A Systematic Review on Machine Learning Techniques for Early Detection of Mental, Neurological and Laryngeal Disorders Using Patient's Speech

Mohammadjavad Sayadi [1], Vijayakumar Varadarajan [2,3,4,*], Mostafa Langarizadeh [1,*], Gholamreza Bayazian [5] and Farhad Torabinezhad [6]

[1] Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran 13, Iran

[2] School of Computer Science and Engineering, The University of New South Wales, Sydney, NSW 1466, Australia

[3] Dean International, Ajeenkya D Y Patil University, Pune 412105, India

[4] Swiss School of Business and Management, 1213 Geneva, Switzerland

[5] ENT and Head &Neck Research Center and Department, The Five Sense Health Institute, Rasoul Akram Medical Complex, Iran University of Medical Sciences, Tehran 13, Iran

[6] Rehabilitation Research Center, Department of Speech Therapy, School of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran 13, Iran

* Correspondence: v.varadarajan@unsw.edu.au (V.V.); langarizadeh.m@iums.ac.ir (M.L.)

Abstract: There is a substantial unmet need to diagnose speech-related disorders effectively. Machine learning (ML), as an area of artificial intelligence (AI), enables researchers, physicians, and patients to solve these issues. The purpose of this study was to categorize and compare machine learning methods in the diagnosis of speech-based diseases. In this systematic review, a comprehensive search for publications was conducted on the Scopus, Web of Science, PubMed, IEEE and Cochrane databases from 2002–2022. From 533 search results, 48 articles were selected based on the eligibility criteria. Our findings suggest that the diagnosing of speech-based diseases using speech signals depends on culture, language and content of speech, gender, age, accent and many other factors. The use of machine-learning models on speech sounds is a promising pathway towards improving speech-based disease diagnosis and treatments in line with preventive and personalized medicine.

Keywords: voice disorder; verbal fluency disorder; machine learning; speech-based disease

1. Introduction

In the United States, 25% of adults, 18% of adolescents and 13% of children have a mental disorder. Although these disorders have larger economic impacts than others, governments still spend less on them [1].

Major depression is one of the most common mental disorders, affecting over 300 million people [2]. The global prevalence of depression was estimated at 4.4% HE in 2015, affecting more women than men over a two-year period [3]. Early depressive symptoms, such as psychomotor retardation and cognitive impairment, are usually associated with language impairment. Thus, depressive speech has been described by clinicians as monotonous, uninteresting, and lacking energy. The difference may allow for the detection of depression by analyzing the acoustics of depressed people's voices [3,4].

Alzheimer's disease is a type of neurological disorder that gradually reduces a patient's mental abilities. The main symptoms of this disease include memory loss, difficulties in making decisions, and the incorrect choice of words. Therefore, speech signal processing in this disease has attracted the attention of many researchers over the past decade. A diagnosis of Alzheimer's disease using audio signals depends on culture, language, language content, gender, age, accent, and many other factors. Parkinson's disease (PD) is the

second most common neurodegenerative disease after Alzheimer's disease (AD). PD is reported to predominate in 0.3% of the general population in developed countries, whereas PD prevalence in the elderly population (60 years and older) is 1%. Voice impairment has been reported to be an early biomarker for this disease. Moreover, the proposed intelligent system can be used as a means of prodromal diagnosis [5,6].

Since medical-decision support systems are being developed in various fields and lead to early diagnosis, much research has been conducted to create intelligent disease diagnosis models using a patient's speech.

This study's objective was to perform an updated systematic review of the available literature to appraise the machine learning models in diagnosing mental, neurological and laryngeal diseases based on a patient's speech/voice.

2. Related Work

Diseases related to the human vocal system, such as laryngeal cancer and polyps, usually have a serious impact on the patient's health and social life. Fortunately, most of these diseases can be cured if detected early. Because laryngeal syndrome actually causes voice abnormalities in the patient (such as wheezing and hoarseness—two major symptoms of voice system dysfunction), some professionals can detect the problem simply by listening to the patient and deciding to prescribe a test such as a laryngoscopy. However, these tests are very expensive and time consuming. Additionally, they are invasive and cause patient discomfort. Therefore, some preliminary research is worthwhile. A major drawback of perceptual studies is their inherent subjectivity, which is unreliable and difficult to quantify. To overcome these problems, researchers have sought reliable methods to distinguish between healthy and abnormal voices, usually based on speech signal processing techniques [4,7].

To be able to diagnose healthy and abnormal voices, we first need to find some distinct features. The speech is then classified as healthy or diseased, using several classification methods such as SVM and neural networks (NN) [4,8].

In modern language technology systems, many methods and algorithms have emerged that rely on the interdisciplinary research fields of signal processing and artificial intelligence. Machine learning helps to build good models using real-world features to do real work. The new techniques developed in the machine learning paradigm have brought enormous improvements to language technology. The main concept of machine learning is learning from a given dataset to analyze, recognize or complete a given task. Moreover, various mathematicians, psychologists, engineers, medical scientists, computer scientists and many others have invented and sometimes rediscovered some methods of solving problems. Therefore, different methods applicable to emotion prediction in speech recognition were presented in comparative frameworks [2,9].

