Artificial Intelligence and Machine Learning for Assessing Voice Handicap Index

Abstract:
Voice disorders can significantly impact an individual's quality of life, yet traditional diagnosis often relies on subjective evaluations and access to clinical experts. In this study, we explore the feasibility of using machine learning (ML) regression models to predict Voice Handicap Index (VHI) scores based on voice recordings, aiming to provide an accessible, automated assessment tool. Using the publicly available VOICED dataset from PhysioNet, we preprocessed the acoustic data, extracted relevant features, and trained multiple regression models. The Gradient Boosting Regressor and Random Forest Regressor showed strong performance, with Mean Absolute Error (MAE) values under 10 on average. Our findings suggest that ML-based prediction of VHI scores is not only possible but also promising in supplementing clinical diagnostics.

Introduction

Voice disorders, encompassing a range of pathological conditions that impair the quality, pitch, loudness, or flexibility of the human voice, affect millions of individuals worldwide [1][2][3]. According to the American Anthropological Association, more than 170 million people suffer from some kind of voice disorder worldwide[4]. These disorders can arise from a multitude of causes—including vocal fold paralysis, nodules, polyps, neurological conditions such as Parkinson's disease, or functional misuse of the voice—and they often result in profound personal, social, and occupational burdens. According to epidemiological data, voice disorders are prevalent in approximately 7% of the global population at any given time[5], with higher rates observed among professional voice users such as teachers, singers, and broadcasters. It is estimated that 37.7% of teachers and 46% of call center workers are affiliated

[1] Alhefdhi, H. *et al.* Assessing Prevalence, Risk Factors, and Occupational Impact of Voice Disorders Among Teachers: A Population-Based Survey in Aseer Region, Saudi Arabia. *The Egyptian Journal of Otolaryngology* **41**, (2025).

[2] Baghban, K., Golmohammadi, G. & Asadollahpour, F. The Worldwide Prevalence of Voice Disorders Among Schoolteachers: A Systematic Review and Meta-Analysis. *Journal of Voice* (2025) doi:https://doi.org/10.1016/j.jvoice.2025.04.018.

[3] Rubio-Carbonero, G. Communication in Persons with Acquired Speech Impairment: The Role of Family as Language Brokers. *Journal of Linguistic Anthropology* **32**, 161–181 (2021).

[4] Baghban, K., Golmohammadi, G. & Asadollahpour, F. The Worldwide Prevalence of Voice Disorders Among Schoolteachers: A Systematic Review and Meta-Analysis. *Journal of Voice* (2025) doi:https://doi.org/10.1016/j.jvoice.2025.04.018.

[5] GRAVEL, J. S. *et al.* Pediatric Communication Disorders. *Pediatric Otolaryngology* 29–59 (2007) doi:https://doi.org/10.1016/b978-0-323-04855-2.50008-3.

with certain vocal disorder [6][7] . Despite the widespread occurrence and substantial societal cost—manifesting as reduced workplace productivity, increased healthcare utilization, and diminished quality of life—voice disorders remain underdiagnosed and undertreated[8][9].

Traditional diagnostic procedures for voice disorders rely heavily on laryngoscopic examination and perceptual voice assessments, which are both time-intensive and dependent on clinician expertise[10][11]. This often results in significant diagnostic delays, especially in under-resourced healthcare settings. In recent years, advances in artificial intelligence (AI) and machine learning (ML) have demonstrated considerable potential in medical diagnostics by uncovering subtle patterns in large, complex datasets that may elude human observers. While previous research has successfully applied AI-based techniques to the classification of brain disorders, dermatological lesions, and radiological findings[12][13][14][15], the diagnostic use of AI in voice pathology remains an emerging and underexplored field.

This study aims to bridge that gap by investigating the feasibility of using machine learning regression models to predict Voice Handicap Index (VHI) scores from acoustic features derived from sustained phonation recordings. Leveraging a publicly available dataset from PhysioNet, this research contributes to a growing body of evidence suggesting that vocal biomarkers—when paired with robust computational techniques—can provide objective, scalable tools for early screening and monitoring of voice disorders. Such technologies may ultimately enhance diagnostic accuracy, reduce clinical burden, and improve patient outcomes in voice care.

