2022 Nov;26(11):5364-5371 [FREE Full text] [doi: 10.1109/JBHI.2022.3197910] [Medline: 35947565]
147. Albadr MAA, Tiun S, Ayob M, Al-Dhief FT. Particle Swarm Optimization-Based Extreme Learning Machine for COVID-19 Detection. Cognit Comput 2022 Oct 12:1-16 [FREE Full text] [doi: 10.1007/s12559-022-10063-x] [Medline: 36247809]
148. Ye W, Jiang Z, Li Q, Liu Y, Mou Z. A hybrid model for pathological voice recognition of post-stroke dysarthria by using 1DCNN and double-LSTM networks. Applied Acoustics 2022 Aug;197:108934 [doi: 10.1016/j.apacoust.2022.108934]
149. Svoboda E, Bořil T, Rusz J, Tykalová T, Horáková D, Guttmann CRG, et al. Assessing clinical utility of machine learning and artificial intelligence approaches to analyze speech recordings in multiple sclerosis: A pilot study. Comput Biol Med 2022 Sep;148:105853 [doi: 10.1016/j.compbiomed.2022.105853] [Medline: 35870318]
150. Bertini F, Allevi D, Lutero G, Montesi D, Calzà L. Automatic Speech Classifier for Mild Cognitive Impairment and Early Dementia. ACM Trans Comput Healthcare 2022;3:1-11 [doi: 10.1145/3469089]
151. Chi NA, Washington P, Kline A, Husic A, Hou C, He C, et al. Classifying Autism From Crowdsourced Semistructured Speech Recordings: Machine Learning Model Comparison Study. JMIR Pediatr Parent 2022 Apr 14;5(2):e35406 [FREE Full text] [doi: 10.2196/35406] [Medline: 35436234]
152. Ditthapron A, Lammert AC, Agu EO. Continuous TBI Monitoring From Spontaneous Speech Using Parametrized Sinc Filters and a Cascading GRU. IEEE J Biomed Health Inform 2022 Jul;26(7):3517-3528 [doi: 10.1109/JBHI.2022.3158840] [Medline: 35290191]