Machine learning is an intensive research field with successful applications to solve various problems in health sciences, such as neurological diseases, larynx, stress, depression, autism and Parkinson's disease [6,10–14]. For example, one of the most important uses of machine learning is its use in solving problems for Spanish-speaking Alzheimer's patients, with features such as the number of verbs, nouns and conjunctions with radial basis function kernel [15]. More than 60 percent of speech is voiced and this part very important to develop intelligent systems based on speech data [16].

Materials and Methods

The Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines was followed (Figure 1).
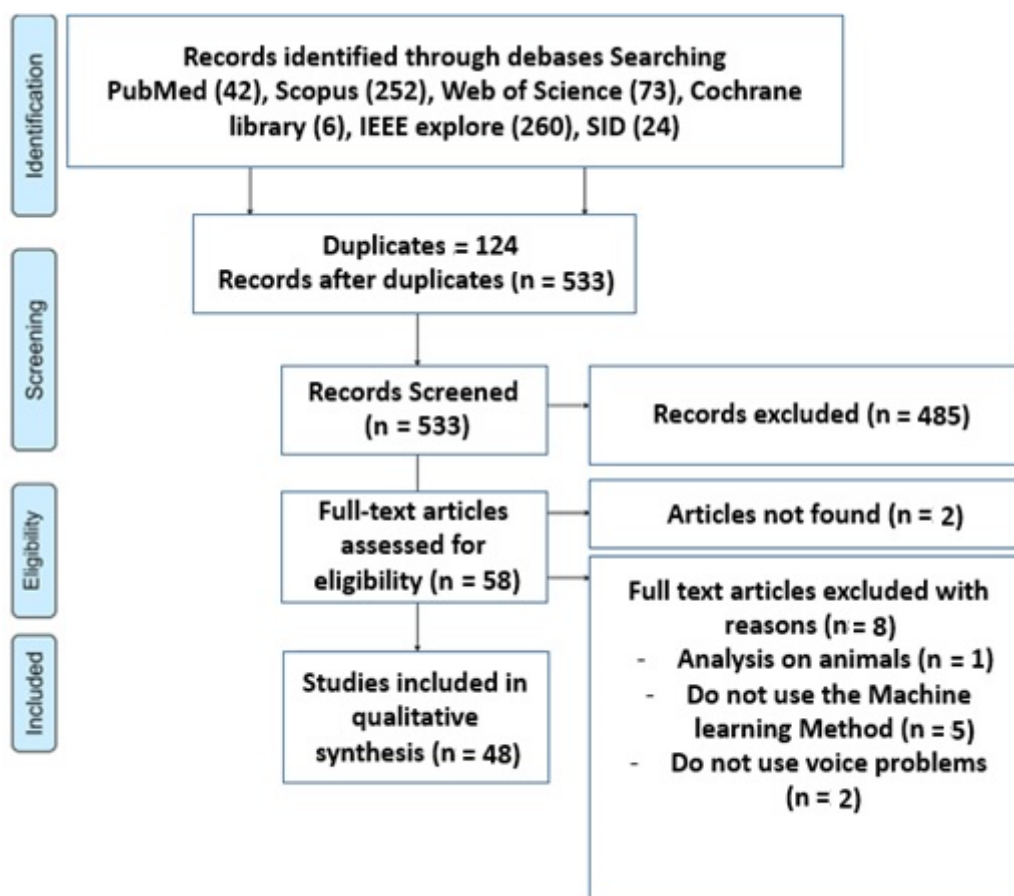
Figure 1. PRISMA flow diagram of study inclusion and exclusion criteria for the systematic review.

The search aimed at identifying articles from the last 10 years that include machine learning methods for the diagnosis of mental, neurological and laryngeal disorders, based on a patient's speech. A comprehensive analysis of the publication was carried out using data from the Scopus, Web of Science (WOS), PubMed, IEEE and Cochrane databases from 2002–2022.

Based on the inclusion and exclusion criteria of this study (which were the time period and required English-language-based articles), articles were added to this study, their information was extracted in a checklist, and the collected data were analyzed using descriptive statistics. Incoming and outgoing articles were finally reported by a PRISMA flowchart. JBI's critical appraisal tools were used for assessing the trustworthiness.

Because of the different terminology of selected databases in indexing papers, and in an attempt to include all relevant articles, thesauruses were used and a systematic record in databases of subject headings were used to index articles. To organize the search systematically, search terms were grouped around four expressions: "Machine Learning", "Lower Gastrointestinal Cancers", "Cancer" and "Diagnosis and Screening". Further elaboration of the search terms used for eligible articles in the four expressions can be seen in Table 1. The search strategy consisted of four terms, respectively: Term1 (Machine Learning), Term2 (Speech), Term3 (Neurological, Mental or Laryngeal disorders) and Term4 (Diagnosis and Screening). The terms within each term were a mix of medical subject heading (MeSH) terms and synonyms. The AND operator was applied between each term and the OR operator was applied between each MeSH term and synonym. Only a few limitations were marked in the search criteria, such as studies written in languages other than English, literature published before 2002 and studies that were done on treatment and follow-up. The following studies were also excluded: (a) those that were used for animals and (b) those that did not use machine learning methods.