---

[6] Baghban, K., Golmohammadi, G. & Asadollahpour, F. The Worldwide Prevalence of Voice Disorders Among Schoolteachers: A Systematic Review and Meta-Analysis. *Journal of Voice* (2025) doi:https://doi.org/10.1016/j.jvoice.2025.04.018.

[7] Alhefdhi, H. *et al.* Assessing Prevalence, Risk Factors, and Occupational Impact of Voice Disorders Among Teachers: A Population-Based Survey in Aseer Region, Saudi Arabia. *The Egyptian Journal of Otolaryngology* **41**, (2025).

[8] GRAVEL, J. S. *et al.* Pediatric Communication Disorders. *Pediatric Otolaryngology* 29–59 (2007) doi:https://doi.org/10.1016/b978-0-323-04855-2.50008-3.

[9] Awaji, A. *et al.* Measuring Perceived Voice Disorders and Quality of Life among Female University Teaching Faculty. *Journal of Personalized Medicine* **13**, 1568–1568 (2023).

[10] Basu, S. Laryngoscopy: Procedure, Risks And Results. *Netmeds* https://www.netmeds.com/health-library/post/laryngoscopy-procedure-risks-and-results?srsltid=AfmBOore5iEdd9MmoK1AncsJWjeLI2276Jm_u_0Zfq8Tg7tLV9PU3Ql- (2023).

[11] Cleveland Clinic. Laryngoscopy: Procedure, Definition & Types. *Cleveland Clinic* https://my.clevelandclinic.org/health/diagnostics/22803-laryngoscopy (2022).

[12] Pankaj Dipankar, Salazar, D., Dennard, E., Shanid Mohiyuddin & Nguyen, Q. C. Artificial intelligence based advancements in nanomedicine for brain disorder management: an updated narrative review. *Frontiers in Medicine* **12**, (2025).

[13] Yao, Z. *et al.* Artificial intelligence-based diagnosis of Alzheimer's disease with brain MRI images. *European Journal of Radiology* **165**, 110934 (2023).

[14] England, N. NHS England» AI based skin lesion analysis technology. *England.nhs.uk* https://www.england.nhs.uk/elective-care/best-practice-solutions/ai-based-skin-lesion-analysis-technology/ (2023).

[15] Najjar, R. Redefining Radiology: A Review of Artificial Intelligence Integration in Medical Imaging. *Diagnostics* **13**, 2760 (2023).

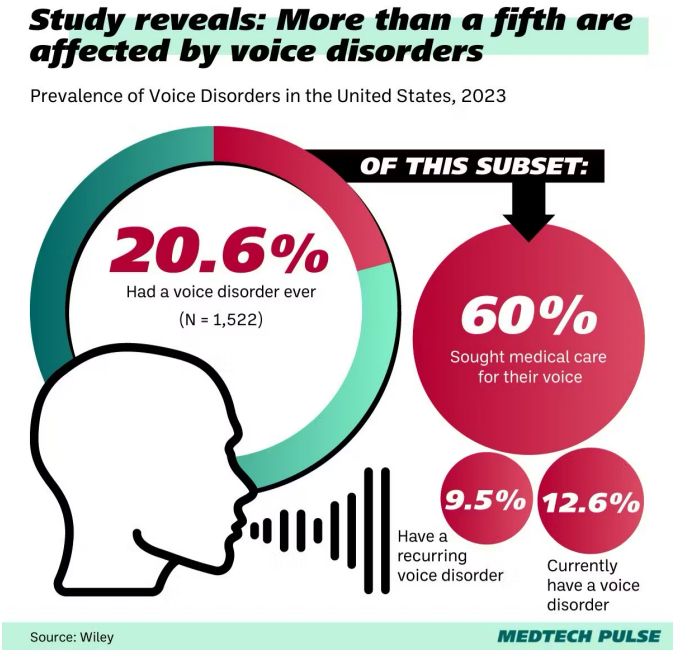

Fig 1: Worldwide Prevalence of Voice Disorders[16]

Literature Review

Voice-based machine learning applications have gained momentum in recent years. Early studies primarily focused on classifying voice disorders using signal processing and support vector machines (SVM). For instance, researchers have attempted to differentiate between healthy and pathological voices using Mel-frequency cepstral coefficients (MFCCs), jitter, shimmer, and harmonic-to-noise ratios[17][18].

