International Journal of Innovative Technologies in Social Science

e-ISSN: 2544-9435

Scholarly Publisher
RS Global Sp. z O.O.
ISNI: 0000 0004 8495 2390

Dolna 17, Warsaw,
Poland 00-773
+48 226 0 227 03
editorial_office@rsglobal.pl

ARTICLE TITLE	ARTIFICIAL INTELLIGENCE IN THE DIAGNOSIS OF LARYNGEAL CANCER BASED ON ENDOSCOPIC IMAGES: A COMPREHENSIVE NARRATIVE REVIEW
DOI	https://doi.org/10.31435/ijitss.3(47).2025.3838
RECEIVED	24 July 2025
ACCEPTED	03 September 2025
PUBLISHED	23 September 2025
LICENSE	The article is licensed under a **Creative Commons Attribution 4.0 International License.**

ARTIFICIAL INTELLIGENCE IN THE DIAGNOSIS OF LARYNGEAL CANCER BASED ON ENDOSCOPIC IMAGES: A COMPREHENSIVE NARRATIVE REVIEW

Tobiasz Sławiński *(Corresponding Author, Email: tobiasz.slawinski@gumed.edu.pl)*
Provincial Integrated Hospital in Elbląg, Department of Otolaryngology, Królewiecka 146, 82-300 Elbląg, Poland
ORCID ID: 0009-0005-0402-9702

Oliwia Sójkowska-Sławińska
7th Navy Hospital in Gdansk, Polanki 117, 80-305 Gdańsk, Poland
ORCID ID: 0009-0009-1601-3985

Anna Leśniewska
COPERNICUS, Nicolaus Copernicus Hospital in Gdansk, Nowe Ogrody 1/6, 80-803 Gdańsk, Poland
ORCID ID: 0009-0006-3272-3467

Patryk Macuk
University Clinical Center in Gdansk, Dębinki 7, 80-952 Gdańsk, Poland
ORCID ID: 0009-0005-1615-5036

Natalia Rutecka
University Clinical Center in Gdansk, Dębinki 7, 80-952 Gdańsk, Poland
ORCID ID: 0009-0002-8118-6400

ABSTRACT

Background: Laryngeal cancer represents a significant global health burden with early detection being crucial for improved patient outcomes. The integration of artificial intelligence (AI) into medical diagnostics has shown tremendous potential in enhancing the accuracy and efficiency of laryngeal cancer detection through endoscopic image analysis.
Objective: This narrative review comprehensively examines the current state of AI applications in laryngeal cancer diagnosis using endoscopic imaging, focusing on deep learning methodologies, diagnostic performance, clinical validation and future perspectives.
Methods: A systematic analysis of peer-reviewed literature was conducted, examining studies published between 2016 and 2025 that investigated AI-based approaches for laryngeal cancer detection and classification using endoscopic images. The review encompassed various AI techniques including convolutional neural networks (CNNs), transfer learning, multimodal approaches and novel architectures applied to white light imaging, narrow-band imaging and other advanced endoscopic modalities.
Results: Current AI systems demonstrate remarkable diagnostic accuracy with sensitivity ranging from 71-98% and specificity from 86-98% across different studies and methodologies. Deep learning approaches, particularly CNNs such as ResNet, DenseNet, and YOLO architectures, have shown performance comparable to or exceeding that of experienced clinicians. Multimodal AI systems combining white light and narrow-band imaging data achieve superior performance compared to single-modality approaches. Real-time processing capabilities with inference times as low as 0.01-0.03 seconds per image enable practical clinical implementation.
Conclusions: AI-assisted endoscopic diagnosis represents a transformative technology for laryngeal cancer detection, offering the potential to standardize diagnostic protocols, reduce inter-observer variability and improve access to expert-level diagnostic capabilities globally.

KEYWORDS
Artificial Intelligence, Deep Learning, Laryngeal Cancer, Endoscopy, Convolutional Neural Networks, Narrow-Band Imaging

CITATION

Tobiasz Sławiński, Oliwia Sójkowska-Sławińska, Anna Leśniewska, Patryk Macuk, Natalia Rutecka (2025) Artificial Intelligence in the Diagnosis of Laryngeal Cancer Based on Endoscopic Images: A Comprehensive Narrative Review. *International Journal of Innovative Technologies in Social Science*. 3(47). doi: 10.31435/ijitss.3(47).2025.3838

Introduction

Laryngeal cancer represents the second most common malignancy of the head and neck region globally with an estimated annual incidence of over 180, 000 new cases worldwide. The prognosis for laryngeal cancer is heavily dependent on early detection and staging with five-year survival rates exceeding 90% for early-stage disease compared to less than 50% for advanced stages (Xiong et al., 2019). Traditional diagnostic approaches rely primarily on visual inspection during laryngoscopy followed by histopathological confirmation, a process that is highly dependent on clinician experience and can be subject to significant inter-observer variability.

