Protocol of the Systematic Literature Review

**A study on researches that utilize machine learning techniques to detect or monitor voice affecting conditions or disorders**

Alper Idrisoglu

V:1.0

2021-12-14

**Blekinge Institute of Technology**

Department of health (TIHA)

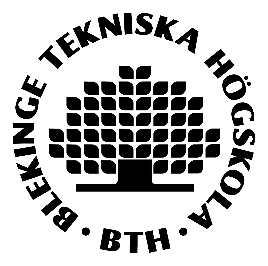


Table of contents

[1. Background 3](#_Toc90280720)

[2 Aim of the research 3](#_Toc90280721)

[2.1 Research Question 3](#_Toc90280722)

[2.2 Search Procedure 4](#_Toc90280723)

[2.3 Study Selection 4](#_Toc90280724)

[Inclusion criteria: 5](#_Toc90280725)

[Exclusion criteria: 5](#_Toc90280726)

[2.4 Study Evaluation 6](#_Toc90280727)

[2.5 Data Extraction 6](#_Toc90280728)

[2.6 Data Synthesis 7](#_Toc90280729)

[References 8](#_Toc90280730)

## Background

The Voice is a tool that enables verbal communication in order to express and share ideas, emotions, thoughts, and sentiments [1]. The airflow created by the pressure in the lungs reaches the vocal folds in the larynx and vibrates the vocal folds which is the basic mechanic of voice creation [2]. Voice production can be affected by various voice disorders such as structural, inflammatory, traumatic, systemic, aerodigestive, psychiatric and psychological, neurological, and functional [3]. Changes in vocal quality are the typical outcome of voice disorders, which not only causes an impaired quality of life (QOL) for individuals but even contribute to high direct health care costs. When we sum up the side costs such as absence from work or loos of jobb, It becomes a large burden for society [1] [4] [5] [6]. Quivering sound, hoarseness, choppy or strained sound, weak sound, changed pitch, and speech quality is some of the symptoms of disordered voice and can be diagnosed by a health care specialist through several examinations and tests [7].

As mentioned in the previous section, the voice or speech can be affected by several conditions and disorders which is a factor that contributes to decreased QOL. Nevertheless, being so sensitive could be beneficial and rise new possibilities for earlier diagnosis of  voice disorders through the use of voice as a biomarker [8]. Earlier diagnosis is a fundamental factor for experiencing a better QOL for people with mild cognitive impairments (MCI) [9] [10]. Existing non-invasive methodologies are not sensitive enough to detect MCI in the early stages, while machine learning (ML) technologies, based on digital biomarkers extracted from voice or speech could offer huge potential for earlier detection. [11] [12].

Traditional biological markers, i.e., biomarkers are currently being used to detect molecular changes associated with a disease and have been integrated with clinical practices for decades [13] [14]. Digital biomarkers are a developing landscape that shares the same objectives as traditional biomarkers in answering health-related questions [15]. However, since the collection of voice and speech is a non-invasive process that can be done at a low cost [16], the voice as a digital biomarker could be a diagnostic and prognostic resource with the potential of being more economically viable, in addition to being a more ecologically measure than many of the currently employed clinical alternatives for the assessment of cognition and function. Digital biomarkers extracted from voice or speech are potential biomarkers that can be used together with machine learning (ML) techniques for the detection or monitoring of several diseases [8]. 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. Two most commonly employed ML techniques in this context are supervised and unsupervised learning. In supervised learning, an ML technique is trained using labelled datasets (training set), and its accuracy is evaluated using unlabelled datasets (validation and test set). The labelled data contains the actual diagnostic information that allows the ML technique to compare its output and adjust its parameters for improved accuracy. Unsupervised learning involves applying clustering methods on training data without labels to group data through one or several clustering algorithms. There are several types of research done to investigate the potential of voice-based digital biomarkers (voice biomarkers) for detecting cognitive impairments such as Dementia, Alzheimer and Parkinson and the results are satisfying [17] [18] [19].

Emerging ML techniques are becoming more commonly used tools to support decision-making in treatment and diagnosis in health care [20]. A short explanation of the working principles of an ML algorithm is the features extracted from voice feed into an ML algorithm to classify disorders or discriminate if the voice is pathologic or normal. Prior studies provide extended information on feature extraction and their application [21] [22] [23]. Encouraging results of ML classifiers with voice biomarkers brings them into the focus of researchers. A meta-analysis on applied ML techniques on voice disorders summarizes used classifiers, respective features, and accuracy [6]. The prior studies show many diverse accuracy achievements which is an indication of demand for further development [24].