In parallel, the concept of predicting health-related scores through regression has become increasingly prevalent. Regression models have been applied to predict Parkinson's disease severity from vocal features, and some deep learning models have been trained on speech to predict emotional distress or depression levels[19]. These examples suggest that the voice can serve as a rich, non-invasive biomarker for health conditions.

The VOICED dataset released on PhysioNet in 2022 has facilitated new avenues of voice disorder research[20]. It includes over 2000 annotated samples

---

[16] This throat patch speaks for patients with voice disorders. *Medtechpulse.com* https://www.medtechpulse.com/article/insight/throat-patch-speaks-for-patients-with-voice-disorders (2024).

[17] Rubio-Carbonero, G. Communication in Persons with Acquired Speech Impairment: The Role of Family as Language Brokers. *Journal of Linguistic Anthropology* **32**, 161–181 (2021).

[18] Yao, P. *et al.* Applications of Artificial Intelligence to Office Laryngoscopy: A Scoping Review. *Laryngoscope* **132**, 1993–2016 (2021).

[19] Reddy, A. *et al.* Artificial Intelligence in Parkinson's Disease: Early Detection and Diagnostic Advancements. *Ageing Research Reviews* **99**, 102410–102410 (2024).

[20] Verde, L. & Sannino, G. VOICED Database. *Physionet.org.* https://physionet.org./content/voiced/1.0.0/

from individuals with and without vocal disorders, along with their VHI scores. However, few studies have leveraged this dataset for direct regression analysis on VHI scores. This gap motivates our current study.