The advent of advanced endoscopic imaging technologies, particularly narrow-band imaging (NBI), has enhanced the visualization of mucosal and vascular patterns associated with malignant transformation (Esmaeili et al., 2020). However, the interpretation of these subtle morphological changes requires considerable expertise and training, limiting their effectiveness in settings with less experienced practitioners or resource constraints. Furthermore, the increasing volume of endoscopic examinations worldwide, driven by improved access to healthcare and growing awareness of head and neck cancer screening, has created a pressing need for standardized, objective diagnostic tools that can assist in accurate lesion characterization while reducing diagnostic delays and improving workflow efficiency.

Artificial intelligence, particularly deep learning methodologies, has emerged as a revolutionary approach to medical image analysis, demonstrating remarkable success across numerous diagnostic domains. The application of AI to laryngeal endoscopic image analysis represents a natural convergence of clinical need and technological capability, offering the potential to enhance diagnostic accuracy, reduce inter-observer variability and facilitate real-time decision-making during endoscopic procedures.

The unique characteristics of laryngeal endoscopic imaging present both opportunities and challenges for AI implementation. The high-resolution visualization of mucosal surfaces, vascular patterns and tissue architecture provides rich morphological information that can be leveraged by machine learning algorithms. However, the variability in imaging conditions, anatomical diversity and the subtle nature of early malignant changes require sophisticated AI approaches capable of handling complex, multi-dimensional data (Wellenstein et al., 2023).

This comprehensive narrative review aims to provide a thorough examination of the current state of AI applications in laryngeal cancer diagnosis through endoscopic imaging. We analyze the evolution of deep learning methodologies, evaluate diagnostic performance across different approaches, examine clinical validation studies and identify key challenges and future directions for this rapidly advancing field.

Methodology

This narrative review was conducted through a comprehensive analysis of peer-reviewed literature published between 2016 and 2025, focusing on artificial intelligence applications for laryngeal cancer diagnosis using endoscopic images. The search strategy encompassed multiple electronic databases and included studies investigating various AI methodologies applied to laryngoscopic image analysis.

Search Strategy and Selection Criteria

The literature search was performed across major medical and engineering databases, targeting studies that investigated AI-based diagnostic systems for laryngeal pathology detection and classification. Inclusion criteria comprised original research articles published in English that presented empirical data on AI system performance for laryngeal cancer detection using endoscopic imaging modalities. Studies were required to provide sufficient methodological detail and quantitative performance metrics to enable meaningful analysis.

Data Extraction and Analysis

From each included study, we systematically extracted information regarding study design, dataset characteristics, AI methodologies employed, imaging modalities utilized, performance metrics, and clinical validation approaches. Particular attention was paid to deep learning architectures, training methodologies, dataset sizes, validation strategies and comparative performance against human experts where available.

Classification Framework

The reviewed studies were categorized according to several key dimensions: (1) AI architectural approaches (CNNs, transfer learning, multimodal systems, novel architectures), (2) imaging modalities (white light, narrow-band imaging, contact endoscopy), (3) diagnostic tasks (detection, classification, segmentation), and (4) validation methodologies (retrospective, prospective, comparative studies with clinicians).

This framework enabled systematic evaluation of technological evolution, performance trends across different approaches and identification of key challenges and opportunities for future development. Special emphasis was placed on studies providing external validation, multicenter assessment and direct comparison with clinical experts to evaluate real-world applicability and clinical translation potential.

Results

Overview of AI Approaches in Laryngeal Cancer Diagnosis

The application of artificial intelligence to laryngeal cancer diagnosis has evolved rapidly over the past decade with the majority of recent studies employing deep learning methodologies. Our analysis reveals a clear progression from traditional machine learning approaches to sophisticated deep neural network architectures with convolutional neural networks (CNNs) emerging as the dominant paradigm for endoscopic image analysis (Zurek et al., 2022).

Convolutional Neural Network Architectures

Single-Architecture CNN Approaches

Multiple studies have investigated the application of established CNN architectures to laryngeal cancer detection. A comprehensive evaluation using DenseNet201 demonstrated exceptional performance, achieving 98.5% training accuracy, 92.0% internal validation accuracy and 86.3% external validation accuracy across 2, 254 laryngoscopic images from 428 patients in two medical centers, with the AI system matching or exceeding the performance of experienced clinicians in external validation (Xu et al., 2023).

A InceptionV3 CNN achieved sensitivity of 90.1% and specificity of 91.5% for laryngeal squamous cell carcinoma detection, outperforming human specialists in processing speed (0.01 seconds vs. 5.5 seconds per image) while maintaining comparable accuracy (He et al., 2021).

