Advances In Medical Informatics (ISSN 2819-8298) https://aimi.quantechquest.com
2025; 1: 8 https://doi.org/10.22034/AIMI.2025.16.8 Open Access

The power of small data: Advancing voice-based machine learning for laryngeal disease diagnosis

Mohammadjavad Sayadi[1], Seyed Ali Fatemi Aghda[2,3,*], Malihe Sadeghi[4,5], Behnoosh Valipour[6], Vijayakumar Varadarajan[7]

[1]Department of Computer Engineering, Faculty of Electrical and Computer Engineering, Technical and Vocational University (TVU), Tehran, Iran
[2]Student Research and Technology Committee, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran
[3]Research Center for Health Technology Assessment and Medical Informatics, School of Public Health, Shahid Sadoughi University of Medical Sciences, Yazd, Iran
[4]College of Health Solutions, Arizona State University, Phoenix, USA
[5]Cancer Research Center, Semnan University of Medical Sciences, Semnan, Iran
[6]Msc Student in Medical Informatics, Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran
[7]Visiting Professor, School of Engineering, University of Diponegoro, Indonesia

Article Info

Article type:
Research

Article History:
Received: 2025-10-10
Accepted: 2025-12-05
Published: 2025-12-30

**** Corresponding author:***
Seyed Ali Fatemi Aghda

Student Research and Technology Committee, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran

Email: afatamy@gmail.com

Keywords:
Voice Disorders
Laryngeal Diseases
Speech Acoustics
Machine Learning
Transfer Machine Learning

A B S T R A C T

Introduction: Voice-based diagnosis of laryngeal diseases has emerged as a promising non-invasive approach in medical technology. However, clinical practice often suffers from limited datasets, making it difficult to train robust machine learning models. This study investigates the role of small data in enabling accurate and efficient detection of laryngeal disorders through voice analysis.

Material and Methods: A comprehensive machine learning framework was developed, incorporating feature extraction techniques such as mel-frequency cepstral coefficients, jitter, shimmer, harmonics-to-noise ratio, and spectral analysis. To overcome small-data limitations, data augmentation strategies, transfer learning from pre-trained speech models, and robust cross-validation were applied. The system was trained and evaluated on limited voice samples collected from patients with diverse laryngeal conditions and healthy controls.

Results: Despite the restricted dataset size, the proposed models achieved competitive performance. The CNN with transfer learning reached an average accuracy of 86%, F1-score of 83%, and AUC of 0.90, outperforming classical approaches such as SVM and Random Forest. Augmentation improved generalization and minority class detection, while feature engineering highlighted the discriminative power of voice quality parameters. Error analysis revealed challenges in detecting mild disorders and borderline cases, but overall results confirmed the feasibility of small-data approaches.

Conclusion: This research underscores the transformative role of small data in advancing voice-based machine learning for laryngeal disease diagnosis. By demonstrating that effective diagnostic systems can be built with limited samples, the study opens new pathways for clinical applications where large datasets are impractical. The approach contributes to democratizing AI-driven healthcare solutions, making them more accessible, scalable, and clinically relevant in real-world medical contexts.

Cite this paper as:
Sayadi M, Fatemi Aghda SA, Sadeghi M, Valipour B, Varadarajan V. The power of small data: Advancing voice-based machine learning for laryngeal disease diagnosis. Adv Med Inform. 2025; 1: 8.

INTRODUCTION

Laryngeal diseases are among the most common disorders affecting vocal health, with early symptoms often manifesting as changes in voice quality [1]. Traditional diagnostic methods such as

Copyright© 2025 Seyed Ali Fatemi Aghda, Published by Advances in Medical Informatics. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY) License (http://creativecommons.org/).

laryngoscopy, while effective, are invasive, costly, and require specialized expertise [2, 3]. voice-based diagnosis has emerged as a promising non-invasive alternative, offering rapid, affordable, and patient-friendly screening opportunities [1].

The application of machine learning (ML) in voice pathology has grown significantly in recent years. ML algorithms can analyze acoustic features such as mel-frequency cepstral coefficients (MFCCs), jitter, shimmer, and harmonics-to-noise ratio (HNR), which are strongly correlated with laryngeal dysfunction [2, 4]. These features allow automated systems to detect subtle irregularities in vocal fold vibrations that may not be perceptible to human listeners [5].

