

Author

Licence

Julia Zofia Tomaszewska

CC BY 4.0

New Vistas

Volume 12, Issue 1, 2026, UWL Annual Doctoral Students' Conference 2025

Abstract

This research aims to develop a non-invasive and accurate system for detecting and classifying laryngeal pathologies – especially cancerous and precancerous lesions – by combining audio and laryngeal bioimpedance signals in a multimodal deep learning classification architecture. It addresses the central research question: Can a multimodal deep learning framework using audio and electroglottographic (EGG) signals outperform traditional and unimodal approaches in classifying laryngeal conditions?

To answer this, a novel dataset was collected, featuring simultaneous recordings of audio and laryngeal bioimpedance from healthy individuals and patients with a variety of laryngeal disorders. The methodology includes feature extraction using Equivalent Rectangular Bandwidth-based methods, and the design of both unimodal and multimodal deep learning models based on Convolutional Neural Networks (CNNs). The performance of Recurrent Neural Networks (RNNs) for laryngeal pathology classification is also investigated.

The final system adopts a late fusion architecture: two one-dimensional CNNs independently process each modality, and their outputs are combined using a stacked generalisation approach with an ECOC-based meta-classifier. Models were trained and evaluated across two datasets – a custom database and the Saarbruecken Voice Database (SVD) – to ensure generalisability of the developed system.

The results show that the multimodal classification approach significantly outperforms unimodal baselines, particularly in malignant case detection, achieving $89.23\% \pm 1.95$ accuracy, $82.32\% \pm 2.78$ precision, and $83.27\% \pm 3.43$ sensitivity. For general pathology detection, the system achieved even higher classification metrics ($94.92\% \pm 2.82$ accuracy and $96.67\% \pm 2.90$ precision). Continuous speech outperformed the sustained phonation, and ERB-based feature provided richer pathological discrimination than other feature representations.

This study contributes a robust, accurate, and clinically relevant framework for automated laryngeal pathology screening. Its findings support the integration of multimodal voice analysis into future diagnostic tools and lay the groundwork for real-time, non-invasive applications in clinical settings.

Keywords

laryngeal pathologies, bioimpedance, multimodal deep learning classification architecture