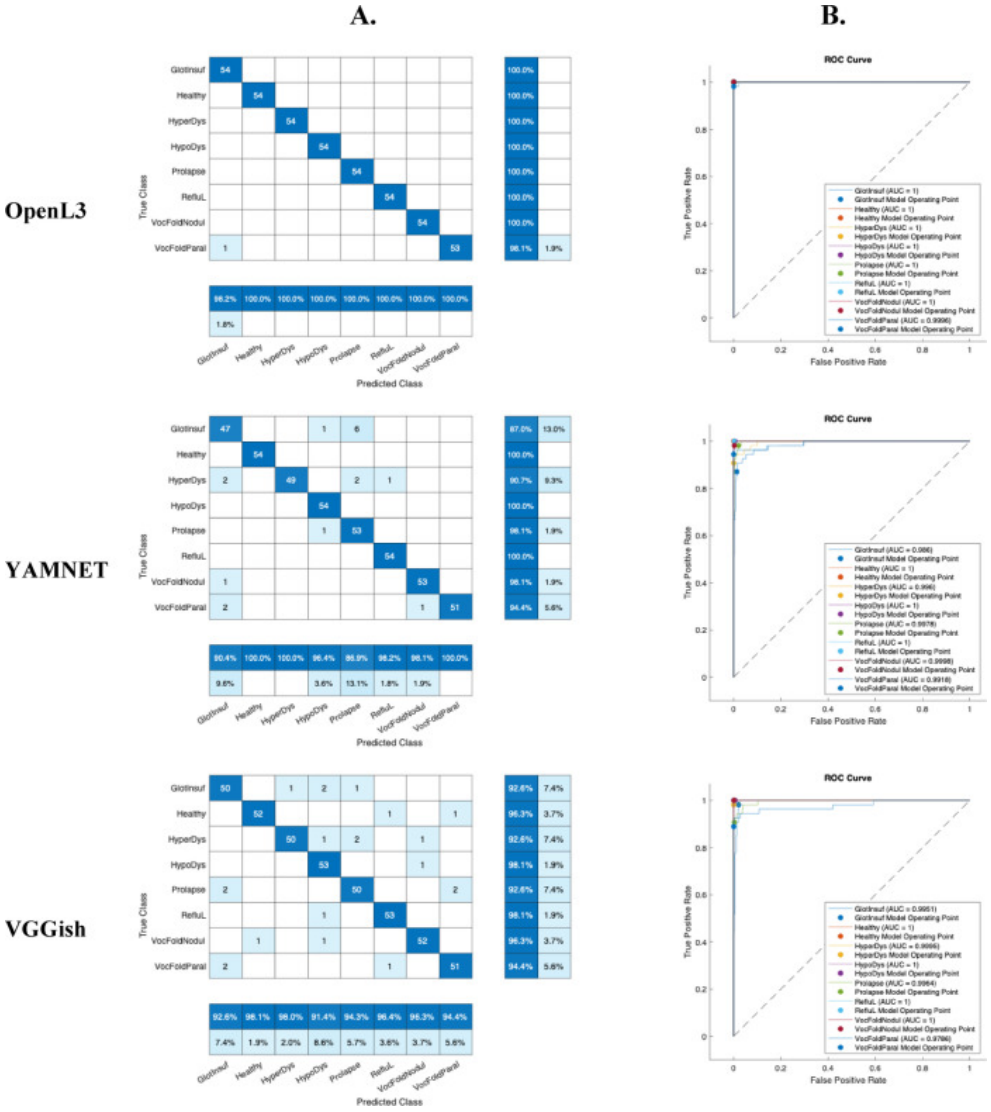

Fig 2: Results for voice classification based on fine-tuning of OpenL3, YAMNET and VGGish. These results concern a single test without cross-validation. Eight classes are presented: Glottic Insufficiency, Hyperkinetic Dysphonia, Hypokinetic Dysphonia, Prolapse, Reflux Laryngitis, Vocal Fold Nodules, Vocal Fold Paralysis and Healthy. (**A**) The confusion matrices show the actual classes (rows) and the predicted classes (columns). The diagonal cells show the correctly classified observations. The measures shown at the bottom, in dark blue, are called precision. The measures on the right, shown in dark blue are called recall or sensitivity (false negative rates are in light blue). (**B**) ROC curves (different colours) and AUC values for the eight classes[21].

(2018).

[21] Fatma Özcan. Differentiability of voice disorders through explainable AI. *Scientific Reports* **15**, (2025).

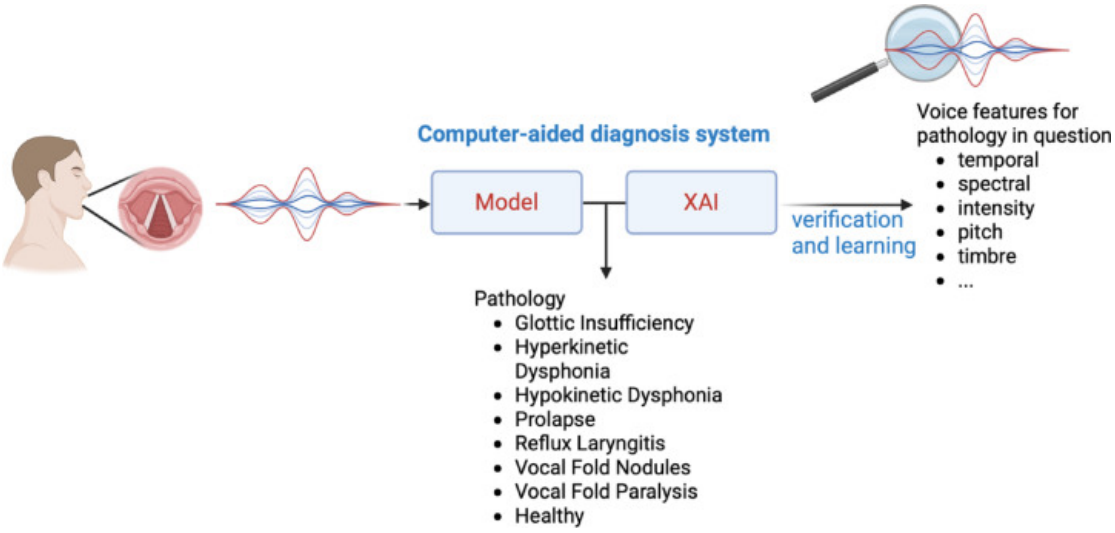

Fig3: Principle of voice pathology detection and differentiation. Created in BioRender[22].