153. Cai T, Ni H, Yu M, Huang X, Wong K, Volpi J, et al. DeepStroke: An efficient stroke screening framework for emergency rooms with multimodal adversarial deep learning. Med Image Anal 2022 Aug;80:102522 [doi: 10.1016/j.media.2022.102522] [Medline: 35810587]
154. Farrús M, Codina-Filbà J, Reixach E, Andrés E, Sans M, Garcia N, et al. Speech-Based Support System to Supervise Chronic Obstructive Pulmonary Disease Patient Status. Applied Sciences 2021 Aug 29;11(17):7999 [doi: 10.3390/app11177999]
155. Ben AR, Ben AY. Speech Processing for Early Alzheimer Disease Diagnosis: Machine Learning Based Approach. New York, NY: IEEE; 2018 Presented at: 2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA); October 28-November 1, 2018; Aqaba, Jordan p. 1-8 [doi: 10.1109/AICCSA.2018.8612831]
156. Verde L, De Pietro G, Sannino G. Artificial Intelligence Techniques for the Non-invasive Detection of COVID-19 Through the Analysis of Voice Signals. Arab J Sci Eng 2021 Oct 08:1-11 [FREE Full text] [doi: 10.1007/s13369-021-06041-4] [Medline: 34642613]
157. Verde L, De Pietro G, Ghoneim A, Alrashoud M, Al-Mutib KN, Sannino G. Exploring the Use of Artificial Intelligence Techniques to Detect the Presence of Coronavirus Covid-19 Through Speech and Voice Analysis. IEEE Access 2021;9:65750-65757 [FREE Full text] [doi: 10.1109/ACCESS.2021.3075571] [Medline: 35256922]
158. Rong P. A novel hierarchical framework for measuring the complexity and irregularity of multimodal speech signals and its application in the assessment of speech impairment in amyotrophic lateral sclerosis. J Speech Lang Hear Res. 2021;64:2996-3014 [doi: 10.1044/2021/JSLHR/20/00743]
159. Wang J, Kothalkar PV, Kim M, Bandini A, Cao B, Yunusova Y, et al. Automatic prediction of intelligible speaking rate for individuals with ALS from speech acoustic and articulatory samples. Int J Speech Lang Pathol 2018 Nov;20(6):669-679 [FREE Full text] [doi: 10.1080/17549507.2018.1508499] [Medline: 30409057]
160. Stegmann GM, Hahn S, Duncan CJ, Rutkove SB, Liss J, Shefner JM, et al. Estimation of forced vital capacity using speech acoustics in patients with ALS. Amyotroph Lateral Scler Frontotemporal Degener 2021;22(sup1):14-21 [doi: 10.1080/21678421.2020.1866013] [Medline: 34348537]
161. Rejaibi E, Komaty A, Meriaudeau F, Agrebi S, Othmani A. MFCC-based Recurrent Neural Network for automatic clinical depression recognition and assessment from speech. Biomedical Signal Processing and Control 2022 Jan;71:103107 [doi: 10.1016/j.bspc.2021.103107]
162. Rao Mv A, Yamini BK, Ketan J, Preetie Shetty A, Pal PK, Shivashankar N, et al. Automatic Classification of Healthy Subjects and Patients With Essential Vocal Tremor Using Probabilistic Source-Filter Model Based Noise Robust Pitch Estimation. J Voice 2023 May;37(3):314-321 [doi: 10.1016/j.jvoice.2021.01.009] [Medline: 33579623]
163. Suppa A, Asci F, Saggio G, Di Leo P, Zarezadeh Z, Ferrazzano G, et al. Voice Analysis with Machine Learning: One Step Closer to an Objective Diagnosis of Essential Tremor. Mov Disord 2021 Jun;36(6):1401-1410 [doi: 10.1002/mds.28508] [Medline: 33528037]
164. Rozenstoks K, Novotny M, Horakova D, Rusz J. Automated Assessment of Oral Diadochokinesis in Multiple Sclerosis Using a Neural Network Approach: Effect of Different Syllable Repetition Paradigms. IEEE Trans Neural Syst Rehabil Eng 2020 Jan;28(1):32-41 [doi: 10.1109/TNSRE.2019.2943064] [Medline: 31545738]
165. Al-Hameed S, Benaissa M, Christensen H, Mirheidari B, Blackburn D, Reuber M. A new diagnostic approach for the identification of patients with neurodegenerative cognitive complaints. PLoS One 2019;14(5):e0217388 [FREE Full text] [doi: 10.1371/journal.pone.0217388] [Medline: 31125389]
166. Roldan-Vasco S, Orozco-Duque A, Suarez-Escudero JC, Orozco-Arroyave JR. Machine learning based analysis of speech dimensions in functional oropharyngeal dysphagia. Comput Methods Programs Biomed 2021 Sep;208:106248 [doi: 10.1016/j.cmpb.2021.106248] [Medline: 34260973]
167. Daqrouq K, Al-Qawasmi AR, Balamesh A, Alghamdi AS, Al-Amoudi MA. The Use of Arabic Vowels to Model the Pathological Effect of Influenza Disease by Wavelets. Comput Math Methods Med 2019;2019 [FREE Full text] [doi: 10.1155/2019/4198462] [Medline: 31915460]
168. Lamba R, Gulati T, Alharbi H, Jain A. A hybrid system for Parkinson's disease diagnosis using machine learning techniques. Int J Speech Technol 2021 Apr 14;25(3):583-593 [doi: 10.1007/s10772-021-09837-9]
169. Benba A, Jilbab A, Hammouch A. Voice assessments for detecting patients with neurological diseases using PCA and NPCA. Int J Speech Technol 2017 Jul 8;20(3):673-683 [doi: 10.1007/s10772-017-9438-9]
170. Arora S, Lo C, Hu M, Tsanas A. Smartphone Speech Testing for Symptom Assessment in Rapid Eye Movement Sleep Behavior Disorder and Parkinson's Disease. IEEE Access 2021;9:44813-44824 [doi: 10.1109/ACCESS.2021.3057715]
171. Jeancolas L, Mangone G, Petrovska-Delacrétaz D, Benali H, Benkelfat BE, Arnulf I, et al. Voice characteristics from isolated rapid eye movement sleep behavior disorder to early Parkinson's disease. Parkinsonism Relat Disord 2022 Feb;95:86-91 [FREE Full text] [doi: 10.1016/j.parkreldis.2022.01.003] [Medline: 35063866]
172. König A, Satt A, Sorin A, Hoory R, Toledo-Ronen O, Derreumaux A, et al. Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. Alzheimers Dement (Amst) 2015 Mar;1(1):112-124 [FREE Full text] [doi: 10.1016/j.dadm.2014.11.012] [Medline: 27239498]