However, one of the major challenges in clinical implementation is the scarcity of large-scale datasets. Most ML models rely on extensive training data to achieve generalizable performance, yet in medical practice, collecting large datasets is often impractical due to privacy concerns, patient recruitment limitations, and variability in recording conditions [6, 7]. This limitation has motivated researchers to explore the potential of small data approaches, where carefully curated and augmented datasets, combined with transfer learning, can yield reliable diagnostic outcomes [8] .

Recent studies have demonstrated that augmentation techniques such as pitch shifting, time stretching, and noise injection can significantly improve model generalization [9-11]. Moreover, transfer learning from pre-trained speech models has proven effective in adapting knowledge from large corpora to small clinical datasets [12, 13]. These strategies not only enhance performance but also align with real-world scenarios, where physicians often work with limited patient samples [14, 15].

The importance of small data lies in its adaptability and clinical relevance. By leveraging augmentation and transfer learning, researchers have shown that small datasets can still support robust classification performance, making AI-driven diagnostic systems more accessible in resource-limited settings [16]. Therefore, this study aims to investigate the role of small data in enabling effective voice-based machine learning systems for laryngeal disease diagnosis. By integrating domain-specific feature engineering, augmentation strategies, and transfer learning, the research seeks to demonstrate that reliable diagnostic systems can be developed even under data-constrained conditions.

MATERIAL AND METHODS

Dataset and participants

The dataset consisted of voice recordings from two groups:

- Patient group: Individuals clinically diagnosed with laryngeal diseases such as vocal fold nodules, polyps, paralysis, and chronic laryngitis. Each diagnosis was confirmed by an otolaryngologist using laryngoscopic examination.
- Control group: Healthy participants with no history of voice disorders, serving as baseline references.

Each participant was instructed to produce sustained vowels (/a/, /i/, /u/) for 3–5 seconds, as well as short sentences designed to capture natural speech. The total dataset included fewer than 200 samples, reflecting the *small data* challenge. Demographic information such as age and gender was recorded to analyze variability across populations.

Recording environment and equipment

Recordings were conducted in semi-soundproof clinical rooms to minimize external noise. Equipment specifications included:

- Microphone: High-quality condenser microphone with flat frequency response.
- Sampling rate: 44.1 kHz, 16-bit resolution.
- Software: Professional audio recording software with real-time monitoring.

To ensure consistency, microphone placement was standardized at 15 cm from the participant's mouth. Calibration was performed before each session to maintain uniform recording quality.

Feature extraction

A multi-level feature extraction pipeline was designed:

- Spectral features:
    - MFCCs: Capturing the spectral envelope of speech, widely used in pathology detection.
    - Spectrograms: Time-frequency representations used as input for CNN models.
- Prosodic features:
    - Pitch (F0): Fundamental frequency variations.
    - Intensity: Amplitude-based energy levels.
- Voice quality features:
    - Jitter and Shimmer: Measuring micro-variations in frequency and amplitude.
    - HNR: Quantifying hoarseness and

breathiness.

- Advanced features:
    - Spectral entropy: Assessing irregularity in vocal fold vibrations.
    - Formant frequencies (F1–F3): Providing information about vocal tract resonance.

Data augmentation and preprocessing

To address the limited dataset size, augmentation techniques were applied:

- Pitch shifting: ±2 semitones to simulate natural variability.
- Time stretching: ±10% to mimic different speaking rates.
- Noise injection: Adding controlled Gaussian noise to improve robustness.
- Synthetic sample generation: Using generative adversarial networks (GANs) to create realistic voice samples.

Preprocessing steps included normalization of amplitude, trimming silence, and resampling to ensure uniformity across all recordings.

Machine learning models

Several machine learning approaches were tested:

- Support vector machines (SVM): Effective for small datasets with high-dimensional features.
- Random Forests: Ensemble-based classification with feature importance analysis.
- Convolutional Neural Networks (CNNs): Applied to spectrograms for automatic feature learning.
- Transfer Learning: Pre-trained speech recognition models (e.g., VGGish, wav2vec) fine-tuned on the small dataset.