Table 1. The terms below show the search strategy used in this research. Each term consists of MeSH terms and synonyms.

Machine Learning	Speech	Neurological, Mental or Laryngeal	Diagnosis or Screening
Machine learning OR artificial intelligence OR deep learning OR neural networks OR data mining OR text mining	Speech or voice	Neurological or mental or laryngeal	Prediction OR diagnosis OR detection OR screening OR predict

A review of 657 papers was undertaken to determine the most related ones. After removing duplicates, only 533 remained. Finally, 48 articles were selected against the eligibility criteria.

Screening was performed by all authors by reading the title and abstract. From each article, the following features were synthesized if available: disorders, sample size, presence of control group, age, clinically assessed or self-assessed, clinical scales used for diagnosis, tasks to obtain speech, predictive model, highest performance or statistical significance, type of validation or test set, and other relevant findings (especially if it was stated which features were predictive).

Finally, JBI's critical appraisal tool was used for assessing the trustworthiness, relevance and results of published papers [17].

3. Results

A general description of the search results is shown in Table 2. Based on the studies, the articles can be divided into three general categories: neurological disorders, laryngeal disorders and mental disorders. According to Table 2, the frequency of each category is specified; this shows that most of the articles that used voice features for diagnosis are related to mental disorders.

Table 2. Summary of systematic review results.

Row	Disorders		Articles % (*n*)
1	Neurological disorders	PD–autism	25 (12)
2	Laryngeal disorders	Aphasia–speech	12.5 (6)
3	Mental disorders	MDD–anxiety and depression–schizophrenia–bipolar	62.5 (30)

Table 3 shows the papers included in this systematic review, along with information from each paper that was the target of this systematic review. It is shown in this table each of the included papers focused on one of the targeted disease by this systematic review. In front of the name of the disease, the name of the category of that disease is also mentioned. This table also shows which machine learning algorithm each article used to diagnose its target disease. For a better and faster comparison, the most important evaluation metric of each study is stated in front of the name of the machine learning algorithm. Another important issue in this systematic review is which category of speech/sound features have been used by each of the included papers.

A general review on this table shows that the SVM algorithm has been the most used among all machine learning algorithms. Regardless of the algorithm used, only acoustic features have been used to diagnose laryngeal diseases, and mainly prosodic features have been used for neurological and mental diseases.

Table 3. Classification of voice features and detection algorithms based on speech disorders.