Most health problems could benefit from an early diagnosis for better treatment and management. Increased lifetime expectance and an aging population are an indication of the need for extended health care resources in the future, where we can use the benefits of ML techniques on reasonable digital biomarkers to catch health problems before they are visibly remarkable and affect the person's life.

In order to be able to adapt voice-based diagnosis and prognosis into clinical practices, it requires solid evidence and research to clinically validate the usability of voice biomarkers and the performance of ML classifiers. This paper aims to study the global trend of research done in the field of voice-based diagnosis and prognosis with emphasis on ML utilization.

## 2 Aim of the research

The purpose of this systematic literature review (SLR) is to study the utilization of machine learning techniques to detect or monitor voice affecting disorders in recent studies, which are not directly related to the voice production mechanism, but indirectly lead to the deterioration of vocal quality. The SLR targets research done within ten years.

## 2.1 Research Question

“How are machine learning (ML) techniques being utilized for diagnosing voice affecting disorders through voice?” is the central question this SLR strives to find an answer through several sub-questions presented below.

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

## 2.2 Search Procedure

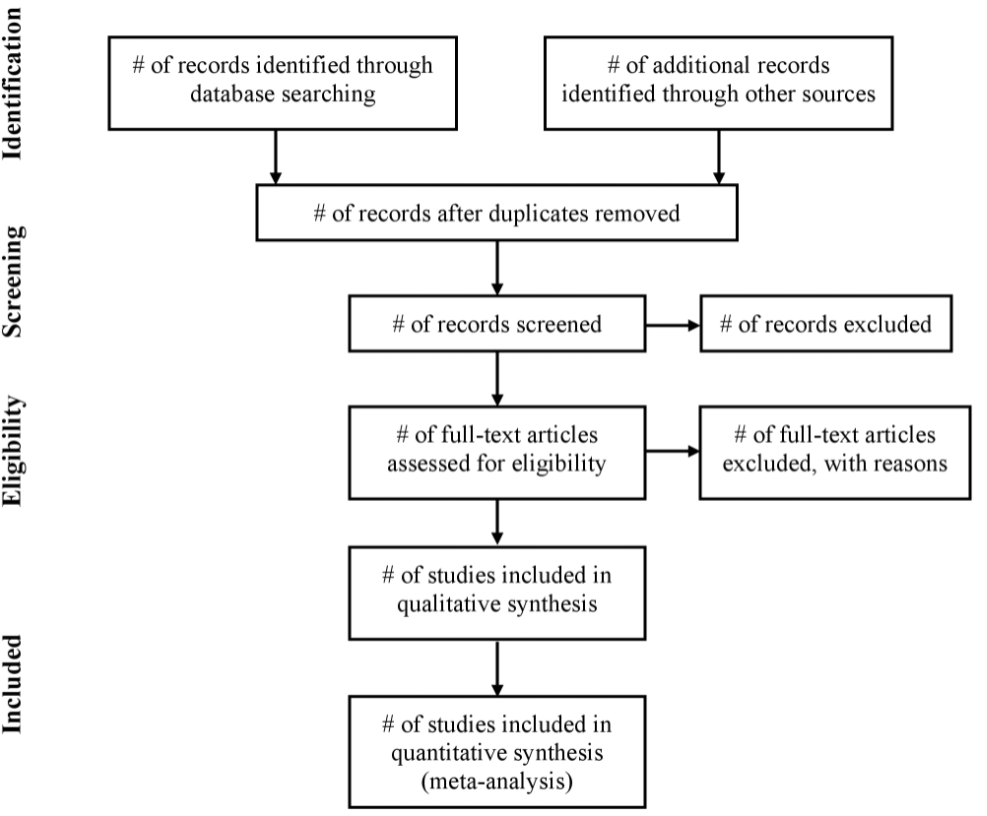
Automated search will be performed on databases PubMed, Scopus, and Web of Science.

By using the PICO approach, a search string based on the central question and sub-questions will be optimized for relevant results [25]. The appliance of the PICO approach is explained below:

* **Population:** Studies that investigate disorders that affect the voice.
* **Intervention:** With the use of ML techniques for diagnosis and/or monitoring of disorders through voice samples.
* **Comparison:** Machine learning techniques.
* **Outcome:** Reported quantities or results such as precision and accuracy.