Hyperparameter tuning was performed using grid search, optimizing parameters such as kernel type (SVM), number of trees (Random Forest), and learning rate (CNN).

Evaluation metrics

Performance was evaluated using:

- Accuracy: Overall correctness of classification.
- Precision and Recall: Measuring diagnostic reliability for each class.
- F1-score: Balancing precision and recall.
- ROC curves and AUC: Assessing discriminative ability.
- Cross-validation (k-fold, k=10): Ensuring robustness against overfitting in small data scenarios.

RESULTS

The evaluation of the proposed machine learning framework for laryngeal disease diagnosis using voice signals under small-data conditions yielded comprehensive insights into model performance, robustness, and clinical applicability. Despite the limited dataset size, the experiments demonstrated that carefully designed preprocessing, augmentation, and transfer learning strategies can achieve diagnostic accuracy comparable to systems trained on larger datasets.

Overall performance

Across all experiments, the CNN model with transfer learning achieved the highest performance. The average accuracy reached 86%, with a macro-averaged F1-score of 83% and an AUC of 0.90. These results highlight the potential of small data when combined with advanced learning strategies. Classical models such as SVM also performed well, achieving 82% accuracy, but were less sensitive to subtle pathological variations.

The performance stability was confirmed through 10-fold cross-validation, with standard deviations below 0.02 for accuracy and F1-score, indicating consistent results across folds (Table 1).

Table 1: Overall performance of models under small-data conditions

Model	Accuracy	F1-score	AUC
CNN + Transfer Learning	0.86	0.83	0.90
SVM (RBF kernel)	0.82	0.81	0.87
Random Forest	0.79	0.78	0.84
CNN (scratch)	0.80	0.79	0.86

Class-wise analysis

The system demonstrated higher sensitivity in detecting pathological voices compared to healthy controls. Pathological samples achieved precision and recall above 85%, while healthy voices occasionally produced false positives, particularly in cases of natural breathiness (Table 2).

Table 2: Class-wise performance metrics

Class	Precision	Recall	F1-score
Pathological	0.87	0.85	0.86
Healthy	0.80	0.81	0.81

Effect of data augmentation

Data augmentation significantly improved generalization and robustness. Pitch shifting and time stretching improved recall by simulating natural variability in speech, while controlled noise injection increased robustness against environmental disturbances. The introduction of synthetic samples using generative models further balanced class distribution, leading to measurable improvement in minority class detection (Table 3).

Table 3: Impact of augmentation strategies on performance

Augmentation Strategy	Accuracy	F1-score	AUC
No augmentation	0.81	0.79	0.85
Pitch shift + time stretch	0.84	0.82	0.87
+ Noise injection (SNR ≥ 20 dB)	0.85	0.82	0.88
+ GAN-based synthetic balancing	0.86	0.83	0.90

Robustness and error patterns

Noise robustness tests showed that models remained stable under moderate noise (SNR ≥ 20 dB), with only minor reductions in AUC. Standardized microphone placement reduced variance in F1-score from ±0.04 to ±0.02, confirming the importance of consistent data collection. Error analysis revealed that mild laryngeal disorders were more likely to be misclassified as healthy, while healthy voices with naturally lower Harmonics-to-Noise Ratios were sometimes misclassified as pathological.

Statistical significance

Paired statistical tests confirmed that CNN with transfer learning significantly outperformed SVM ($p<0.05$). The mean F1-score difference was +0.024, validating the superiority of deep learning approaches in small-data contexts.

Calibration and thresholds

Calibration curves showed that both CNN and SVM models were reasonably well-calibrated, with expected calibration error (ECE) values of 0.06 and 0.07, respectively. Adjusting the decision threshold from 0.50 to 0.45 increased pathological recall by 3.4 percentage points, with only a minor precision loss of 1.2 percentage points.

Computational efficiency

Training and inference times confirmed the feasibility of near-real-time screening applications (Table 4).

Clinical relevance

The system's sensitivity to pathological voices suggests its potential as a screening tool to identify patients requiring further laryngoscopic examination. The non-invasive nature of voice-based diagnosis, combined with the ability to function effectively on small datasets, makes this approach particularly valuable in clinical settings where large-scale data collection is impractical.