Author (Ref#)	Country	Year	Disease (Category)	Algorithm (Metric)	Features
Low et al. [1]		2017	MDD (Mental)	SVM (AUC = 89%)	Prosodic
Shin et al. [2]	Korea	2021	Anxiety, stress and minor depression (Mental)	MLP (AUC = 65.9%)	Acoustic
Espinola et al. [3]	Brazil	2021	MDD (Mental)	SVM (ACC = 89.14%)	Prosodic
Ghasemzadeh et al. [4]	Iran	2015	Voca l(Laryngeal)	SVM (ACC = 99.3%)	Acoustic
Rahman et al. [5]	Pakistan	2021	Parkinson (Neurological)	SVM	Acoustic
Vigneswari et al. [6]	Canada	2021	Parkinson (Neurological)	Gradient boosting (ACC = 95.16%)	Acoustic
Osman et al. [7]		2014	Anxiety, stress and minor depression (Mental)	SVM (ACC = 77.5)	Acoustic
Farhoumandi et al. [8]	Iran	2021	MDD (Mental)	SVM (ACC = 81.81%)	Acoustic
Punithavathi et al. [9]		2022	Anxiety, stress and minor depression (Mental)	Deep learning (ACC = 90.5)	Acoustic
Ming ET AL. [10]		2019	Autism (Neurological)	SVM	Prosodic
Sumali et al. [11]	Korea	2021	Anxiety, stress and minor depression (Mental)	SVM (ACC = 67.2)	Acoustic
Izumi et al. [12]	China	2021	Anxiety, stress and minor depression (Mental)	Gradient boosting	Prosodic
Weintraub et al. [13]	USA	2021	Bipolar (Mental)	Decision tree (ACC = 81.8%)	Prosodic
Xu et al. [14]	China	2020	Parkinson (Neurological)	Deep learning (ACC = 91.25%)	Acoustic
Benba et al. [18]		2016	Parkinson (Neurological)	SVM	Acoustic
Vasquez-Correa et al. [19]	Colombia	2019	Parkinson (Neurological)	Deep learning (ACC = 97.3%)	Acoustic
Jeancolas et al. [20]	France	2022	Parkinson (Neurological)	SVM (ACC = 89%)	Prosodic
Zahid et al. [21]	Pakistan	2020	Parkinson (Neurological)	Deep learning (ACC = 99.7%)	Acoustic
Berus et al. [22]		2018	Parkinson (Neurological)	Artificial neural networks	Acoustic
Chao et al. [23]		2014	Parkinson (Neurological)	SVM	Acoustic
Eni et al. [24]	Israel	2020	Autism (Neurological)	Deep learning (RMSE = 4.65)	Prosodic
Lin et al. [25]	Taiwan	2020	Autism (Neurological)	SVM (ACC = 66.8)	Acoustic
Mahmoud et al. [26]		2021	Aphasia (Laryngeal)	Deep learning (ACC = 99.3)	Acoustic
Fonseca et al. [27]	Brazil	2007	Aphasia (Laryngeal)	SVM (ACC = 90%)	Acoustic
Verikas et al. [28]	Litvania	2010	Aphasia (Laryngeal)	SVM (ACC = 72.1%)	Acoustic
Antti et al. [29]		2011	Aphasia (Laryngeal)	KNN (90.5%)	Acoustic
Ali et al. [30]		2016	Speech disorders (Laryngeal)	SVM (ACC = 100%)	Acoustic
Corcoran et al. [31]	USA	2018	MDD (Mental)	UCLA classifier	Acoustic
Behroozi et al. [32]		2016	MDD (Mental)	SVM (ACC = 87.5%)	Acoustic
Bedi et al. [33]	USA	2014	MDD (Mental)	SVM (ACC = 88%)	Prosodic
Zhao et al. [34]	China	2020	MDD (Mental)	Neural network (RMSE = 5.51)	Prosodic
Bedi et al. [35]	USA	2015	Schizophrenia	LSA	Prosodic
Rezaii et al. [36]	USA	2019	MDD (Mental)	Neural network (ACC = 93%)	Prosodic
Gavrilescu et al. [37]	Romania	2019	MDD (Mental)	Neural network (ACC = 80.75)	Prosodic
Goldberg et al. [38]	USA	2020	MDD (Mental)	NLP	Acoustic
Zhang et al. [39]	China	2021	MDD (Mental)	Deep learning	Acoustic
Song et al. [40]	Singapore	2014	MDD (Mental)	SVM	Acoustic
Fischer et al. [41]	Germany	2011	MDD (Mental)		Acoustic
Di et al. [42]	China	2021	MDD (Mental)	i-Vector	Acoustic
McGinnis et al. [43]	India	2019	Anxiety, stress and minor depression (Mental)	SVM (ACC = 80%)	Acoustic
Wang et al. [44]		2019	Anxiety, stress and minor depression (Mental)	MANCOVA	Acoustic
He et al. [45]	China	2018	Anxiety, stress and minor depression (Mental)	CNN (RMSE = 10.44)	Acoustic
Jenei et al. [46]	Hungary	2021	Anxiety, stress and minor depression (Mental)	CNN (ACC = 85.2%)	Acoustic
Huang et al. [47]	China	2020	Schizophrenia (Mental)	Deep learning (ACC = 84%)	Prosodic
Xu et al. [48]		2022	Schizophrenia (Mental)	Time series	Prosodic
Fisher et al. [49]	Canada	2008	Schizophrenia (Mental)	Mismatch negative (ACC = 82–99%)	Prosodic
Juanita et al. [50]		2007	Schizophrenia (Mental)	Mismatch negative	Prosodic
Arevian et al. [51]	USA	2020	Schizophrenia (Mental)	SVM (AUC = 81%)	Prosodic

CNN: Convolutional Neural Network; SVM: Support Vector Machine; RF: Random Forest; DNN: Deep Neural Network; MLP: Multilayer Perceptron; GA: Genetic Algorithm; NN: Neural Network; LSTM: Long Short-Term Memory; LSA: Latent Semantic Analysis; MMN: Mismatch Negativity; DT: Decision Tree; LS-SVM: Least Squares Support Vector Machines; FFNN: Feedforward Neural Network.

4. Discussion

To our knowledge, there was no precise classification of the types of machine learning algorithms in speech-based disease diagnosis that would specify which machine learning algorithms can be used for this purpose. We did not find any systematic study that checked which speech-based disorders and what origin and symptoms can be taken into consideration, what the characteristics are for each of the used algorithms, or what evaluation metrics have been used. Therefore, this systematic study was planned regarding the relationship among three important matters. Basic symptoms of the patient population, speech disorders and the machine learning methods for the development of an early detection model.

In this systematic review, four main issues were considered. First was the name of the disease related to speech; second was the features of speech that were affected by the given disease and therefore used in modeling; third was the machine learning algorithms that were used for modeling; and finally, the evaluation metrics for machine learning models used to detect the speech-based disorders.

Considering the three different categories of diseases that have been investigated in this systematic review, the results of this study are discussed independently for each category.

Neurological diseases: Twelve of the included studies investigated machine learning tools for diagnosing neurological diseases based on patient speech [5,6,10,14,18–25]. Nine papers presented a machine learning model for the early detection of Parkinson's disease [5,6,14,18–23]. Five articles in this category were better than all the other articles, and the results of their evaluation metrics were also better [14,19–21]. The deep learning algorithm had been used the most [14,19,21] and the best accuracy was reported in the study of Zahid et al. in 2020 in Pakistan, wherein they conducted their study on a Spanish language dataset using deep learning with 99.7% of accuracy [21]. Only one study used prosodic features to diagnose Parkinson's, which is related to Hammer et al. from France in 2022. In this study, speech fluency and speech rhythm were used in SVM to implement a diagnostic model on a French language data set, and the accuracy of this model was 89% [20]. The rest of the studies used acoustic features for analysis and modeling. Three of the included articles deal with the diagnosis of autism using machine learning [10,24,25]. Two studies of this category presented better results, both of which used deep learning [24,25]. A quantitative study by Eni and colleagues conducted in 2020 in Israel determined the severity of autism in patients using the prosodic features of speech. In this study, RMSE = 4.65 and mean correlation = 0.72 [24]. In the qualitative study of Lin et al. in 2020, the presence of autism in patients was diagnosed using the acoustic characteristics of the patient's voice with an accuracy of 66.8 [25].