173. Gosztolya G, Vincze V, Tóth L, Pákáski M, Kálmán J, Hoffmann I. Identifying Mild Cognitive Impairment and mild Alzheimer's disease based on spontaneous speech using ASR and linguistic features. Computer Speech & Language 2019 Jan;53:181-197 [doi: 10.1016/j.csl.2018.07.007]
174. Alkenani AH, Li Y, Xu Y, Zhang Q. Predicting Alzheimer's Disease from Spoken and Written Language Using Fusion-Based Stacked Generalization. J Biomed Inform 2021 Jun;118 [FREE Full text] [doi: 10.1016/j.jbi.2021.103803] [Medline: 33965639]
175. Sumali B, Mitsukura Y, Liang K, Yoshimura M, Kitazawa M, Takamiya A, et al. Speech Quality Feature Analysis for Classification of Depression and Dementia Patients. Sensors (Basel) 2020 Jun 26;20(12) [FREE Full text] [doi: 10.3390/s20123599] [Medline: 32604728]
176. Mirzaei S, El Yacoubi M, Garcia-Salicetti S, Boudy J, Kahindo C, Cristancho-Lacroix V, et al. Two-Stage Feature Selection of Voice Parameters for Early Alzheimer's Disease Prediction. IRBM 2018 Dec;39(6):430-435 [doi: 10.1016/j.irbm.2018.10.016]
177. Themistocleous C, Eckerström M, Kokkinakis D. Identification of Mild Cognitive Impairment From Speech in Swedish Using Deep Sequential Neural Networks. Front Neurol 2018;9:975 [FREE Full text] [doi: 10.3389/fneur.2018.00975] [Medline: 30498472]
178. Konig A, Satt A, Sorin A, Hoory R, Derreumaux A, David R, et al. Use of Speech Analyses within a Mobile Application for the Assessment of Cognitive Impairment in Elderly People. Curr Alzheimer Res 2018;15(2):120-129 [doi: 10.2174/1567205014666170829111942] [Medline: 28847279]
179. Ying Y, Yang T, Zhou H. Multimodal fusion for alzheimer's disease recognition. Appl Intell 2022 Dec 01;53(12):16029-16040 [doi: 10.1007/s10489-022-04255-z]
180. Song J, Lee JH, Choi J, Suh MK, Chung MJ, Kim YH, et al. Detection and differentiation of ataxic and hypokinetic dysarthria in cerebellar ataxia and parkinsonian disorders via wave splitting and integrating neural networks. PLoS One 2022;17(6):e0268337 [FREE Full text] [doi: 10.1371/journal.pone.0268337] [Medline: 35658000]
181. Chen F, Yang C, Khishe M. Diagnose Parkinson's disease and cleft lip and palate using deep convolutional neural networks evolved by IP-based chimp optimization algorithm. Biomedical Signal Processing and Control 2022 Aug;77:103688 [doi: 10.1016/j.bspc.2022.103688]
182. Alam MZ, Simonetti A, Brillantino R, Tayler N, Grainge C, Siribaddana P, et al. Predicting Pulmonary Function From the Analysis of Voice: A Machine Learning Approach. Front Digit Health 2022;4:750226 [FREE Full text] [doi: 10.3389/fdgth.2022.750226] [Medline: 35211691]
183. Nilashi M, Ibrahim O, Ahmadi H, Shahmoradi L, Farahmand M. A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. Biocybernetics and Biomedical Engineering 2018;38(1):1-15 [doi: 10.1016/j.bbe.2017.09.002]
184. Tsanas A, Little MA, Fox C, Ramig LO. Objective Automatic Assessment of Rehabilitative Speech Treatment in Parkinson's Disease. IEEE Trans Neural Syst Rehabil Eng 2014 Jan;22(1):181-190 [doi: 10.1109/TNSRE.2013.2293575] [Medline: 26271131]
185. Jain A, Abedinpour K, Polat O, Çalışkan MM, Asaei A, Pfister FMJ, et al. Voice Analysis to Differentiate the Dopaminergic Response in People With Parkinson's Disease. Front Hum Neurosci 2021;15:667997 [FREE Full text] [doi: 10.3389/fnhum.2021.667997] [Medline: 34135742]
186. Pană MA, Busnatu SS, Serbanoiu LI, Vasilescu E, Popescu N, Andrei C. Reducing the heart failure burden in romania by predicting congestive heart failure using artificial intelligence: proof of concept. Appl Sci Switz 2021;11(24):11728 [doi: 10.3390/app112411728]
187. Gao X, Ma K, Yang H, Wang K, Fu B, Zhu Y, et al. A rapid, non-invasive method for fatigue detection based on voice information. Front Cell Dev Biol 2022;10:994001 [FREE Full text] [doi: 10.3389/fcell.2022.994001] [Medline: 36176279]
188. Bárcenas R, Fuentes-García R, Naranjo L. Mixed kernel SVR addressing Parkinson's progression from voice features. PLoS One 2022;17(10):e0275721 [FREE Full text] [doi: 10.1371/journal.pone.0275721] [Medline: 36206238]
189. Lab L. Coswara-Data. GitHub. 2022. URL: https://github.com/iiscleap/Coswara-Data [accessed 2022-10-31]
190. Setia MS. Methodology Series Module 3: Cross-sectional Studies. Indian J Dermatol 2016;61(3):261-264 [FREE Full text] [doi: 10.4103/0019-5154.182410] [Medline: 27293245]
191. Caruana E, Roman M, Hernández-Sánchez J, Solli P. Longitudinal studies. J Thorac Dis 2015 Nov;7(11):E537-E540 [FREE Full text] [doi: 10.3978/j.issn.2072-1439.2015.10.63] [Medline: 26716051]
192. White RT, Arzi HJ. Longitudinal Studies: Designs, Validity, Practicality, and Value. Res Sci Educ 2005 Mar;35(1):137-149 [doi: 10.1007/s11165-004-3437-y]
193. Scherer RW, Saldanha IJ. How should systematic reviewers handle conference abstracts? A view from the trenches. Syst Rev 2019 Nov 07;8(1):264 [FREE Full text] [doi: 10.1186/s13643-019-1188-0] [Medline: 31699124]