The ResNet-101 architecture's reported 96.24% overall accuracy across five diagnostic categories, significantly outperforming 12 otolaryngologists in direct comparison studies (Ren et al., 2020).

Object Detection Architectures

The adaptation of object detection frameworks to laryngeal cancer diagnosis has yielded promising results for real-time clinical applications. YOLOv5-based detection systems achieved sensitivities between 71-78% for carcinoma detection and 70-82% for benign lesions, with processing speeds of 63 frames per second suitable for real-time endoscopic guidance (Wellenstein et al., 2023).

A sophisticated YOLO-based ensemble system incorporating test time augmentation achieved 0.66 precision and 0.62 recall for laryngeal squamous cell carcinoma detection with processing times of 0.026 seconds per frame, demonstrating the feasibility of real-time lesion detection during videolaryngoscopy procedures (Azam et al., 2022).

The Mask R-CNN architecture has been applied to pharyngeal cancer detection, achieving 92% sensitivity for cancer detection with real-time processing capabilities (0.03 seconds per image). However, specificity remained moderate at 47%, highlighting the ongoing challenges in distinguishing malignant from inflammatory conditions (Kono et al., 2021).

Transfer Learning and Feature Extraction Approaches

Transfer learning has emerged as a crucial strategy for overcoming limited dataset sizes in medical AI applications. A comprehensive evaluation of six pre-trained CNN architectures (VGG16, InceptionV4, ResNet variants, Inception-ResNet V2) for informative frame selection in laryngoscopic videos demonstrated that transfer learning significantly outperformed training from scratch, with VGG16 achieving the best performance when combined with support vector machine classification (Patrini et al., 2019).

ResNet-18 with transfer learning for automatic informative frame selection achieved 94.4% precision on test sets while processing significantly faster than human annotators, addressing a critical clinical need by automatically identifying diagnostically relevant frames from lengthy endoscopic video sequences (Yao et al., 2021).

Multimodal and Advanced AI Approaches

Multimodal Integration

The integration of multiple imaging modalities represents a significant advancement in AI-assisted laryngeal diagnosis. The Laryngopharyngeal Artificial Intelligence Diagnostic System (LPAIDS) combines white light imaging and narrow-band imaging features through a sophisticated U-Net-based architecture, achieving 95.6% accuracy on image test sets and 94.9% on video test sets, demonstrating superior performance compared to single-modality approaches (Li et al., 2023).

The multimodal approach addresses the complementary nature of different imaging techniques with white light imaging providing anatomical detail and narrow-band imaging enhancing vascular pattern visualization crucial for malignancy detection (Li et al., 2023).

Super-Resolution Enhancement

SRE-YOLO, a novel approach combining YOLOv8 with a decoupled super-resolution branch during training, improves small lesion detection capabilities while maintaining real-time performance by removing the super-resolution component during inference. This approach achieved 15% improvement in average precision for small lesions compared to baseline YOLOv8 (Baldini et al., 2024).

Multi-Instance Learning

Multi-instance learning applied to vocal fold leukoplakia diagnosis developed a fusion strategy that aggregates image-level predictions to patient-level diagnoses, improving AUC from 0.775 at the image level to 0.869 at the patient level, demonstrating the value of considering multiple images per patient in diagnostic decision-making (Wang et al., 2024).

Specialized Applications and Novel Approaches

Anatomical Site Recognition and Quality Control

The standardization of endoscopic examinations represents a critical clinical need addressed by AI systems. The Intelligent Laryngoscopy Monitoring Assistant (ILMA) using Inception-ResNet-v2 combined with Squeeze-and-Excitation Networks achieved 97.60% accuracy for 20 anatomical site classifications and 99.40% for six regional classifications with real-time processing capabilities enabling quality assurance during endoscopic procedures (Zhu et al., 2023).

ILMA demonstrated exceptional performance metrics with average sensitivity of 100%, specificity of 99.87%, positive predictive value of 97.65% and negative predictive value of 99.87% across all anatomical classifications. The system processes individual images in approximately 21 milliseconds with frame rates of 48 frames per second, enabling real-time monitoring and guidance during endoscopic procedures. Multicenter validation across five independent hospitals demonstrated accuracy of ≥95% for anatomical site identification in clinical video sequences, confirming the system's robustness across different institutional settings and equipment configurations (Zhu et al., 2025).

Vocal Fold Motion Analysis

Advanced AI applications extend beyond static image analysis to dynamic vocal fold assessment. A Fully Convolutional Regression Network for automated vocal fold segmentation in endoscopic videos enables quantitative analysis of vocal fold motion patterns, addressing the complex challenge of tracking dynamic structures in challenging imaging conditions (Hamad et al., 2019).