## 2.3 Study Selection

The most preferred reporting item for SLRs, the PRISMA statement will be used for study selection [26]. The articles obtained from automated search results will be reviewed by more than one reviewer based on the PRISMA tree diagram shown in figure 1.



***Figure 1. PRISMA tree diagram for SLR***

Several database search results from pre-named sources will be stored in the reference management program Zotero [27]. Further on a duplication check and removal will be performed in order to ensure that only exist one example of each article is in the Zotero reference list. The third stage will be the screening of titles and abstracts based on inclusion and exclusion criteria. The articles that pass the third stage will be fully readied and assessed for eligibility in the fourth stage. Whit a quality assessment based on the questions in section 2.4, the fifth process will be completed. Data extraction progress will be executed in the last stage of the PRISMA tree.

## Inclusion criteria:

1. Be a journal study.
2. Be a primary study written in English.
3. Be research published not earlier than 2012.
4. Be research that uses voice as the input data.
5. Be research that employs at least one ML algorithm.
6. Be research that aims to diagnosis and monitor at least one voice-affecting disorders not related to the larynx.

## Exclusion criteria:

1. Be a non-peer reviewed and primary study.
2. Be research written in other languages than English.
3. Be research that uses medical imaging with voice involvement.
4. Be research published before 2012 and after 2022.
5. Be research uses voice as a secondary input.
6. Be research that classifies voice affecting dis- ease without an ML approach.
7. Be research that classifies voice disorders related to the larynx.

## 2.4 Study Selection

In addition to inclusion and exclusion criteria, in order to investigate and prevent the high-quality differences, the articles chosen for full-text reading will be put into quality assessment through the questions in table 1 [28].

Table 1. Quality assessment questionary.

|  |
| --- |
| Questions |
| **General questions** | **Assessment**: **0***/***0***.***5***/***1** |
| 1. Are the aims clearly stated? 2. Is the targeted population described? 3. Is it discussed the contribution of the study? 4. Are gender and age considered? 5. Is/are the technique(s) being implemented clearly described? |  |
| **Data analysis** | **Assessment**: **0***/***0***.***5***/***1** |
| 1. Is the origin of data given? 2. Is the type of data clearly described? 3. Is the data consist of voice recordings? 4. Is the data validation method given? 5. is there a discussion on if the data size can be generalized for the targeted population? |  |
| **Result** | **Assessment**: **0***/***0***.***5***/***1** |
| 1. Is/are the result(s) clearly discussed? 2. Are all aims or questions answered? 3. Was the outcome related to the target population? 4. Are the limitations discussed? 5. Did results compare with previous rapports? |  |
| **Score** |  |

## 2.5 Data Extraction

After quality assessment, the articles will be readied carefully, and data extraction will be cared out by records in table 2.

Table 2. Data attributes will be extracted and explained.

|  |  |
| --- | --- |
| Attribute | Definition |
| ISSN | International standard serial number recorded. |
| Title | Ful title of the research. |
| Journal | Publication venue record. |
| Authors | All author’s names. |
| Publication date | The date of paper published. |
| Publication Type | The type of publication. |
| Origin of publication | Geographical location of first author’s institution. |
| Targeted disorder | Investigated disorder. |
| Database | Source of the data. |
| Origin of data | The geographical location of data sources. |
| The characteristics of the data | Type of voice recordings |
| Additional data | Used additional data except for voice recordings. |
| Datasets | The number of participants. |
| Sample size | The number of recordings. |
| Aim of the study | Purpose of study. |
| Age Range | The considered age range of the participants. |
| Gender | The number of participants differentiated 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 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 best outcome. |

## 2.6 Data Synthesis

* **SubQ.1:** What are the aims of pathologic voice evaluation?

Table 3. Categorisation of studies based the aim of the study.

|  |  |  |
| --- | --- | --- |
| **Nr:** | **Targeted disorder(s)** | **Aim of the study** |
|  |  |  |
|  |  |  |

* **SubQ.2:** Which ML techniques are being employed for the diagnosis and monitoring of VAD through voice and which VADs are being investigated?

Table 4. Targeted disorders and ML techniques.

|  |  |  |
| --- | --- | --- |
| **Nr:** | **Targeted disorder(s)** | **Proposed ML algorithm(s)** |
|  |  |  |
|  |  |  |

* **SubQ.3:** What are the time and geographical trends of publications in the scope of SLR?