Laryngeal diseases: Six of the included articles presented machine learning models for the diagnosis of laryngeal diseases [4,26–30]. All these studies used acoustic features for modeling, and five articles presented the results well [4,26–28,30]. All the studies of this category used the SVM algorithm for modeling, except Mahmoud et al.'s study in 2021, which implemented a diagnostic model with deep learning (ACC = 99.2) [26]. The best performance was related to the study of Ali et al. in 2016, where the accuracy of their model was 100% [30].

Mental disease: Thirty of the reviewed articles dealt with the diagnosis of mental illnesses using speech [1–3,7–9,11,13,31–51]. In fourteen studies, machine learning algorithms were used to diagnose Major Depressive Disorder (MDD) [1,3,8,31–34,36–42]. Five studies presented better results than the others [3,33,34,36,37]. In all these studies, prosodic features of speech and artificial neural networks were used for disease diagnosis. The article by Bedi et al. is a quantitative study to calculate MDD severity in 2020 in China for a Chinese-language dataset. In this study, RMSE = 5.51 and MAE = 4.2 were obtained [34]. Among the qualitative studies, the article by Rezaii et al., which was conducted in 2019 in the United States, obtained the highest accuracy with artificial neural networks (ACC = 93%) [36]. This level of accuracy is much better compared to the article by Gavrilescu et al. [37] in 2019 in Romania, which used artificial neural networks with an accuracy of 80.75%. The article by Espinola et al. [3] in 2021 in Brazil and the article by Bedi et al. [33] in 2014 in the United

States used the SVM algorithm for modeling, and the accuracy of their models was 89.14% and 88%, respectively.

Eight of the included studies presented machine learning models for diagnosing anxiety and minor depression [2,7,9,11,43–46]. Among these articles, five articles obtained better results, in all of which acoustic sound vignettes were used to diagnose the disease [2,11,43,45,46]. He et al.'s article is a quantitative study to detect disease severity using the SVM algorithm, which was conducted in China in 2018. In this article, RMSE = 10.44 and MAE = 8.60 [45]. The best accuracy was related to the study by Jenei et al., which was conducted in Hungary in 2021 and obtained an accuracy of 85.2% using CNN [46]. McGinnis et al.'s article in 2019 in India [43] and Sumali et al.'s article in 2020 in Japan [11] used the SVM algorithm, and their accuracy was reported as 80% and 67.2%, respectively. The most recent paper in this category belongs to Shin et al. in 2021 in Korea, which obtained an accuracy of 65.9% using MLP.

Five articles have examined the diagnosis of schizophrenia using machine learning [35,47–50]. Among them, two studies presented better results [47,49]. Huang et al. from China in 2020, using deep learning, presented a disease diagnosis model with 84% accuracy [47]; Fisher et al. from Canada in 2008, with MMN (Mismatch Negative), provided a variable between 82% and 99%. Both studies used prosodic features.

Two of the included articles are related to the diagnosis of bipolar disease using machine learning [13,51]. In the study by Weintraub et al., which was conducted in 2021 in America, they implemented a diagnostic model with an accuracy of 81.8% using a decision tree and the prosodic features of speech [13]. Arevian and colleagues also presented a diagnostic model with the prosodic features of speech. In this study, AUC was 81% [51].

Although the diagnosis of neurological diseases using speech deep learning has been the most repeated in the reviewed studies, the findings show that, in general, SVM is the most used for the diagnosis of speech-related diseases. All the studies that have dealt with the diagnosis of laryngeal diseases have used acoustic features—for the diagnosis of neurological and mental diseases more so than for the features of voice speech. This means that, in order to diagnose mental and neurological diseases, it is necessary for the patient to speak, while for laryngeal diseases, the voice of the larynx is usually sufficient.

One of the problems that make it difficult to compare studies is that the prosodic features are different from one language to another, and therefore one cannot expect an algorithm to provide the same results in two languages. On the other hand, studies that have been conducted in the same language usually use different features.

This study had some limitations. First, unpublished and non-English studies were not included in this review. Second, the quality of the included studies was not assessed. Also, heterogeneities between studies prevented the conducting of a meta-analysis. Finally, there was not a standard for the feature selection or the dataset in the reviewed studies.

Despite these limitations, the current review's findings provide crucial recommendations for further research.

First, future study should focus on implementing a standard dataset and also some indicators for each language, based on the acoustic and prosodic features of patient speech for each category of mentioned disease. Second, some research is needed to focus on how to generalize the use of the results of these studies, considering that in the studies conducted, the patient is evaluated in special conditions and by saying pre-determined sentences.

5. Conclusions

The present systematic review provides evidence of how the prosodic or acoustic features of a patient's speech can be affected by mental, neurological or laryngeal diseases, as well as a comparison of machine learning methods in diagnosing these diseases.