Abbreviations

AD: Alzheimer disease
ADBC: Alzheimer Dementia Bank blog corpus

ALS: amyotrophic lateral sclerosis
ANN: artificial neural network
BLA: base line acoustic
CI: cognitive impairment
CFS: collected for study
CHRSD: Corona Hack Respiratory Sound data set
DT: decision tree
GB: gradient boosting
GMM: Gaussian mixture model
HC: healthy control
KNN: K-nearest neighbor
LR: logic regression
MCI: mild cognitive impairment
MeML: mixed effect machine learning
MeSH: Medical Subject Headings
MFCC: Mel-frequency cepstral coefficients
ML: machine learning
NB: naïve Bayes
NCC: neurodegenerative cognitive complaint
NCVS: National Center for Voice and Speech
ND: neurological disease
NG: not given
PA: passive active
PARCZ: Czech Parkinsonian Speech Database
PD: Parkinson disease
PICO: population, intervention, comparison, and outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RF: random forest
RP: recurrence plot
SLR: systematic literature review
SQ: subquestion
SVM: support vector machine
SVR: support vector regression
TBIBank: Traumatic Brain injury bank
TQWT: tunable Q-factor wavelet transform
UCI: University of California, Irvine

Edited by A Mavragani; submitted 30.01.23; peer-reviewed by F Chu, V Martin; comments to author 09.03.23; revised version received 26.04.23; accepted 23.05.23; published 19.07.23

Please cite as:
Idrisoglu A, Dallora AL, Anderberg P, Berglund JS
Applied Machine Learning Techniques to Diagnose Voice-Affecting Conditions and Disorders: Systematic Literature Review
J Med Internet Res 2023;25:e46105
URL: https://www.jmir.org/2023/1/e46105
doi: 10.2196/46105
PMID: