BMJ Open

Voice disorder recognition using machine learning: a scoping review protocol

Rijul Gupta [1], Dhanshree R Gunjawate [2], Duy Duong Nguyen [2], Craig Jin [1], Catherine Madill [2]

To cite: Gupta R, Gunjawate DR, Nguyen DD, *et al*. Voice disorder recognition using machine learning: a scoping review protocol. *BMJ Open* 2024;**14**:e076998. doi:10.1136/bmjopen-2023-076998

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2023-076998).

Received 22 June 2023
Accepted 10 January 2024

[1]School of Electrical and Information Engineering, The University of Sydney Faculty of Engineering and Information Technologies, Sydney, New South Wales, Australia
[2]Sydney School of Health Sciences, The University of Sydney Faculty of Medicine and Health, Sydney, New South Wales, Australia

Correspondence to
Dr Catherine Madill;
cate.madill@sydney.edu.au

ABSTRACT

Introduction Over the past decade, several machine learning (ML) algorithms have been investigated to assess their efficacy in detecting voice disorders. Literature indicates that ML algorithms can detect voice disorders with high accuracy. This suggests that ML has the potential to assist clinicians in the analysis and treatment outcome evaluation of voice disorders. However, despite numerous research studies, none of the algorithms have been sufficiently reliable to be used in clinical settings. Through this review, we aim to identify critical issues that have inhibited the use of ML algorithms in clinical settings by identifying standard audio tasks, acoustic features, processing algorithms and environmental factors that affect the efficacy of those algorithms.

Methods We will search the following databases: Web of Science, Scopus, Compendex, CINAHL, Medline, IEEE Explore and Embase. Our search strategy has been developed with the assistance of the university library staff to accommodate the different syntactical requirements. The literature search will include the period between 2013 and 2023, and will be confined to articles published in English. We will exclude editorials, ongoing studies and working papers. The selection, extraction and analysis of the search data will be conducted using the 'Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for scoping reviews' system. The same system will also be used for the synthesis of the results.

Ethics and dissemination This scoping review does not require ethics approval as the review solely consists of peer-reviewed publications. The findings will be presented in peer-reviewed publications related to voice pathology.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This review focuses on the quality of the dataset, the variability of the audio tasks and the confidence in voice disorder diagnoses, and analyses the effect of input features, environmental conditions, model algorithms and evaluation procedures on the efficacy of detecting voice disorders.
⇒ The search strategy for the review will be set as broad as possible to maximise the coverage of various techniques for voice disorder detection.
⇒ Articles published in languages other than English will be excluded.
⇒ Studies investigating speech disorders with no involvement of voice disorders will be excluded, even though they may be closely related.

INTRODUCTION

Voice disorders are defined as deviations from the voice quality, pitch and loudness from the expected values for someone's age, gender and cultural background.[1–3] Voice disorders affect a significant portion of the population, with approximately 30% of individuals reporting having suffered from a voice disorder at some point in their lifetime and 20% experiencing chronic issues.[4] Voice disorders can range from minor vocal discomfort to severe dysphonia and from functional to malignant conditions.[5] Voice disorders are diagnosed using a range of different processes including case history questionnaires and interviews, patient-reported outcome measures, auditory–perceptual judgements, acoustic analysis and laryngoscopy. In cases of malignancy suspicion, biopsy of the larynx may be required to establish a diagnosis. A proportion of voice-disordered patients undergo late or inaccurate diagnosis[6 7] that further impacts the efficacy of treatment and increases the costs of healthcare.[8] It has been suggested that AI/machine learning (ML) is useful in assisting clinicians with early and more accurate detection of voice disorders.[9 10] During a typical visit to a voice clinic, patients are requested to perform vocal tasks such as vocalisation of vowels, reading out written passages and engaging in conversation with the clinician; these interactions are often recorded along with the patients' demographic data and are collated into private[11–15] and publicly available datasets.[16–19] ML requires voice databases as a source to train its algorithms.[20–23]

The availability of these datasets has allowed researchers to explore the application of ML in differentiating voice disorders from non-voice disorders. Several different features have been used as input data within the various ML algorithms. Some of these features include fundamental frequency (F0), perturbation measures (jitter and shimmer), the harmonics-to-noise ratio,[20–23] Mel frequency cepstrum coefficients,[24–28] wavelet sub-band features,[29–32] Multi Dimensional Voice Program parameters[33–35] and glottal flow estimation parameters.[14 36 37] Several ML algorithms have been explored, such as support vector machines,[34 37–39] deep neural networks (DNNs),[28 40–42] k-nearest neighbours[12 43 44] and others.[23 33 45–47]

Despite numerous studies investigating voice disorder recognition, it remains unclear how much progress has been made towards achieving the successful incorporation of ML into the clinical practice of diagnosing and treating voice disorders. More significantly, the motivation and requirements for the next steps to achieve progress in this direction remain unclear. It is unclear which audio tasks, demographic data points, input acoustic features or ML algorithms best detect any specific pathology or collection of pathologies, which further hinders future work as we do not know how to benchmark new work. Therefore, we plan to conduct a scoping review to address these issues. A preliminary search has been conducted to verify that no similar reviews are underway, and previously conducted surveys and reviews do not provide answers to all our questions.

Reviewing seven identified previous works, in[48] only published journal articles that used supervised ML for the binary classification of healthy versus pathological voice was included leading them to review only 13 studies[49]; is a non-systematic review and only has data on ML systems used for voice disorder recognition up to 2018[50]; only includes articles that use Saarbruecken Voice Database (SVD),[17] Massachusetts Eye and Ear Infirmary (MEEI)[16] and Arabic Voice Pathology Database (AVPD)[18] as databases and only work with voice filtering and segmentation techniques[51 52]; only review a subset of the literature with no identifiable search strategy[53]; focuses on the use of internet of things technologies to augment ML systems and does not systematically review the literature on ML systems used for voice disorder recognition[54]; studies the management of voice disorders using voice analyser applications but does not review voice disorders recognition systems.

The aim of this review is to identify the key obstacles that have prevented the widespread adoption of ML algorithms in voice disorder recognition in clinical practice by examining the methodologies, processes and outcomes reported in the literature.

Review questions

This review addresses several new questions regarding the application of ML in voice disorder recognition. These questions include:

1. Which vocal tasks have been used for ML application in voice disorder detection?
2. What is the reported impact of the recording environmental factors (such as acoustic characteristics of the recording studio, ambient noise levels) on the audio recordings?
3. What patient demographic characteristics and case history data have been used for ML for voice disorder detection?
4. What is the frequency of different voice/laryngeal diseases and demographics within the datasets used in previous studies?

Additional questions are included to address the testing methodology, availability of experimental codes, existence of baseline model implementations and the comparison between the various input features used in existing ML systems. A full list of questions that are intended to be addressed by this review are provided in online supplemental appendix I.

METHODS

This review will follow all the points mentioned under the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for scoping reviews (PRISMA-ScR)[55] guidelines for conducting the review and use the Joanna Briggs Institute (JBI) scoping review template to present the results. A version of PRISMA-ScR as it applies to this protocol document has been added in online supplemental appendix II.

Eligibility criteria

Participants

The review will include studies with participants of any language, region, gender, age or ethnicity. There will be no exclusion criteria related to demographics.

Concept

The review will include articles that explore the application of ML algorithms for screening, identification and assessment of voice disorders.

Types of sources

The review will include conference proceedings and articles published in peer-reviewed journals.

Search strategy

We have developed a search strategy based on previous literature. Search terms were used to develop a search strategy with the consensus of all the authors and assistance from multiple librarians at the university's library. The search strategy has been further modified to accommodate the syntactical requirements of the different databases, as indicated in the search strategy presented in online supplemental appendix III. This process involved breaking words into subwords (eg, psych* voice disorder*), using appropriate wildcards and keywords (eg, (MH "Classification+")), as needed by the different databases.

Only studies published in English between the years 2013 and 2023 will be included. We believe this should capture all of the relevant ML techniques and algorithms used for voice disorder detection, as there has

 Gupta R, *et al. BMJ Open* 2024;**14**:e076998. doi:10.1136/bmjopen-2023-076998

been an explosion in ML research and techniques since AlexNet[56] demonstrated a 10% improvement in classification performance by using DNNs in the ImageNet 2012 (ILSVRC2012) competition.[57]

The search will be conducted using the Web of Science, Scopus, Compendex, CINAHL, Medline, IEEE Explore and Embase databases. Only peer-reviewed literature will be included in this review.

Study/source of evidence selection

Following the literature search, all of the bibliographic data will be exported from the associated databases and added to the online literature review tool 'Covidence'.[58] We will remove duplicates using the Covidence automated duplicate detection system. Two authors will then screen the title and abstract of all the articles. The first author will screen all the articles and the rest of the screening work will be divided among the other authors. Screening conflicts will be resolved via group discussion among authors.

The articles selected for full-text review will similarly be assessed by two authors. Again, the first author will review all of the articles with the remaining work to be divided among the other authors. Conflicts will again be resolved via group discussion.

The results of the search and inclusion process will be presented using a PRISMA-ScR flow diagram.[55]

Data extraction

The agreed list of included articles will be uploaded onto the Zotero reference managing system. Data extraction will be conducted by all authors. The data will be stored and analysed using custom programs (online supplemental appendix IV) and Excel software package.

The data to be extracted will include all of the following:

- Vocal tasks
- Recording equipment details
  - Microphone(s) and their specifications
  - Recording software
  - Sampling rate of recording
  - Bit rate of recording
  - Number of channels in the recording
- Acoustic characteristics of the recording environment (eg, ambient noise levels)
- Demographics of participants
- Patient history
- Participant self-reported symptoms
- Methods for voice disorder diagnosis
- Confidence level of the diagnosis
- Disease/disorder name and type
- Inclusion of control groups
- Whether data were collected in a single session or across multiple sessions
- Whether the training and testing datasets were cleanly separated
- The details and parameters used in the algorithms
- The availability of the ML code used for the analysis
- Evaluation metrics
- A description of ablation studies done on the models
- A description of comparison baseline models
- The usage of demographics and symptom data within the algorithms
- Whether the classification is broad spectrum (normal vs pathological) or granular (dysphonia vs nodules vs normal)
- Hardware requirements for running the algorithms
- Description of out-of-domain testing

Data analysis

We will perform analysis using the data extracted from the literature and present the results in the form of descriptive statistics and frequencies. The search analysis data will be made available as a table of information that can be easily searched to answer further questions. The results will be presented in the form of tables and graphs. Quantitative findings will be presented in the form of frequencies and percentages. We will present possible future research directions to improve on the existing methods so that they can be applied in clinical practice.

Planned dates

We plan to conduct this review between January and July 2024.

Twitter Rijul Gupta @rijulg, Dhanshree R Gunjawate @DhanshreeRG_PhD, Duy Duong Nguyen @VoiceLab_Usyd and Catherine Madill @Cate_Madill

Acknowledgements We received support from Doctor Liang Voice Program. Tess Aitken, Isabelle Raisin and Vanah Montgomery from The University of Sydney Library assisted us with refining the search terms. This work contributes towards a PhD in Engineering for the author RG.

Contributors CM proposed the study, which was then developed by RG and DRG with the help of all authors. RG identified the search terms and established the search strategy with DRG and the university library. RG and DRG are responsible for data collection. RG will screen and analyse all the articles and the work will be divided among the rest of the authors for secondary screening and analysis. RG prepared the first two drafts of the manuscript. The final manuscript was prepared by iteratively refining through group discussions between RG, DRG, DDN, CJ and CM.

Funding The research was funded by Doctor Liang Voice Program (Grant number: NA) and the Faculty of Engineering Research Scholarship (Grant number: NA).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs
Rijul Gupta http://orcid.org/0000-0003-3356-3531
Dhanshree R Gunjawate http://orcid.org/0000-0002-2464-2580
Duy Duong Nguyen http://orcid.org/0000-0001-8097-8938
Craig Jin http://orcid.org/0000-0003-4636-753X
Catherine Madill http://orcid.org/0000-0001-8114-1427

REFERENCES

1 Aronson AE, Bless DM. *Clinical Voice Disorders*. 4th ed. New York: Thieme, 2009: 301.
2 Boone DR, McFarlane SC, Von Berg SL, *et al*. *The Voice and Voice Therapy*. Ninth edition. Boston: Pearson, 2014: 636.
3 Lee L, Stemple JC, Glaze L, *et al*. Quick screen for voice and supplementary documents for identifying pediatric voice disorders. *Lang Speech Hear Serv Sch* 2004;35:308–19.
4 Roy N, Merrill RM, Gray SD, *et al*. Voice disorders in the general population: prevalence risk factors, and occupational impact. *The Laryngoscope* 2005;115:1988–95.
5 Koufman JA, Isaacson G. The spectrum of vocal dysfunction. *Otolaryngol Clin North Am* 1991;24:985–8.
6 Paul BC, Chen S, Sridharan S, *et al*. Diagnostic accuracy of history, laryngoscopy, and stroboscopy. *Laryngoscope* 2013;123:215–9.
7 Cohen SM, Kim J, Roy N, *et al*. Factors influencing referral of patients with voice disorders from primary care to otolaryngology. *Laryngoscope* 2014;124:214–20.
8 Cohen SM, Kim J, Roy N, *et al*. Delayed otolaryngology referral for voice disorders increases health care costs. *Am J Med* 2015;128:426.
9 Hu H-C, Chang S-Y, Wang C-H, *et al*. Deep learning application for vocal fold disease prediction through voice recognition: preliminary development study. *J Med Internet Res* 2021;23:e25247.
10 Fang S-H, Tsao Y, Hsiao M-J, *et al*. Detection of pathological voice using cepstrum vectors: a deep learning approach. *J Voice* 2019;33:634–41.
11 Rani K U, Holi MS. Wavelet transform features to hybrid classifier for detection of neurological-disordered voices. *J Clin Eng* 2017;42:89–98.
12 Benba A, Jilbab A, Hammouch A. Voice assessments for detecting patients with neurological diseases using PCA and NPCA. *Int J Speech Technol* 2017;20:673–83.
13 Wang Z, Yu P, Yan N, *et al*. Automatic assessment of pathological voice quality using multidimensional acoustic analysis based on the grbas scale. *J Sign Process Syst* 2016;82:241–51.
14 Forero M LA, Kohler M, Vellasco MMBR, *et al*. Analysis and classification of voice pathologies using glottal signal parameters. *J Voice* 2016;30:549–56.
15 Vásquez-Correa JC, Arias-Vergara T, Orozco-Arroyave JR, *et al*. Automatic detection of Parkinson's disease from continuous speech recorded in non-controlled noise conditions. Interspeech 2015; ISCA, 2015:105–9
16 Massachusetts Eye and Ear Infirmary. *Voice disorders database, version. 1.03 (cd-rom)*. Lincoln Park, NJ: Kay Elemetrics Corporation.
17 Woldert-Jokisz B. Saarbruecken voice database. 2007.
18 Mesallam TA, Farahat M, Malki KH, *et al*. Development of the arabic voice pathology database and its evaluation by using speech features and machine learning algorithms. *J Healthc Eng* 2017;2017:8783751.
19 Cesari U, De Pietro G, Marciano E, *et al*. A new database of healthy and pathological voices. Comput Electr Eng 2018;68:310–21.
20 Verde L, De Pietro G, Sannino G. Voice disorder identification by using machine learning techniques. *IEEE Access* 2018;6:16246–55.
21 Cesari U, De Pietro G, Marciano E, *et al*. Voice disorder detection via an m-health system: Design and results of a clinical study to evaluate vox4health. *Biomed Res Int* 2018;2018:8193694.
22 Teixeira JP, Fernandes PO, Alves N. Vocal acoustic analysis – classification of dysphonic voices with artificial neural networks. Procedia Comput Sci 2017;121:19–26.
23 Moran RJ, Reilly RB, de Chazal P, *et al*. Telephony-based voice pathology assessment using automated speech analysis. *IEEE Trans Biomed Eng* 2006;53:468–77.
24 Vavrek L, Hires M, Kumar D, *et al*. Deep Convolutional neural network for detection of pathological speech. 2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI); Herl'any, Slovakia. IEEE, August 18, 2021:000245–50
25 Milani MGM, Ramashini M, Krishani M. A real-time application to detect human voice disorders. 2020 International Conference on Decision Aid Sciences and Application (DASA); Sakheer, Bahrain.2020
26 Guan H, Lerch A. Learning strategies for voice disorder detection. 2019 IEEE 13th International Conference on Semantic Computing (ICSC); Newport Beach, CA, USA.2019
27 Takashima Y, Takiguchi T, Ariki Y. End-to-end Dysarthric speech recognition using multiple databases. ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); Brighton, United Kingdom. IEEE, October 3, 2019:6395–9
28 Ghulam M, Alhamid MF, Alsulaiman M, *et al*. Edge computing with cloud for voice disorder assessment and treatment. *IEEE Commun Mag* 2018.
29 Fonseca ES, Guido RC, Silvestre AC, *et al*. Discrete Wavelet transform and support vector machine applied to pathological voice signals identification. Seventh IEEE International Symposium on Multimedia (ISM'05); Irvine, CA, USA. IEEE, November 2005:785–9
30 Nayak J, Bhat PS, Acharya R, *et al*. Classification and analysis of speech abnormalities. *ITBM-RBM* 2005;26:319–27.
31 Behroozmand R, Almasganj F. Optimal selection of wavelet-packet-based features using genetic algorithm in pathological assessment of patients' speech signal with unilateral vocal fold paralysis. *Comput Biol Med* 2007;37:474–85.
32 Saidi P, Almasganj F. Voice disorder signal classification using m-band wavelets and support vector machine. *Circuits Syst Signal Process* 2015;34:2727–38.
33 Ali Z, Elamvazuthi I, Alsulaiman M, *et al*. Automatic voice pathology detection with running speech by using estimation of auditory spectrum and cepstral coefficients based on the all-pole model. *J Voice* 2016;30:757.
34 Al-Nasheri A, Muhammad G, Alsulaiman M, *et al*. An investigation of multidimensional voice program parameters in three different databases for voice pathology detection and classification. *J Voice* 2017;31:113.
35 Dibazar AA, Narayanan S, Berger TW. Feature analysis for automatic detection of pathological speech. Second Joint EMBS-BMES Conference 2002 24th Annual International Conference of the Engineering in Medicine and Biology Society. Annual Fall Meeting of the Biomedical Engineering Society; Houston, TX, USA. IEEE, November 2002:182–3
36 Ben Aicha A, Ezzine K. Cancer Larynx detection using glottal flow parameters and statistical tools. 2016 International Symposium on Signal, Image, Video and Communications (ISIVC); Tunis, Tunisia. IEEE, November 23, 2016:65–70
37 Muhammad G, Alsulaiman M, Ali Z, *et al*. Voice pathology detection using interlaced derivative pattern on glottal source excitation. Biomed Signal Process Control 2017;31:156–64.
38 Al-Nasheri A, Muhammad G, Alsulaiman M, *et al*. Investigation of voice pathology detection and classification on different frequency regions using correlation functions. *J Voice* 2017;31:3–15.
39 Muhammad G, Altuwaijri G, Alsulaiman M, *et al*. Automatic voice pathology detection and classification using vocal tract area irregularity. Biocybern Biomed Eng 2016;36:309–17.
40 Alhussein MA, Muhammad G. Voice Pathology Detection Using Deep Learning on Mobile Healthcare Framework. *IEEE Access* 2018;6:41034–41.
41 Gumelar AB, Yuniarno EM, Anggraeni W, *et al*. Enhancing detection of pathological voice disorder based on deep VGG-16 CNN. 2020 3rd International Conference on Biomedical Engineering (IBIOMED); Yogyakarta, Indonesia. IEEE, 2020
42 Al-Dhief FT, Baki MM, Latiff NMA, *et al*. Voice pathology detection and classification by adopting online sequential extreme learning machine. *IEEE Access* 2021;9:77293–306.
43 Alemami Y, Almazaydeh L. Pathological voice signal analysis using machine learning based approaches. *CIS* 2018;11:8.
44 López-de-Ipiña K, Solé-Casals J, Eguiraun H, *et al*. Feature selection for spontaneous speech analysis to aid in Alzheimer's disease diagnosis: a fractal dimension approach. Comput Speech Lang 2015;30:43–60.
45 Cordeiro H, Fonseca J, Meneses C. Spectral envelope and periodic component in classification trees for pathological voice diagnostic. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Chicago, IL: IEEE, 2014:4607–10 Available: http://ieeexplore.ieee.org/document/6944650/
46 Saldanha JC, Ananthakrishna T, Pinto R. Vocal fold pathology assessment using PCA and LDA. 2013 International Conference on Intelligent Systems and Signal Processing (ISSP); Gujarat. IEEE, November 23, 2013:140–4
47 Arias-Londoño JD, Godino-Llorente JI, Sáenz-Lechón N, *et al*. An improved method for voice pathology detection by means of a HMM-based feature space transformation. *Pattern Recognition* 2010;43:3100–12.

48 Al-Hussain G, Shuweihdi F, Alali H, *et al*. The effectiveness of supervised machine learning in screening and diagnosing voice disorders: Systematic review and meta-analysis. *J Med Internet Res* 2022;24:e38472.
49 Hegde S, Shetty S, Rai S, *et al*. A survey on machine learning approaches for automatic detection of voice disorders. *J Voice* 2019;33:947.
50 Syed S, Rashid M, Hussain S. Meta-analysis of voice disorders databases and applied machine learning techniques. *Math Biosci Eng MBE* 2020;17:7958–79.
51 Selvakumari N, Radha V. A survey on optimization techniques in voice disorder classification. Available: https://www.semanticscholar.org/paper/A-Survey-on-Optimization-Techniques-in-Voice-Selvakumari-Radha/fc3a0188628d5339cbc5524c85ef2e0a4ff82ac1 [Accessed 24 Aug 2022].
52 Abdulmajeed NQ, Al-Khateeb B, Mohammed MA. A review on voice pathology: taxonomy, diagnosis, medical procedures and detection techniques, open challenges, limitations, and recommendations for future directions. J Intell Syst 2022;31:855–75.
53 Al-Dhief FT, Latiff NMA, Malik NNNAbd, *et al*. A survey of voice pathology surveillance systems based on internet of things and machine learning algorithms. *IEEE Access* 2020;8:64514–33.
54 Munnings AJ. The current state and future possibilities of mobile phone “voice analyser” applications, in relation to otorhinolaryngology. *J Voice* 2020;34:527–32.
55 PRISMA. Extension for scoping reviews (PRISMA-SCR): checklist and explanation | annals of internal medicine. Available: https://www.acpjournals.org/doi/10.7326/M18-0850 [Accessed 3 May 2023].
56 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2017;60:84–90.
57 Russakovsky O, Deng J, Su H, *et al*. ImageNet large scale visual recognition challenge. *Int J Comput Vis* 2015;115:211–52.
58 Covidence systematic review software. Melbourne, Australia Veritas Health Innovation; Available: www.covidence.org