According to our study, the results demonstrate a classification of machine learning tools for detecting and screening severe disorders in a cost-effective manner. New mechanisms are needed to enhance the medical diagnosis of disease. Also, our findings show the relevant feature group for each disease category, which is an important phase for implementing a machine learning model.

This approach has considerable potential to address a critical gap in the diagnosing of some severe diseases using patient speech; nevertheless, the findings from these primary studies support further research in implementing machine learning model as a clinical decision support system based on speech/voice.

References

1. Guntuku, S.C.; Yaden, D.B.; Kern, M.L.; Ungar, L.H.; Eichstaedt, J.C. Detecting depression and mental illness on social media: An integrative review. *Curr. Opin. Behav. Sci.* **2017**, *18*, 43–49. [CrossRef]
2. Shin, D.; Cho, W.I.; Park, C.H.K.; Rhee, S.J.; Kim, M.J.; Lee, H.; Kim, N.S.; Ahn, Y.M. Detection of minor and major depression through voice as a biomarker using machine learning. *J. Clin. Med.* **2021**, *10*, 3046. [CrossRef] [PubMed]
3. Espinola, C.W.; Gomes, J.C.; Pereira, J.M.S.; dos Santos, W.P. Detection of major depressive disorder using vocal acoustic analysis and machine learning—An exploratory study. *Res. Biomed. Eng.* **2021**, *37*, 53–64. [CrossRef]
4. Ghasemzadeh, H.; Tajik Khass, M.; Khalil Arjmandi, M.; Pooyan, M. Detection of vocal disorders based on phase space parameters and Lyapunov spectrum. *Biomed. Signal Process Control* **2015**, *22*, 135–145. [CrossRef]
5. Rahman, A.; Rizvi, S.S.; Khan, A.; Abbasi, A.A.; Khan, S.U.; Chung, T.S. Parkinson's disease diagnosis in cepstral domain using MFCC and dimensionality reduction with SVM classifier. *Mob. Inf. Sys.* **2021**, *2021*, 8822069. [CrossRef]
6. Vigneswari, D.A.; Aravinth, J. (Eds.) Parkinson's disease Diagnosis using Voice Signals by Machine Learning Approach. In Proceedings of the 2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), Karnataka, India, 27–28 August 2021.
7. Abdel-Hamid, O.; Mohamed, A.R.; Jiang, H.; Deng, L.; Penn, G.; Yu, D. Convolutional neural networks for speech recognition. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2014**, *22*, 1533–1545. [CrossRef]
8. Farhoumandi, N.; Mollaey, S.; Heysieattalab, S.; Zarean, M.; Eyvazpour, R. Facial emotion recognition predicts alexithymia using machine learning. *Comput. Intell. Neurosci.* **2021**, *2021*, 2053795. [CrossRef]
9. Punithavathi, R.; Sharmila, M.; Avudaiappan, T.; Raj, I.I.; Kanchana, S.; Mamo, S.A. Empirical investigation for predicting depression from different machine learning based voice recognition techniques. *Evid. Based Complement. Altern. Med. eCAM* **2022**, *2022*, 6395860. [CrossRef]
10. Li, M.; Tang, D.; Zeng, J.; Zhou, T.; Zhu, H.; Chen, B.; Zou, X. An automated assessment framework for atypical prosody and stereotyped idiosyncratic phrases related to autism spectrum disorder. *Comput. Speech Lang.* **2019**, *56*, 80–94. [CrossRef]
11. Sumali, B.; Mitsukura, Y.; Liang, K.C.; Yoshimura, M.; Kitazawa, M.; Takamiya, A.; Fujita, T.; Mimura, M.; Kishimoto, T. Speech quality feature analysis for classification of depression and dementia patients. *Sensors* **2020**, *20*, 3599. [CrossRef]
12. Izumi, K.; Minato, K.; Shiga, K.; Sugio, T.; Hanashiro, S.; Cortright, K.; Kudo, S.; Fujita, T.; Sado, M.; Maeno, T.; et al. Unobtrusive sensing technology for quantifying stress and well-being using pulse, speech, body motion, and electrodermal data in a workplace setting: Study concept and design. *Front. Psychiatry* **2021**, *12*, 611243. [CrossRef] [PubMed]
13. Weintraub, M.J.; Posta, F.; Arevian, A.C.; Miklowitz, D.J. Using machine learning analyses of speech to classify levels of expressed emotion in parents of youth with mood disorders. *J. Psychiatr. Res.* **2021**, *136*, 39–46. [CrossRef] [PubMed]
14. Xu, Z.J.; Wang, R.F.; Wang, J.; Yu, D.H. Parkinson's disease detection based on spectrogram-deep convolutional generative adversarial network sample augmentation. *IEEE Access* **2020**, *8*, 206888–206900. [CrossRef]
15. Hernández-Domínguez, L.; García-Cano, E.; Ratté, S.; Sierra, G. (Eds.) Detection of Alzheimer's disease based on automatic analysis of common objects descriptions. In Proceedings of the 7th Workshop on Cognitive Aspects of Computational Language Learning, Berlin, Germany, 11 August 2016.
16. Bachu, R.; Kopparthi, S.; Adapa, B.; Barkana, B. (Eds.) Separation of voiced and unvoiced using zero crossing rate and energy of the speech signal. In *American Society for Engineering Education (ASEE) Zone Conference Proceedings*; Society for Engineering Education: Washington, DC, USA, 2008.
17. Porritt, K.; Gomersall, J.; Lockwood, C. JBI's systematic reviews: Study selection and critical appraisal. *AJN Am. J. Nurs.* **2014**, *114*, 47–52. [CrossRef]
18. Benba, A.; Jilbab, A.; Hammouch, A. Analysis of multiple types of voice recordings in cepstral domain using MFCC for discriminating between patients with Parkinson's disease and healthy people. *Int. J. Speech Technol.* **2016**, *19*, 449–456. [CrossRef]
19. Vasquez-Correa, J.C.; Arias-Vergara, T.; Orozco-Arroyave, J.R.; Eskofier, B.; Klucken, J.; Noth, E. Multimodal assessment of parkinson's disease: A deep learning approach. *IEEE J. Biomed. Health Informat.* **2016**, *23*, 1618–1630. [CrossRef]

20. Jeancolas, L.; Mangone, G.; Petrovska-Delacrétaz, D.; Benali, H.; Benkelfat, B.E.; Arnulf, I.; Corvol, J.C.; Vidailhet, M.; Lehéricy, S. Voice characteristics from isolated rapid eye movement sleep behavior disorder to early Parkinson's disease. *Park. Relat. Disord.* **2022**, *95*, 86–91. [CrossRef]

21. Zahid, L.; Maqsood, M.; Durrani, M.Y.; Bakhtyar, M.; Baber, J.; Jamal, H.; Mehmood, I.; Song, O.-Y. A spectrogram-based deep feature assisted computer-aided diagnostic system for parkinson's disease. *IEEE Access* **2020**, *8*, 35482–35495. [CrossRef]

22. Berus, L.; Klancnik, S.; Brezocnik, M.; Ficko, M. Classifying parkinson's disease based on acoustic measures using artificial neural networks. *Sensors* **2019**, *19*, 16. [CrossRef]

23. Ma, C.; Ouyang, J.; Chen, H.L.; Zhao, X.H. An efficient diagnosis system for parkinson's disease using kernel-based extreme learning machine with subtractive clustering features weighting approach. *Comput. Math. Methods Med.* **2014**, *2014*, 985789. [CrossRef]

24. Eni, M.; Dinstein, I.; Ilan, M.; Menashe, I.; Meiri, G.; Zigel, Y. Estimating Autism Severity in Young Children From Speech Signals Using a Deep Neural Network. *IEEE Access* **2020**, *8*, 139489–139500. [CrossRef]

25. Lin, Y.; Gau, S.S.; Lee, C. A multimodal interlocutor-modulated attentional BLSTM for classifying autism subgroups during clinical interviews. *IEEE J. Sel. Top. Signal Process.* **2020**, *14*, 299–311. [CrossRef]

26. Mahmoud, S.S.; Kumar, A.; Li, Y.; Tang, Y.; Fang, Q. Performance evaluation of machine learning frameworks for aphasia assessment. *Sensors* **2021**, *21*, 2582. [CrossRef] [PubMed]

27. Fonseca, E.S.; Guido, R.C.; Scalassara, P.R.; Maciel, C.D.; Pereira, J.C. Wavelet time-frequency analysis and least squares support vector machines for the identification of voice disorders. *Comput. Biol. Med.* **2007**, *37*, 571–578. [CrossRef]

28. Verikas, A.; Gelzinis, A.; Bacauskiene, M.; Hållander, M.; Uloza, V.; Kaseta, M. Combining image, voice, and the patient's questionnaire data to categorize laryngeal disorders. *Artif. Intell. Med.* **2010**, *49*, 43–50. [CrossRef]

29. Järvelin, A.; Juhola, M. Comparison of machine learning methods for classifying aphasic and non-aphasic speakers. *Comput. Methods Programs Biomed.* **2011**, *104*, 349–357. [CrossRef]

30. Ali, S.M.; Karule, P.T. MFCC, LPCC, formants and pitch proven to be best features in diagnosis of speech disorder using neural networks and SVM. *Int. J. Appl. Eng. Res.* **2016**, *11*, 897–903.

31. Corcoran, C.M.; Carrillo, F.; Fernández-Slezak, D.; Bedi, G.; Klim, C.; Javitt, D.C.; Bearden, C.E.; Cecchi, G.A. Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World Psychiatry* **2018**, *17*, 67–75. [CrossRef]

32. Behroozi, M.; Sami, A. A Multiple-classifier framework for parkinson's disease detection based on various vocal tests. Int. *J. Telemed. Appl.* **2016**, *2016*, 6837498. [CrossRef]

33. Bedi, G.; Cecchi, G.A.; Slezak, D.F.; Carrillo, F.; Sigman, M.; de Wit, H. A window into the intoxicated mind? Speech as an index of psychoactive drug effects. *Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol.* **2014**, *39*, 2340–2348. [CrossRef]

34. Zhao, Z.; Bao, Z.; Zhang, Z.; Deng, J.; Cummins, N.; Wang, H.; Tao, J.; Schuller, B. Automatic assessment of depression from speech via a hierarchical attention transfer network and attention autoencoders. *IEEE J. Sel. Top. Signal Process.* **2020**, *14*, 423–434. [CrossRef]