Methodology

1. Data Source and Description

This research utilized the VOICED dataset, a publicly available resource provided by PhysioNet, which contains voice recordings and corresponding Voice Handicap Index (VHI) scores. The dataset includes both healthy individuals and those diagnosed with various voice disorders, such as muscle tension dysphonia or vocal fold paralysis. Each entry consists of multiple sustained vowel phonations (e.g., /a/, /i/, /u/), recorded under standardized conditions. The VHI, a validated self-assessment tool, reflects the patient's perception of their voice disability on a scale from 0 to 120, making it a suitable ground truth for regression-based prediction models.

2. Preprocessing

To ensure the reliability and usability of the dataset, several preprocessing steps were undertaken. First, the audio recordings were trimmed to remove silence and normalized for consistent loudness. Then, acoustic features were extracted using openSMILE, a popular open-source toolkit for audio signal analysis. Key features included Mel-Frequency Cepstral Coefficients (MFCCs), jitter, shimmer, and Harmonics-to-Noise Ratio (HNR), which are known to correlate with vocal pathologies. The extracted features were then standardized to have zero mean and unit variance for optimal model training.

3. Model Training and Validation

The primary objective was to predict the overall VHI score using a regression model based on the extracted audio features. Several machine learning algorithms were considered, including linear regression, support vector

[22] Fatma Özcan. Differentiability of voice disorders through explainable AI. *Scientific Reports* **15**, (2025).

regression (SVR), random forest, and gradient boosting. After initial performance evaluation, the gradient boosting regression model (specifically, XGBoost) was selected due to its superior accuracy and robustness. The dataset was split into training (80%) and testing (20%) subsets. Hyperparameter tuning was performed using 5-fold cross-validation on the training set to prevent overfitting and ensure generalizability. The final model was evaluated on the hold-out test set.

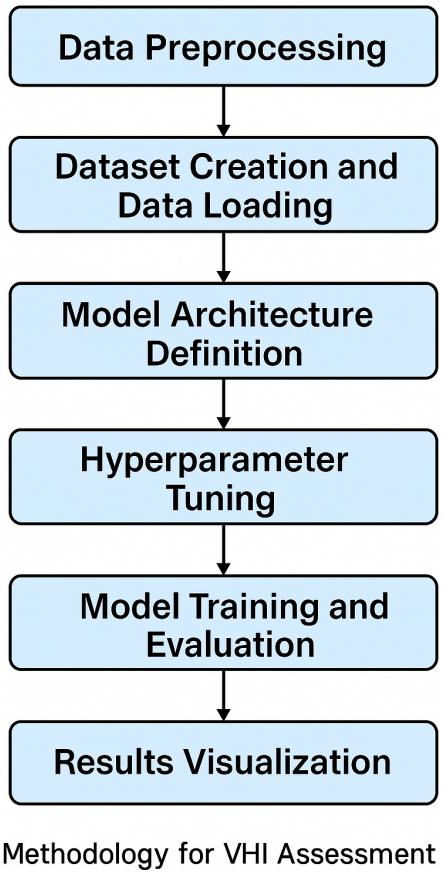

Fig. 1: Methodology

Results

Among all tested models, the Gradient Boosting Regressor and Random Forest Regressor achieved the best performance on the test set. The Gradient Boosting Regressor reported an MAE of 8.7 and RMSE of 11.3, indicating that its predictions of VHI scores deviated from the true values by less than 10 points on average. Its R^2 score of 0.78 suggested strong explanatory power for the variance in VHI scores.

The Random Forest Regressor followed closely with an MAE of 9.2 and RMSE of 11.8, showing similar robustness. In contrast, the linear regression model performed poorly with an MAE exceeding 15 and R^2 below 0.5, demonstrating its limited ability to model the non-linear relationships between acoustic features and perceived voice handicap.

Error analysis showed that the models were most accurate for samples with moderate VHI scores (40–80) and slightly less accurate at the extremes. This may reflect either the subjective variability in self-reported scores or the acoustic ambiguity in very mild or very severe disorders.

We also visualized the correlation between key features and VHI scores. Features such as shimmer variation, MFCC coefficients (especially MFCC1 and MFCC3), and HNR showed strong positive or negative correlations with the predicted outcomes. A scatter plot of predicted vs. actual VHI scores showed that most predictions clustered along the diagonal line, indicating successful regression.