CNN-LSTM architectures for glottis and vocal fold segmentation in high-speed videos using U-Net-based approaches with bi-directional ConvLSTM layers achieved Dice coefficients of 0.85-0.91 for different anatomical structures, demonstrating the value of temporal information in dynamic tissue analysis (Fehling et al., 2020).

Voice Analysis Integration

The integration of acoustic analysis with endoscopic imaging represents an emerging frontier. One-dimensional CNNs applied to voice signals achieved 85% accuracy in distinguishing pathological voices associated with laryngeal cancer with performance exceeding that of experienced laryngologists, offering potential for screening applications and telemedicine implementations (Kim et al., 2020).

Performance Metrics and Clinical Validation

Diagnostic Accuracy Across Studies

Meta-analytic evidence demonstrates consistently high performance across AI systems for laryngeal cancer detection. Pooled sensitivity and specificity of 91% and 97% respectively for healthy tissue identification, and 91% and 94% for benign versus malignant differentiation have been reported (Zurek et al., 2022). These findings are confirmed by additional systematic reviews reporting pooled sensitivity of 78% and specificity of 86% across 15 studies, with CNN-based models significantly outperforming non-CNN approaches (Alabdalhussein et al., 2025).

Comprehensive meta-analytic evidence demonstrates AI-assisted endoscopy achieved 91% sensitivity for benign-malignant differentiation and 92% accuracy for lesion detection, confirming the clinical utility of these technologies across diverse study populations and methodologies (Marrero-Gonzalez et al., 2024).

Comparison with Human Experts

Direct comparisons between AI systems and human experts have consistently demonstrated competitive or superior AI performance. DenseNet201 models achieved comparable performance to clinicians with 30 years of experience (AUC 0.926 vs. 0.927) while significantly outperforming less experienced practitioners (Xu et al., 2023).

Extensive validation against 12 otolaryngologists ranging from residents to chief physicians showed ResNet-101 algorithms achieving superior accuracy, sensitivity, and specificity across all diagnostic categories. Particularly striking was the improvement in difficult diagnostic categories such as vocal nodules (98% vs. 45%) and leukoplakia (91% vs. 65%) (Ren et al., 2020).

The LPAIDS multimodal system achieved accuracy of 94.0% when compared against experienced laryngologists on 200 video sequences, with performance statistically comparable to expert specialists ($p = 0.063$) while significantly outperforming senior physicians, residents, and trainees ($p \leq 0.001$). Inter-observer reliability analysis demonstrated that AI systems maintain consistent performance levels equivalent to expert clinicians across repeated assessments (Li et al., 2023).

Real-Time Processing Capabilities

The clinical implementation of AI systems requires real-time processing capabilities compatible with endoscopic workflow. Multiple studies have demonstrated inference times suitable for clinical use, with processing speeds ranging from 0.01 seconds to 0.03 seconds per image (He et al., 2021; Kono et al., 2021). These processing speeds exceed typical endoscopic frame rates, enabling real-time diagnostic assistance during procedures.

Imaging Modality Considerations

White Light versus Narrow-Band Imaging

The comparative effectiveness of AI systems across different imaging modalities has been extensively investigated. While narrow-band imaging theoretically provides enhanced visualization of vascular patterns associated with malignancy, meta-analytic evidence suggests no statistically significant differences in AI performance between white light and narrow-band imaging modalities (Zurek et al., 2022; Marrero-Gonzalez et al., 2024).

However, multimodal approaches combining both imaging techniques have demonstrated superior performance compared to single-modality systems, suggesting complementary information content that can be leveraged by sophisticated AI architectures (Li et al., 2023).

Advanced Imaging Techniques

Several studies have explored AI applications with specialized imaging modalities. Stimulated Raman scattering microscopy combined with ResNet34 for rapid histological analysis achieved 100% accuracy in distinguishing neoplastic from normal tissue (Zhang et al., 2019). Hyperspectral imaging applications pioneered CNN-based systems achieving AUCs of 0.85-0.95 for head and neck cancer detection across multiple anatomical sites (Halicek et al., 2019).

Dataset Characteristics and Generalizability

Dataset Sizes and Diversity

The reviewed studies encompass datasets ranging from hundreds to tens of thousands of images with larger datasets generally associated with improved performance and generalizability. Studies utilizing 13, 721 training images across four diagnostic categories and 24, 667 images spanning five conditions represent some of the largest datasets in the field (Xiong et al., 2019; Ren et al., 2020).

Geographic and institutional diversity remains a concern with most studies originating from single institutions or limited geographic regions. Multicenter validation provides more robust evidence for clinical generalizability (Xu et al., 2023; Li et al., 2023).

External Validation and Generalizability

External validation represents a critical component of AI system validation. Studies consistently demonstrate performance degradation when models are tested on external datasets, highlighting the importance of diverse training data and robust validation strategies. Accuracy reduction from 92.0% (internal validation) to 86.3% (external validation) has been observed while maintaining clinically relevant performance levels (Xu et al., 2023).

Discussion

Current State of the Field

The comprehensive analysis of AI applications in laryngeal cancer diagnosis reveals a rapidly maturing field with multiple successful implementations demonstrating clinical-grade performance. The convergence of several technological advances—including powerful CNN architectures, transfer learning methodologies and multimodal integration—has enabled AI systems to achieve diagnostic accuracy comparable to or exceeding that of experienced clinicians across multiple diagnostic tasks.

The predominance of CNN-based approaches reflects the particular suitability of convolutional architectures for medical image analysis with their ability to learn hierarchical feature representations from complex visual patterns. The success of transfer learning approaches addresses a critical challenge in medical AI by enabling effective model training despite limited labeled datasets, leveraging knowledge from large-scale natural image datasets.

Clinical Impact and Healthcare Technology Integration

The demonstrated real-time processing capabilities of current AI systems position them for seamless integration into existing endoscopic workflows. Processing speeds of 0.01-0.03 seconds per image enable real-time diagnostic assistance without disrupting clinical procedures, potentially transforming the endoscopic examination from a purely observational procedure to an AI-augmented diagnostic experience.

The potential impact extends beyond diagnostic accuracy to include standardization of care quality, reduction of inter-observer variability and enhancement of diagnostic capabilities in resource-limited settings where specialized expertise may be unavailable. AI-assisted endoscopy could democratize access to expert-level diagnostic capabilities, particularly relevant given the geographic disparities in otolaryngological expertise globally.

Furthermore, the integration of AI systems with telemedicine platforms could enable remote diagnostic consultations, expanding access to specialized care in underserved regions. The objective nature of AI-based assessments could facilitate second opinion consultations and enable more standardized diagnostic protocols across different healthcare systems.

Technological Advances and Innovation

Architectural Innovations

The evolution from basic CNN implementations to sophisticated multimodal systems represents significant technological advancement. The development of multimodal architectures capable of integrating white light and narrow-band imaging data demonstrates the field's progression toward more comprehensive diagnostic systems that leverage complementary information sources (Li et al., 2023).

Novel approaches such as super-resolution enhancement and multi-instance learning represent innovative solutions to specific challenges in medical image analysis, including small lesion detection and patient-level diagnostic aggregation (Baldini et al., 2024; Wang et al., 2024). These developments indicate continued algorithmic innovation addressing clinical needs.

Quality Assurance and Interpretability

The development of quality control systems for endoscopic procedures represents an important advancement toward comprehensive AI-assisted endoscopy (Zhu et al., 2023; Kist et al., 2025). These systems address the critical need for standardized examination protocols and could significantly improve the consistency and completeness of endoscopic evaluations.

The integration of interpretability techniques, such as gradient-weighted class activation mapping (Grad-CAM), provides clinicians with insights into AI decision-making processes, enhancing trust and enabling educational applications. This transparency is crucial for clinical acceptance and regulatory approval of AI systems.

Challenges and Limitations

Dataset Limitations and Bias

Despite impressive performance metrics, several limitations constrain current AI systems. Dataset sizes, while substantial for medical AI applications, remain limited compared to other computer vision domains. The geographic concentration of studies, predominantly from developed countries with advanced healthcare systems, raises questions about generalizability to diverse populations and healthcare settings.

The retrospective nature of most studies introduces potential selection bias with datasets potentially enriched for clear-cut cases that may not represent the full spectrum of clinical presentations encountered in practice. The predominance of static image analysis, despite some progress in video analysis, may not fully capture the dynamic nature of endoscopic examination.

Technical Challenges

Several technical challenges persist in the field. The variability in imaging conditions, including illumination changes, camera positioning and presence of artifacts (mucus, blood, saliva), continues to challenge AI system robustness. While data augmentation and preprocessing techniques address some of these issues, real-world performance may still be affected by conditions not adequately represented in training data.

The subtle morphological changes associated with early malignant transformation remain challenging for AI detection, evidenced by the moderate performance in distinguishing precancerous from malignant lesions in several studies. This limitation is particularly concerning given the clinical importance of early detection.

Clinical Integration Challenges

The integration of AI systems into clinical practice faces multiple barriers beyond technical performance. Regulatory approval processes for medical AI devices remain complex and time-consuming with requirements for extensive clinical validation and safety demonstration. The lack of standardized evaluation protocols for medical AI systems complicates comparison between different approaches and regulatory assessment.

Clinician acceptance and trust represent additional challenges, particularly given the”black box" nature of deep learning systems. The need for adequate training and education of healthcare providers in AI system use and interpretation represents a significant implementation challenge.

Future Directions and Research Priorities

Prospective Clinical Validation

The transition from retrospective validation to prospective clinical trials represents a critical next step for the field. Such studies would provide more robust evidence of clinical utility and address questions regarding real-world performance, clinical workflow integration, and impact on patient outcomes. The design of appropriate control groups and outcome measures for such trials presents methodological challenges that require careful consideration. Large-scale, multicenter prospective studies are needed to validate the generalizability of AI systems across diverse populations, healthcare settings and geographic regions. Such studies would provide evidence for regulatory approval and clinical guideline development.

Advanced AI Architectures

The continued development of more sophisticated AI architectures offers potential for further performance improvements. Vision transformers, which have shown promise in general computer vision tasks, represent a potential alternative to CNN-based approaches that may better capture long-range spatial relationships in endoscopic images.

The integration of multimodal data beyond imaging, including patient demographics, clinical history and genomic information, could enhance diagnostic accuracy and enable personalized risk assessment. The development of AI systems capable of integrating diverse data sources represents an important research frontier.

Real-Time Learning and Adaptation

The development of AI systems capable of continuous learning and adaptation to local patient populations and imaging conditions could address generalizability challenges. Federated learning approaches could enable model improvement through collaboration between institutions while maintaining patient privacy.

The integration of active learning techniques could optimize the utilization of expert annotations by identifying the most informative cases for labeling, potentially reducing the annotation burden for new implementations.

Comprehensive Diagnostic Systems

Future AI systems may evolve toward comprehensive diagnostic platforms that integrate multiple aspects of laryngeal assessment, including anatomical evaluation, functional analysis through voice processing and longitudinal monitoring through serial imaging. Such systems could provide holistic patient assessment and support clinical decision-making across the entire care continuum.

The integration of AI-assisted endoscopy with treatment planning and outcome prediction could enable more personalized and effective therapeutic interventions. Machine learning approaches for treatment response prediction and prognostic assessment represent important areas for future development. The development of AI systems specifically designed for screening and early detection applications, potentially incorporating simplified imaging platforms or smartphone-based technologies, could democratize access to advanced diagnostic capabilities in underserved populations and resource-limited healthcare settings. Such approaches could have particularly significant impact in regions with high disease burden but limited access to specialized medical expertise.

Public Health and Global Health Implications

The potential impact of AI-assisted laryngeal cancer diagnosis extends beyond individual patient care to population health outcomes. The standardization of diagnostic capabilities could reduce geographic and socioeconomic disparities in cancer detection and outcomes, particularly relevant in regions with limited access to specialized otolaryngological expertise.

The cost-effectiveness of AI-assisted screening programs requires economic evaluation, considering both the costs of technology implementation and the potential savings from earlier detection and reduced treatment complexity. Such analyses would inform healthcare policy decisions and resource allocation strategies.

The development of AI systems specifically designed for resource-limited settings, potentially using simplified imaging equipment or smartphone-based platforms, could democratize access to advanced diagnostic capabilities in underserved populations globally.

Conclusions

Artificial intelligence has emerged as a transformative technology for laryngeal cancer diagnosis through endoscopic image analysis with current systems demonstrating clinical-grade performance across multiple diagnostic tasks. The field has progressed rapidly from basic machine learning approaches to sophisticated deep learning systems capable of real-time processing and multimodal data integration.

The evidence base demonstrates that AI systems can achieve diagnostic accuracy comparable to or exceeding that of experienced clinicians with particular advantages in processing speed, consistency and standardization. The development of multimodal approaches, quality assurance systems and novel architectural innovations indicates continued technological advancement addressing clinical needs.

However, significant challenges remain before widespread clinical adoption can be realized. The need for prospective clinical validation, addressing generalizability concerns, and overcoming implementation barriers represents critical priorities for the field. The development of robust evaluation frameworks and regulatory pathways for medical AI systems requires continued attention from researchers, clinicians, and policymakers.

The potential impact of AI-assisted laryngeal cancer diagnosis extends beyond individual patient care to population health outcomes, offering the possibility of reducing diagnostic disparities and improving access to expert-level care globally. The continued development of this field holds promise for transforming laryngeal cancer diagnosis and improving patient outcomes worldwide.

Future research should prioritize prospective clinical validation, development of more sophisticated multimodal systems and addressing implementation challenges to realize the full potential of AI-assisted endoscopic diagnosis. The convergence of technological capability and clinical need positions this field for continued rapid advancement and eventual widespread clinical adoption.

The integration of artificial intelligence into laryngeal cancer diagnosis represents not merely a technological advancement but a fundamental shift toward more precise, standardized, and accessible healthcare delivery. The field continues to mature with the promise of AI-assisted endoscopic diagnosis to improve patient outcomes and transform clinical practice becoming increasingly attainable.

All authors have read and agreed with the published version of the manuscript.

Funding Statement: The article did not receive any funding.
Institutional Review and Board Statement: Not applicable.
Informed Consent Statement: Not applicable.
Data Availability Statement: Not applicable.
Conflict of Interest Statement: No conflicts of interest to declare.
Acknowledgements: None.

REFERENCES

1. Alabdalhussein, A., Al-Khafaji, M. H., Al-Busairi, R., Al-Dabbagh, S., Khan, W., Anwar, F., Raheem, T. S., Elkrim, M., Sahota, R. B., & Mair, M. (2025). Artificial intelligence in laryngeal cancer detection: A systematic review and meta-analysis. Current Oncology, 32(1), 338. https://doi.org/10.3390/curroncol32060338
2. Azam, M. A., Sampieri, C., Ioppi, A., Africano, S., Vallin, A., Mocellin, D., Fragale, M., Guastini, L., Moccia, S., Piazza, C., Mattos, L. S., & Peretti, G. (2022). Deep learning applied to white light and narrow band imaging videolaryngoscopy: Toward real-time laryngeal cancer detection. The Laryngoscope, 132(8), 1798-1806. https://doi.org/10.1002/lary.29960
3. Baldini, C., Migliorelli, L., Sampieri, C., Ioppi, A., & Mattos, L. S. (2024). Improving real-time detection of laryngeal lesions in endoscopic images using a decoupled super-resolution enhanced YOLO. Computer Methods and Programs in Biomedicine, 260, 108539. https://doi.org/10.1016/j.cmpb.2024.108539
4. Esmaeili, N., Illanes, A., Boese, A., Davaris, N., Arens, C., Navab, N., & Friebe, M. (2020). Laryngeal lesion classification based on vascular patterns in contact endoscopy and narrow band imaging: Manual versus automatic approach. Sensors, 20(14), 4049. https://doi.org/10.3390/s20144018
5. Fehling, M. K., Grosch, F., Schuster, M. E., Schick, B., & Lohscheller, J. (2020). Fully automatic segmentation of glottis and vocal folds in endoscopic laryngeal high-speed videos using a deep convolutional LSTM network. PLoS ONE, 15(2), e0227791. https://doi.org/10.1371/journal.pone.0227791

6. Halicek, M., Dormer, J. D., Little, J. V., Chen, A. Y., Myers, L., Sumer, B. D., & Fei, B. (2019). Hyperspectral imaging of head and neck squamous cell carcinoma for cancer margin detection in surgical specimens from 102 patients using deep learning. Cancers, 11(9), 1367. https://doi.org/10.3390/cancers11091367
7. Hamad, A., Haney, M., Lever, T. E., & Bunyak, F. (2019). Automated segmentation of the vocal folds in laryngeal endoscopy videos using deep convolutional regression networks. Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition Workshops, 89-97. https://doi.org/10.1109/CVPRW.2019.00023
8. He, Y., Cheng, Y., Huang, Z., Xu, W., Hu, R., Cheng, L., He, S., Yue, C., Qin, G., Wang, Y., & Zhong, Q. (2021). A deep convolutional neural network-based method for laryngeal squamous cell carcinoma diagnosis. Annals of Translational Medicine, 9(20), 1553. https://dx.doi.org/10.21037/atm-21-6458
9. 9. Kim, H., Lee, S., Jeon, J., Im, S., Han, Y. J., Joo, Y., & Lee, J. (2020). Convolutional neural network classifies pathological voice change in laryngeal cancer with high accuracy. Journal of Clinical Medicine, 9(11), 3415. https://doi.org/10.3390/jcm9113415
10. Kist, A. M., Razi, S., Groh, R., Gritsch, F., & Schützenberger, A. (2025). Predicting semantic segmentation quality in laryngeal endoscopy images. *PLoS ONE*, *20*(1), e0314573. https://doi.org/10.1371/journal.pone.0314573
11. Kono, M., Inoue, T., Matsueda, K., Waki, K., Fukuda, H., Shimamoto, Y., Fujiwara, Y., & Tada, T. (2021). Diagnosis of pharyngeal cancer on endoscopic video images by Mask region-based convolutional neural network. Digestive Endoscopy, 33(4), 569-576. https://doi.org/10.1111/den.13800
12. Li, Y., Gu, W., Yue, H., Lei, G., Guo, W., Wen, Y., Tang, H., Luo, X., Tu, W., Ye, J., Hong, R., Cai, Q., Gu, Q., Liu, T., Miao, B., Wang, R., Ren, J., & Lei, W. (2023). Real-time detection of laryngopharyngeal cancer using an artificial intelligence-assisted system with multimodal data. Journal of Translational Medicine, 21, 698. https://doi.org/10.1186/s12967-023-04572-y
13. Marrero-Gonzalez, A. R., Meenan, K., O'Rourke, A., Diemer, T. J., Nguyen, S. A., & Camilon, T. J. M. (2024). Application of artificial intelligence in laryngeal lesions: A systematic review and meta- analysis. European Archives of Oto-Rhino-Laryngology, 282(3), 1543-1555. https://doi.org/10.1007/s00405-024-09075-0
14. Patrini, I., Ruperti, M., Moccia, S., Mattos, L. S., Frontoni, E., & De Momi, E. (2019). Transfer learning for informative-frame selection in laryngoscopic videos through learned features. Medical & Biological Engineering & Computing, 57(6), 1225-1238. https://doi.org/10.5281/zenodo.1162784
15. Ren, J., Jing, X., Wang, J., Ren, X., Xu, Y., Yang, Q., Ma, L., Sun, Y., Xu, C., Yang, R., Liu, B., Xiang, M., Liu, J., & Zhao, B. (2020). Automatic recognition of laryngoscopic images using a deep-learning technique. The Laryngoscope, 130(11), E686-E693. https://doi.org/ 10.1002/lary.28539
16. Wang, M. L., Tie, C. W., Wang, J. H., Zhu, J. Q., Chen, B. H., Li, Y., Zhang, S., Liu, L., Guo, L., Yang, L.,Yang, L. Q., Wei, J., Jiang, F., Zhao, Z. Q., Wang, G. Q., Zhang, W., Zhang, Q. M., & Ni, X. G. (2024). Multi-instance learning based artificial intelligence model to assist vocal fold leukoplakia diagnosis: A multicentre diagnostic study. American Journal of Otolaryngology–Head and Neck Medicine and Surgery, 45(4), 104342. https://doi.org/10.1016/j.amjoto.2024.104342
17. Wellenstein, D. J., Marres, H. A. M., Woodburn, J., & van den Broek, G. B. (2023). Detection of laryngeal carcinoma during endoscopy using artificial intelligence. Head & Neck, 45(8), 1943-1952. https://doi.org/10.1002/hed.27441
18. Xiong, H., Lin, P., Yu, J. G., Ye, J., Xiao, L., Tao, Y., Jiang, Z., Lin, W., Liu, M., Xu, J., Hu, W., Lu, Y., Liu, H., Li, Y., Zheng, Y., & Yang, H. (2019). Computer-aided diagnosis of laryngeal cancer via deep learning based on laryngoscopic images. EBioMedicine, 48, 92-99. https://doi.org/10.1016/j.ebiom.2019.08.075
19. Xu, Z. H., Fan, D. G., Huang, J. Q., Wang, J. W., Wang, Y., & Li, Y. Z. (2023). Computer-aided diagnosis of laryngeal cancer based on deep learning with laryngoscopic images. Diagnostics, 13(11), 1924. https://doi.org/10.3390/diagnostics13243669
20. Yao, P., Witte, D., Gimonet, H., German, A., Andreadis, K., Sulica, L., Elemento, O., Cheng, M., Barnes, J., & Rameau, A. (2021). Automatic classification of informative laryngoscopic images using deep learning. Laryngoscope Investigative Otolaryngology, 7(2), 313-322. https://doi.org/10.1002/lio2.754
21. Zhang, L., Wu, Y., Zheng, B., Su, L., Chen, Y., Ma, S., Hu, Q., Zou, X., Yao, L., Yang, Y., Chen, L., Mao, Y., Chen, Y., & Ji, M. (2019). Rapid histology of laryngeal squamous cell carcinoma with deep- learning based stimulated Raman scattering microscopy. Theranostics, 9(9), 2541-2554. https://doi.org/10.7150/thno.32655
22. Zhu, J. Q., Wang, M. L., Li, Y., Zhang, W., Li, L. J., Liu, L., Zhang, Y., Han, C. J., Tie, C. W., Wang, S. X.,Wang, G. Q., & Ni, X. G. (2025). Convolutional neural network based anatomical site identification for laryngoscopy quality control: A multicenter study. American Journal of Otolaryngology–Head and Neck Medicine and Surgery, 44(1), 103695. https://doi.org/10.1016/j.amjoto.2022.103695
23. Zurek, M., Jasak, K., Niemczyk, K., & Rzepakowska, A. (2022). Artificial intelligence in laryngeal endoscopy: Systematic review and meta-analysis. Journal of Clinical Medicine, 11(10), 2752. https://doi.org/10.3390/jcm11102752