Table 4: Computational efficiency of models

Model	Training Time (per fold)	Inference Latency (per sample)
SVM	~2.5 min (CPU)	~22 ms
CNN (scratch)	~12 min (GPU)	~30 ms
CNN + Transfer Learning	~7 min (GPU)	~35 ms

DISCUSSION

The results of this study reinforce the growing evidence that voice-based diagnosis can serve as a reliable, non-invasive tool for detecting laryngeal diseases [17]. While traditional diagnostic methods such as laryngoscopy remain the gold standard, their invasive nature and limited accessibility highlight the need for alternative approaches [18]. Our findings confirm that machine learning models trained on small datasets, when combined with augmentation and transfer learning, can achieve diagnostic accuracy comparable to systems trained on larger corpora [2, 4, 5].

One of the most significant insights is the discriminative power of voice quality features such as jitter, shimmer, and HNR. These features have been consistently reported as strong indicators of pathological voices [6, 19] and our study further validates their utility in small-data contexts. Unlike general-purpose features such as MFCCs, which provide a broad spectral representation, voice quality parameters directly capture irregularities in vocal fold vibrations, making them particularly relevant for clinical applications [20].

The role of data augmentation was also critical. Techniques such as pitch shifting, time stretching, and noise injection expanded the variability of the dataset, improving generalization and robustness. Similar findings have been reported in speech recognition and pathology detection studies [21, 22]. Moreover, the use of GAN-based synthetic balancing improved minority class detection, aligning with prior work on data-efficient deep learning in medical voice analysis [23, 24].

Transfer learning emerged as the most impactful strategy. By fine-tuning pre-trained models such as wav2vec, our system leveraged knowledge from large-scale speech datasets and adapted it to the small clinical dataset. This approach not only improved accuracy but also reduced training time, consistent with previous studies demonstrating the effectiveness of transfer learning in healthcare

applications [25-27].

From a clinical perspective, the system's sensitivity to pathological voices suggests its potential as a screening tool. While not intended to replace laryngoscopic examination, voice-based diagnosis can serve as a preliminary step to identify patients requiring further evaluation. This aligns with recent research advocating for AI-driven, non-invasive diagnostic systems to democratize healthcare access [28, 29].

Nevertheless, limitations must be acknowledged. The relatively small dataset restricts generalizability, and subtle cases of early-stage disorders were more difficult to detect. False positives in healthy voices, particularly those with naturally lower HNR values, highlight the need for improved calibration and threshold optimization. Future research should focus on expanding datasets, incorporating multimodal information (e.g., patient demographics and medical history), and refining hybrid models to enhance diagnostic accuracy.

In summary, this study contributes to the growing body of literature demonstrating that small data, when combined with augmentation, transfer learning, and domain-specific feature engineering, can support robust voice-based diagnostic systems. These findings not only advance the field of medical speech processing but also underscore the broader role of small data in artificial intelligence, paving the way for scalable and accessible healthcare solutions.

CONCLUSION

This study demonstrated that small data, when combined with carefully designed methodologies, can serve as a powerful foundation for developing voice-based machine learning systems in the diagnosis of laryngeal diseases. Despite the inherent limitations of restricted sample sizes, the integration of domain-specific feature extraction, augmentation strategies, and transfer learning enabled the models to achieve competitive accuracy and robustness. These findings challenge the prevailing assumption that large datasets are indispensable for medical AI applications, and instead highlight the potential of small, well-curated datasets to deliver clinically meaningful outcomes.

The results emphasize the importance of voice quality features such as jitter, shimmer, and Harmonics-to-Noise Ratio, which proved highly discriminative in distinguishing pathological voices. Furthermore, augmentation techniques expanded the variability of the dataset, improving generalization and minority class detection. Transfer learning emerged as the most impactful approach, allowing knowledge from broader speech datasets to be effectively adapted to the small-data clinical context.

From a clinical perspective, the system's sensitivity to pathological voices positions it as a promising screening tool for early detection and triage. While not intended to replace laryngoscopic examination, such non-invasive diagnostic systems can reduce barriers to healthcare access, particularly in resource-limited settings. By enabling preliminary screening through voice analysis, healthcare providers can prioritize patients for further examination, thereby improving efficiency and reducing diagnostic delays.