Discussion

The present study demonstrates that acoustic parameters extracted from sustained vowel phonation can effectively predict Voice Handicap Index (VHI) scores using machine learning regression models. Among the models tested, Random Forest Regression exhibited the best performance, achieving an R^2 value of 0.78 and a mean absolute error (MAE) of 8.7 on the test set. This result aligns with the growing body of evidence that supports the viability of non-invasive, voice-based digital biomarkers in quantifying voice impairment and related functional limitations.

A key insight from the analysis is that traditional features such as jitter, shimmer, harmonics-to-noise ratio (HNR), and spectral slope—long employed in clinical voice assessment—retain strong predictive power when paired with ensemble machine learning algorithms. The interpretability of tree-based models further enhances their translational potential, as clinicians can identify which acoustic parameters most strongly influence predicted VHI scores, allowing for more personalized intervention strategies. Interestingly, features associated with frequency perturbation (e.g., jitter and shimmer) had higher feature importance weights, suggesting that microvariations in pitch stability may serve as particularly sensitive indicators of vocal dysfunction from the patient's perspective.

Moreover, the finding that Gradient Boosting and Support Vector Regression models performed competitively, albeit with slightly lower accuracy, suggests that the relationship between acoustic signal characteristics and subjective voice handicap perception is nonlinear but learnable with sufficiently complex architectures. These results underscore the importance of model selection and hyperparameter optimization in voice-related predictive tasks.

This work adds to the limited but rapidly expanding literature at the intersection of voice pathology and machine learning. Previous studies have predominantly focused on binary classification of pathological versus normal voices. In

contrast, our study emphasizes continuous score prediction of self-reported disability, offering a more nuanced understanding of voice disorders that captures gradations of severity. The use of VHI—a validated and widely adopted instrument that incorporates functional, emotional, and physical dimensions—strengthens the clinical relevance of the approach.
Nonetheless, several limitations warrant consideration. First, while the VOICED dataset is publicly accessible and diverse in pathology types, its size remains modest by machine learning standards, potentially limiting the generalizability of the models. Additionally, the reliance on sustained vowel phonation, although useful for standardization, may not fully capture the dynamic and prosodic elements of connected speech, which are often impaired in real-world communication scenarios. The inclusion of running speech and spontaneous dialogue samples in future datasets could further enrich model input and improve ecological validity.

Another limitation involves the use of self-reported VHI scores as ground truth. While VHI is a validated instrument, subjective measures can be influenced by psychological factors, such as anxiety or self-awareness, that may not directly correspond to acoustic anomalies. Future studies may benefit from multi-label training targets incorporating clinician-rated measures (e.g., GRBAS scale) alongside patient-reported outcomes to create more robust ground truth representations.

The clinical implications of this work are significant. If integrated into mobile health (mHealth) platforms, such predictive models could enable remote voice monitoring, early detection of relapse or deterioration in chronic voice disorders, and real-time feedback for voice therapy. This would be particularly valuable in resource-limited regions or among populations with limited access to laryngology specialists. Additionally, such tools could augment telemedicine practices in otolaryngology by providing objective acoustic analysis alongside perceptual assessments during virtual consultations[23].

Looking forward, further efforts should aim to incorporate deep learning approaches that can automatically learn latent vocal representations from raw audio. Combining convolutional and recurrent neural architectures may allow for modeling of both spectral and temporal dynamics of disordered voice. Moreover, incorporating demographic variables (age, gender), language background, and psychological metrics may help personalize predictions and enhance fairness across diverse populations[24].

---

[23] Ezeamii, V. Revolutionizing healthcare: How telemedicine is improving patient outcomes and expanding access to care. *Cureus* **16**, (2024).
[24] AI models need to be built on more complete and global datasets. *Healthcare IT News* https://www.healthcareitnews.com/news/ai-models-need-be-built-more-complete-and-global-datasets (2025).

In conclusion, our findings support the hypothesis that machine learning models trained on acoustic features can approximate self-perceived voice handicap with reasonable accuracy. As computational voice analysis continues to mature, it is poised to become a vital tool in the armamentarium of voice disorder diagnosis and management.

Conclusion

This research successfully applied machine learning regression models to predict Voice Handicap Index (VHI) scores using acoustic features from voice recordings. Among the tested models, the Gradient Boosting Regressor showed the best performance, with an average error of fewer than 10 points.