Table 5. Investigated disorders by country and date.

|  |  |  |  |
| --- | --- | --- | --- |
| **Nr:** | **Origin of publication** | **Targeted disorder** | **Publication date** |
|  |  |  |  |
|  |  |  |  |

* **SubQ.4:** What are the data characteristics of the sound samples for different disorders and types of studies?

Table 6. Representation of used data characteristics, features, and type of studies.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Nr:** | **Targeted disorder(s)** | **Data Characteristic(s)** | **Feature(s)** | **Database** |
|  |  |  |  |  |
|  |  |  |  |  |

* **SubQ.5:** Are the studies cross-sectional or longitudinal?

Table 7. Summary of type of the studies.

|  |  |  |
| --- | --- | --- |
| **Nr:** | **Targeted disorder(s)** | **Type of the study** |
|  |  |  |
|  |  |  |

* **SubQ.6:** How is performance being evaluated in the studies?

Table 8. Performance summary of each study.

|  |  |  |  |
| --- | --- | --- | --- |
| **Nr:** | **Targeted disorder(s)** | **Proposed ML algorithm(s)** | **Performance** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

## References

[1] R. M. Merrill, N. Roy, and J. Lowe, “Voice-Related Symptoms and Their Effects on Quality of Life,” *Ann. Otol. Rhinol. Laryngol.*, vol. 122, no. 6, pp. 404–411, Jun. 2013, doi: 10.1177/000348941312200610.

[2] Z. Zhang, “Mechanics of human voice production and control,” *J. Acoust. Soc. Am.*, vol. 140, no. 4, pp. 2614–2635, Oct. 2016, doi: 10.1121/1.4964509.

[3] H. Umeno *et al.*, “A summary of the Clinical Practice Guideline for the Diagnosis and Management of Voice Disorders, 2018 in Japan,” *Auris. Nasus. Larynx*, p. 11, 2020.

[4] S. M. Cohen, J. Kim, N. Roy, C. Asche, and M. Courey, “Direct health care costs of laryngeal diseases and disorders,” *The Laryngoscope*, vol. 122, no. 7, pp. 1582–1588, 2012, doi: 10.1002/lary.23189.

[5] S. Misono, S. Marmor, N. Roy, T. Mau, and S. M. Cohen, “Multi-institutional Study of Voice Disorders and Voice Therapy Referral: Report from the CHEER Network,” *Otolaryngol. Neck Surg.*, vol. 155, no. 1, pp. 33–41, Jul. 2016, doi: 10.1177/0194599816639244.

[6] S. A. Syed, M. Rashid, and S. Hussain, “Meta-analysis of voice disorders databases and applied machine learning techniques,” *Math. Biosci. Eng.*, vol. 17, no. 6, pp. 7958–7979, 2020, doi: 10.3934/mbe.2020404.

[7] “Voice Disorders.” https://www.hopkinsmedicine.org/health/conditions-and-diseases/voice-disorders (accessed Dec. 05, 2021).

[8] J. Robin, J. E. Harrison, L. D. Kaufman, F. Rudzicz, W. Simpson, and M. Yancheva, “Evaluation of Speech-Based Digital Biomarkers: Review and Recommendations,” *Digit. Biomark.*, vol. 4, no. 3, pp. 99–108, Oct. 2020, doi: 10.1159/000510820.

[9] M. N. Sabbagh *et al.*, “Rationale for Early Diagnosis of Mild Cognitive Impairment (MCI) supported by Emerging Digital Technologies,” *J. Prev. Alzheimers Dis.*, vol. 7, no. 3, pp. 158–164, Jun. 2020, doi: 10.14283/jpad.2020.19.

[10] J. M. Tracy, Y. Özkanca, D. C. Atkins, and R. Hosseini Ghomi, “Investigating voice as a biomarker: Deep phenotyping methods for early detection of Parkinson’s disease,” *J. Biomed. Inform.*, vol. 104, p. 103362, Apr. 2020, doi: 10.1016/j.jbi.2019.103362.

[11] B. Roark, M. Mitchell, J.-P. Hosom, K. Hollingshead, and J. Kaye, “Spoken Language Derived Measures for Detecting Mild Cognitive Impairment,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 19, no. 7, pp. 2081–2090, Sep. 2011, doi: 10.1109/TASL.2011.2112351.