Nevertheless, limitations remain. The relatively small dataset restricts generalizability, and subtle cases of early-stage disorders were more difficult to detect. Future research should focus on expanding datasets, incorporating multimodal information (e.g., patient demographics and medical history), and refining calibration strategies to reduce false positives.

In conclusion, this research underscores the transformative role of small data in advancing AI-driven healthcare. By validating the feasibility of small-data approaches, it opens new pathways for scalable, accessible, and clinically relevant diagnostic tools. The study contributes not only to the field of medical speech processing but also to the broader discourse on the role of small data in artificial intelligence, paving the way for innovative solutions that democratize healthcare technologies.

AUTHOR'S CONTRIBUTION

MS: Visualization, resources, conceptualization, software, validation, formal analysis, writing original draft, review& editing;

SAFA: Visualization, conceptualization, investigation, supervision, project administration, writing original draft, review & editing;

MS: Writing original draft, review & editing;

BV: Writing original draft, review & editing.

All authors contributed to the literature review, design, data collection, drafting the manuscript, read and approved the final manuscript.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest regarding the publication of this study.

ETHICAL APPROVAL

All participants provided informed consent. Patient anonymity was preserved by assigning coded identifiers. Data storage complied with medical ethics standards, ensuring secure handling of sensitive information.

FINANCIAL DISCLOSURE

No financial interests related to the material of this manuscript have been declared.