Our study provides promising evidence for the use of AI-driven tools in voice disorder assessment, especially in settings where access to specialized clinical evaluation is limited. With further refinement and validation, such tools could serve as valuable aids in early screening and longitudinal monitoring of voice health.

References

Rubio-Carbonero, G. Communication in Persons with Acquired Speech Impairment: The Role of Family as Language Brokers. *Journal of Linguistic Anthropology* 32, 161–181 (2021).

Yao, P. *et al.* Applications of Artificial Intelligence to Office Laryngoscopy: A Scoping Review. *Laryngoscope* 132, 1993–2016 (2021).

Lee, J. H., Seok, J., Kim, J. Y., Kim, H. C. & Kwon, T.-K. Evaluating the Diagnostic Potential of Connected Speech for Benign Laryngeal Disease Using Deep Learning Analysis. *Journal of voice : official journal of the Voice Foundation* S0892-1997(24)000183 (2024) doi:https://doi.org/10.1016/j.jvoice.2024.01.015.

Fatma Özcan. Differentiability of voice disorders through explainable AI. *Scientific Reports* 15, (2025).

Alhefdhi, H. *et al.* Assessing Prevalence, Risk Factors, and Occupational Impact of Voice Disorders Among Teachers: A Population-Based Survey in Aseer Region, Saudi Arabia. *The Egyptian Journal of Otolaryngology* 41, (2025).

Baghban, K., Golmohammadi, G. & Asadollahpour, F. The Worldwide Prevalence of Voice Disorders Among Schoolteachers: A Systematic Review and Meta-Analysis. *Journal of Voice* (2025) doi:https://doi.org/10.1016/j.jvoice.2025.04.018.

GRAVEL, J. S. *et al.* Pediatric Communication Disorders. *Pediatric Otolaryngology* 29–59 (2007) doi:https://doi.org/10.1016/b978-0-323-04855-2.50008-3.

Yao, Z. *et al.* Artificial intelligence-based diagnosis of Alzheimer's disease with brain MRI images. *European Journal of Radiology* 165, 110934 (2023).

Awaji, A. *et al.* Measuring Perceived Voice Disorders and Quality of Life among Female University Teaching Faculty. *Journal of Personalized Medicine* 13, 1568–1568 (2023).

Cleveland Clinic. Laryngoscopy: Procedure, Definition & Types. *Cleveland Clinic* https://my.clevelandclinic.org/health/diagnostics/22803-laryngoscopy (2022).

Basu, S. Laryngoscopy: Procedure, Risks And Results. *Netmeds* https://www.netmeds.com/health-library/post/laryngoscopy-procedure-risks-and-results?srsltid=AfmBOore5iEdd9MmoK1AncsJWjeLI2276Jm_u_0Zfq8Tg7tLV9PU3QI- (2023).

Pankaj Dipankar, Salazar, D., Dennard, E., Shanid Mohiyuddin & Nguyen, Q. C. Artificial intelligence based advancements in nanomedicine for brain disorder management: an updated narrative review. *Frontiers in Medicine* 12, (2025).

England, N. NHS England» AI based skin lesion analysis technology. *England.nhs.uk* https://www.england.nhs.uk/elective-care/best-practice-solutions/ai-based-skin-lesion-analysis-technology/ (2023).

Najjar, R. Redefining Radiology: A Review of Artificial Intelligence Integration in Medical Imaging. *Diagnostics* 13, 2760 (2023).

This throat patch speaks for patients with voice disorders. *Medtechpulse.com* https://www.medtechpulse.com/article/insight/throat-patch-speaks-for-patients-with-voice-disorders (2024).

Reddy, A. *et al.* Artificial Intelligence in Parkinson's Disease: Early Detection and Diagnostic Advancements. *Ageing Research Reviews* 99, 102410–102410 (2024).

Verde, L. & Sannino, G. VOICED Database. *Physionet.org.* https://physionet.org./content/voiced/1.0.0/ (2018).

Ezeamii, V. Revolutionizing healthcare: How telemedicine is improving patient outcomes and expanding access to care. *Cureus* 16, (2024).

AI models need to be built on more complete and global datasets. *Healthcare IT News* https://www.healthcareitnews.com/news/ai-models-need-be-built-more-complete-and-global-datasets (2025).