35. Bedi, G.; Carrillo, F.; Cecchi, G.A.; Slezak, D.F.; Sigman, M.; Mota, N.B.; Ribeiro, S.; Javitt, D.C.; Copelli, M.; Corcoran, C.M. Automated analysis of free speech predicts psychosis onset in high-risk youths. *NPJ Schizophr.* **2015**, *1*, 15030. [CrossRef] [PubMed]

36. Rezaii, N.; Walker, E.; Wolff, P. A machine learning approach to predicting psychosis using semantic density and latent content analysis. *NPJ Schizophr.* **2019**, *5*, 9. [CrossRef] [PubMed]

37. Gavrilescu, M.; Vizireanu, N. Feedforward neural network-based architecture for predicting emotions from speech. *Data* **2019**, *4*, 101. [CrossRef]

38. Goldberg, S.B.; Flemotomos, N.; Martinez, V.R.; Tanana, M.J.; Kuo, P.B.; Pace, B.T.; Villatte, J.L.; Georgiou, P.G.; Van Epps, J.; Imel, Z.E.; et al. Machine learning and natural language processing in psychotherapy research: Alliance as example use case. *J. Couns. Psychol.* **2020**, *67*, 438–448. [CrossRef]

39. Zhang, Y.; Qin, X.; Lin, Y.; Li, Y.; Wang, P.; Zhang, Z.; Li, X. Psychosis speech recognition algorithm based on deep embedded sparse stacked autoencoder and manifold ensemble. *J. Biomed. Eng.* **2021**, *38*, 655–662.

40. Song, I.; Diederich, J. Speech analysis for mental health assessment using support vector machines. In *Mental Health Informatics*; Studies in Computational Intelligence; Springer: Berlin, Germany, 2014; pp. 79–105.

41. Fischer, J.; Hammerschmidt, K. Ultrasonic vocalizations in mouse models for speech and socio-cognitive disorders: Insights into the evolution of vocal communication. *Genes Brain Behav.* **2011**, *10*, 17–27. [CrossRef]

42. Di, Y.; Wang, J.; Li, W.; Zhu, T. Using i-vectors from voice features to identify major depressive disorder. *J. Affective Disord.* **2021**, *288*, 161–166. [CrossRef]

43. McGinnis, E.W.; Anderau, S.P.; Hruschak, J.; Gurchiek, R.D.; Lopez-Duran, N.L.; Fitzgerald, K.; Rosenblum, K.L.; Muzik, M.; McGinnis, R.S. Giving voice to vulnerable children: Machine learning analysis of speech detects anxiety and depression in early childhood. *IEEE J. Biomed. Health Informat.* **2019**, *23*, 2294–2301. [CrossRef]

44. Wang, J.; Zhang, L.; Liu, T.; Pan, W.; Hu, B.; Zhu, T. Acoustic Differences between healthy and depressed people: A cross situation study. *BMC Psychiatry Res. Artic.* **2019**, *8*, 300. [CrossRef]

45. He, L.; Cao, C. Automated depression analysis using convolutional neural networks from speech. *J. Biomed. Inform.* **2018**, *83*, 103–111. [CrossRef]

46. Jenei, A.Z.; Kiss, G. Severity estimation of depression using convolutional neural network. *Period Polytech. Electr. Eng. Comput. sci.* **2021**, *65*, 227–234. [CrossRef]

47. Huang, Y.J.; Lin, Y.T.; Liu, C.C.; Lee, L.E.; Hung, S.H.; Lo, J.K.; Fu, L.C. Assessing schizophrenia patients through linguistic and acoustic features using deep learning techniques. *IEEE Trans. Neural. Syst. Rehabil. Eng.* **2022**, *30*, 947–956. [CrossRef] [PubMed]

48. Xu, W.; Wang, W.; Portanova, J.; Chander, A.; Campbell, A.; Pakhomov, S.; Ben-Zeev, D.; Cohen, T. Fully automated detection of formal thought disorder with Time-series Augmented Representations for Detection of Incoherent Speech (TARDIS). *J. Biomed. Inform.* **2022**, *126*, 103998. [CrossRef] [PubMed]

49. Fisher, D.J.; Labelle, A.; Knott, V.J. Auditory hallucinations and the mismatch negativity: Processing speech and non-speech sounds in schizophrenia. *Int. J. Psychophysiol.* **2008**, *70*, 3–15. [CrossRef]

50. Todd, J.; Michie, P.T.; Schall, U.; Karayanidis, F.; Yabe, H.; Näätänen, R. Deviant matters: Duration, frequency, and intensity deviants reveal different patterns of mismatch negativity reduction in early and late schizophrenia. *Biol. Psychiatry* **2008**, *63*, 58–64. [CrossRef]

51. Arevian, A.C.; Bone, D.; Malandrakis, N.; Martinez, V.R.; Wells, K.B.; Miklowitz, D.J.; Narayanan, S. Clinical state tracking in serious mental illness through computational analysis of speech. *PLoS ONE* **2020**, *15*, e0225695. [CrossRef]