[12] J. H. Van Stan, D. D. Mehta, and R. E. Hillman, “Recent Innovations in Voice Assessment Expected to Impact the Clinical Management of Voice Disorders,” *Perspect. ASHA Spec. Interest Groups*, vol. 2, no. 3, pp. 4–13, Mar. 2017, doi: 10.1044/persp2.SIG3.4.

[13] R. Mayeux, “Biomarkers: Potential uses and limitations,” vol. 1, no. 2, p. 7, 2004.

[14] R. M. Califf, “Biomarker definitions and their applications,” *Exp. Biol. Med.*, vol. 243, no. 3, pp. 213–221, Feb. 2018, doi: 10.1177/1535370217750088.

[15] E. R. Dorsey, S. Papapetropoulos, M. Xiong, and K. Kieburtz, “The First Frontier: Digital Biomarkers for Neurodegenerative Disorders,” *Digit. Biomark.*, Jul. 2017, doi: 10.1159/000477383.

[16] L. M. Babrak *et al.*, “Traditional and Digital Biomarkers: Two Worlds Apart?,” *Digit. Biomark.*, vol. 3, no. 2, pp. 92–102, Aug. 2019, doi: 10.1159/000502000.

[17] H. Lin *et al.*, “Identification of digital voice biomarkers for cognitive health,” *Explor. Med.*, vol. 1, no. 6, pp. 406–417, Dec. 2020, doi: 10.37349/emed.2020.00028.

[18] K. López-de-Ipiña *et al.*, “On the Selection of Non-Invasive Methods Based on Speech Analysis Oriented to Automatic Alzheimer Disease Diagnosis,” *Sensors*, vol. 13, no. 5, pp. 6730–6745, May 2013, doi: 10.3390/s130506730.

[19] “Investigating voice as a biomarker: Deep phenotyping methods for early detection of Parkinson’s disease - ScienceDirect.” httpswww.sciencedirect.com/science/article/pii/S1532046419302825 (accessed Dec. 10, 2021).

[20] “Recent Innovations in Voice Assessment Expected to Impact the Clinical Mana...: EBSCOhost.” https://web.s.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=2&sid=764c3f56-67fc-430e-ac14-7bb64cebf226%40redis (access://ed Dec. 09, 2021).

[21] Department of Electronics and Communication Engineering, S. G. Balekundri Institute of Technology, Belagavi-India, V. Kanabur, S. S. Harakannanavar, and D. Torse, “An Extensive Review of Feature Extraction Techniques, Challenges and Trends in Automatic Speech Recognition,” *Int. J. Image Graph. Signal Process.*, vol. 11, no. 5, pp. 1–12, May 2019, doi: 10.5815/ijigsp.2019.05.01.

[22] A. Garg and P. Sharma, “Survey on acoustic modeling and feature extraction for speech recognition,” in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, Mar. 2016, pp. 2291–2295.

[23] P. K. Kurzekar, R. R. Deshmukh, V. B. Waghmare, and P. P. Shrishrimal, “A Comparative Study of Feature Extraction Techniques for Speech Recognition System.”

[24] F. T. Al-Dhief *et al.*, “Voice Pathology Detection and Classification by Adopting Online Sequential Extreme Learning Machine,” *Ieee Access*, vol. 9, pp. 77293–77306, 2021, doi: 10.1109/ACCESS.2021.3082565.

[25] A. Butler, H. Hall, and B. Copnell, “A Guide to Writing a Qualitative Systematic Review Protocol to Enhance Evidence-Based Practice in Nursing and Health Care,” *Worldviews Evid. Based Nurs.*, vol. 13, no. 3, pp. 241–249, 2016, doi: 10.1111/wvn.12134.

[26] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and T. P. Group, “Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement,” *PLOS Med.*, vol. 6, no. 7, p. e1000097, Jul. 2009, doi: 10.1371/journal.pmed.1000097.

[27] K. Mueen Ahmed and B. Al Dhubaib, “Zotero: A bibliographic assistant to researcher,” *J. Pharmacol. Pharmacother.*, vol. 2, no. 4, pp. 303–305, Oct. 2011, doi: http://dx.doi.org/10.4103/0976-500X.85940.

[28] B. Kitchenham and S. Charters, “Guidelines for performing Systematic Literature Reviews in Software Engineering.” 2007.