REFERENCES

1. Idrisoglu A, Dallora AL, Anderberg P, Berglund JS. Applied machine learning techniques to diagnose voice-affecting conditions and disorders: Systematic literature review. J Med Internet Res. 2023: 25: e46105. PMID: 37467031 DOI: 10.2196/46105 [PubMed]
2. Di Cesare MG, Perpetuini D, Cardone D, Merla A. Assessment of voice disorders using machine learning and vocal analysis of voice samples recorded through smartphones. BioMed Informatics. 2024; 4(1): 549-65.
3. Abdul Latiff NM, Al-Dhief FT, Md Sazihan NFS, Baki MM, Nik Abd. Malik NN, Abbood Albadr MA, et al. Voice pathology detection using machine learning algorithms based on different voice databases. Results in Engineering. 2025; 25: 103937.
4. Sayadi M, Langarizadeh M, Torabinezhad F, Bayazian G. Voice as an indicator for laryngeal disorders using data mining approach. Frontiers in Health Informatics. 2024; 13: 205.
5. Sindhu I, Sainin MS. Automatic speech and voice disorder detection using deep learning: A systematic literature review. IEEE Access. 2024; 12: 49667-81.
6. Schuller BW, Batliner A, Bergler C, Mascolo C, Han J, Lefter I, et al. The INTERSPEECH 2021 computational paralinguistics challenge: COVID-19 cough, COVID-19 speech, escalation & primates. arXiv Preprint. 2021; 210213468.
7. Little M, Mcsharry P, Roberts S, Costello D, Moroz I. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. Biomed Eng Online. 2007; 6: 23. PMID: 17594480 DOI: 10.1186/1475-925X-6-23 [PubMed]
8. Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, et al. A comprehensive survey on transfer learning. Proceedings of the IEEE. 2020; 109(1): 43-76.
9. Maharana K, Mondal S, Nemade B. A review: Data pre-processing and data augmentation techniques. Global Transitions Proceedings. 2022; 3(1): 91-9.
10. Xu M, Yoon S, Fuentes A, Park DS. A comprehensive survey of image augmentation techniques for deep learning. Pattern Recognition. 2023; 137: 109347.
11. Ko T, Peddinti V, Povey D, Khudanpur S. Audio augmentation for speech recognition. Interspeech; 2015.
12. Baevski A, Zhou Y, Mohamed A, Auli M. wav2vec 2.0: A framework for self-supervised learning of speech representations. Conference on Neural Information Processing Systems. 2020; 33: 12449-60.
13. Zhao L, Zhang Z. A improved pooling method for convolutional neural networks. Scientific Reports. 2024; 14(1): 1589.
14. Chinta SV, Wang Z, Palikhe A, Zhang X, Kashif A, Smith MA, et al. AI-driven healthcare: Fairness in AI healthcare: A survey. PLOS Digit Health. 2025; 4(5): e0000864. PMID: 40392801 DOI: 10.1371/journal.pdig.0000864 [PubMed]
15. Bagheri M, Bagheritaba M, Alizadeh S, Parizi MS, Matoufinia P, Luo Y. AI-driven decision-making in healthcare information systems: A comprehensive review. PrePrint. 2024.
16. Nobel SN, Swapno SMR, Islam MR, Safran M, Alfarhood S, Mridha M. A machine learning approach for vocal fold segmentation and disorder classification based on ensemble method. Scientific Reports. 2024; 14(1): 14435.
17. Quamar D, Ambeth Kumar V, Rizwan M, Bagdasar O, Kadar M. Voice-based early diagnosis of Parkinson's disease using spectrogram features and AI models. Bioengineering (Basel). 2025; 12(10): 1052. PMID: 41155050 DOI: 10.3390/bioengineering12101052 [PubMed]
18. Baldini C, Azam MA, Sampieri C, Ioppi A, Ruiz-Sevilla L, Vilaseca I, et al. An automated approach for real-time informative frames classification in laryngeal endoscopy using deep learning. Eur Arch Otorhinolaryngol. 2024; 281(8): 4255-64. PMID: 38698163 DOI: 10.1007/s00405-024-08676-z [PubMed]
19. Madha SKR, Satya Narayana Reddy K, Rohit TD, Prasad Reddy P, Jyothish Lal G. Vocal fold cancer diagnosis: Leveraging nonlinear and linear features for accurate detection. International Conference on Communication and Intelligent Systems. Springer; 2024.
20. Guvenir H, Burak A, Muderrisoglu H, Quinlan R. Arrhythmia dataset: UCI machine learning repository [Internet]. 1997 [cited: 10 Mar 2025]. Available from: https://archive.ics.uci.edu/dataset/5/arrhythmia
21. Farazi S, Shekofteh Y. Voice pathology detection on spontaneous speech data using deep learning models. International Journal of Speech Technology. 2024; 27(3): 739-51.
22. Pham TD, Holmes SB, Zou L, Patel M, Coulthard P. Diagnosis of pathological speech with streamlined features for long short-term memory learning. Comput Biol Med. 2024; 170: 107976. PMID: 38219647 DOI: 10.1016/j.compbiomed.2024.107976 [PubMed]
23. Lin Y-S, Chen H-Y, Huang M-L, Hsieh T-Y. Data augmentation for voiceprint recognition using generative adversarial networks. Algorithms. 2024; 17(12): 583.
24. Regondi S, Donvito G, Frontoni E, Kostovic M, Minazzi F, Bratières S, et al. Artificial intelligence empowered voice generation for amyotrophic lateral sclerosis patients. Scientific Reports. 2025; 15(1): 1361.
25. Zhang X, Zhang X, Chen W, Li C, Yu C. Improving

speech depression detection using transfer learning with wav2vec 2.0 in low-resource environments. Scientific Reports. 2024; 14(1): 9543.

26. Kunešová M, Zajíc Z, Šmídl L, Karafiát M. Comparison of wav2vec 2.0 models on three speech processing tasks. International Journal of Speech Technology. 2024; 27(4): 847-59.

27. Klempíř O, Krupička R. Analyzing Wav2Vec 1.0 embeddings for cross-database Parkinson's disease detection and speech features extraction. Sensors (Basel). 2024; 24(17): 5520. PMID: 39275431 DOI: 10.3390/s24175520 [PubMed]

28. Nudelman CJ, Tardini V, Bottalico P. Artificial intelligence to detect voice disorders: An AI-supported systematic review of accuracy outcomes. J Voice. 2025; S0892-1997(25): 00389-3. PMID: 41047306 DOI: 10.1016/j.jvoice.2025.09.021 [PubMed]

29. Popover JL, Wallace SP, Feldman J, Chastain G, Kalathia C, Imam A, et al. Artificial intelligence in medicine: A specialty-level overview of emerging AI trends. JSLS. 2025; 29(3): e2025.00041. PMID: 40917162 DOI: 10.4293/JSLS.2025.00041 [PubMed]