Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 22 No. 5 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v22i05.59009

PAPER

Modified Multi-Auscultation Device for Swallowing Sound Analysis: A Feasibility Study toward Non-Invasive Dysphagia Assessment

Pedro Galdames[1], Bojan Petrović[2], David Valencia[1], Lazar Milić[2] (✉), Sanja Kojić[2], Guillaume Serandour[1], Goran M. Stojanović[2]

[1]Universidad Austral de Chile, Valdivia, Republic of Chile

[2]University of Novi Sad, Novi Sad, Republic of Serbia

lmilic@uns.ac.rs

ABSTRACT

This work explores a portable multi-auscultation tool to non-invasively analyse swallowing acoustics, aiming to facilitate early dysphagia screening. To repurpose the existing BODYTUNE vascular-monitoring unit, MEMS microphones were integrated, along with ESP32-based processing, into its architecture. Eight healthy volunteers recorded swallowing sounds while ingesting either saliva or water; the device captured intervals across the cervical domain. Signal segments were clinically defined into initial discrete sound, Bolus transit sound and final discrete sound (FDS), which were extracted and interrogated via time- and frequency-based metrics. Support vector machine (SVM), linear discriminant analysis (LDA), random forest, simple neural network and XGBoost, were employed to differentiate substance types of bolus. Audio fidelity across trials remained high, and the participants adhered to the protocol effortlessly. Within the FDS phase, the peak frequency (PF) and the mean power frequency (MPF) of water and saliva trials diverged substantially ($p = 0.007$ and $p = 0.049$, respectively), corroborating the biological relevance of the extracted features. LDA emerged as the most effective classifier, achieving an overall accuracy of 67% and a balanced F1 measure of 0.73. This study highlights the potential of low-cost, wearable acoustic systems for remote swallowing monitoring and paves the way for future integration with electromyography and machine learning (ML) algorithms for clinical dysphagia diagnosis.

KEYWORDS

biomedical equipment, engineering in medicine, machine learning (ML), microphones, sound recognition

1 INTRODUCTION

Swallowing is a sensorimotor process to transport food and liquid from the mouth to the stomach. Swallowing involves intricate control and coordination of three swallowing phases, referred to as oral, pharyngeal and oesophageal [1]. Dysphagia, defined as

Galdames, P., Petrović, B., Valencia, D., Milić, L., Kojić, S., Serandour, G., Stojanović, G. M. (2026). Modified Multi-Auscultation Device for Swallowing Sound Analysis: A Feasibility Study toward Non-Invasive Dysphagia Assessment. *International Journal of Online and Biomedical Engineering (iJOE)*, 22(5), pp. 87–108. https://doi.org/10.3991/ijoe.v22i05.59009

Article submitted 2025-10-03. Revision uploaded 2026-01-06. Final acceptance 2026-01-06.

difficulty in swallowing, can result from neurological disorders, structural abnormalities, or age-related muscular decline. It affects 7–22% of older adults, increasing to 40–55% in long-term care [2]. Dysphagia is a major risk factor for aspiration pneumonia, especially in elderly patients who may experience silent aspiration during sleep [3], [4], [5].

There are numerous assessment techniques for the swallowing function analysis. When it comes to clinical evaluations, bedside assessments are commonly used to identify swallowing difficulties [6]. In the field of instrumental assessments, most widely used video-fluoroscopic and video-endoscopic studies provide detailed insights into swallowing mechanics [3], [5]. Bedside tests, while safe and easily repeatable, have variable sensitivity (42–92%) and specificity (59–91%), and poor interrater reliability [7]. These tests also fail to detect silent aspiration effectively [8]. Video-fluoroscopy and fiberoptic endoscopy provide detailed anatomical and functional information but require specialised equipment and trained staff [7]. High-resolution oesophageal pressure topography and automated integrated impedance manometry show promise for more objective evaluation of dysphagia [9], [10].

Acoustic analysis has emerged as a non-invasive method to complement clinical swallowing evaluation, particularly in the context of dysphagia [11]. Cervical auscultation, which involves placing an acoustic sensor near the larynx, enables the detection of swallowing sounds. These sounds arise from physiological events, glottal closure, laryngeal elevation, and bolus passage, although direct causal relationships between sound components and anatomical movements are still under investigation [12], [13], [14]. In healthy adults, swallowing sounds typically fall within a frequency range of 0–2 kHz, with most signal energy concentrated below 1.2 kHz [15], [16]. According to Nunes et al. [11], saliva swallows produced a median peak frequency (PF) of 691 Hz, while 50 mL water swallows reached 896 Hz. Sex and age also influence these parameters, with women displaying higher peak frequencies and longer swallow durations, likely due to anatomical factors such as higher laryngeal position and smaller airway structures [17], [18].

Although cervical auscultation provides a non-invasive means of capturing swallowing-related acoustic signals, the resulting data are inherently complex, non-stationary, and highly variable across individuals due to differences in anatomy, age, bolus properties, and sensor placement. Traditional signal-processing approaches relying on handcrafted features and fixed thresholds struggle to robustly capture these variations and to reliably distinguish between swallowing phases, bolus types, and non-swallowing events such as speech or coughing. Machine learning (ML) techniques are therefore essential, as they enable data-driven extraction of discriminative patterns from high-dimensional acoustic features, model nonlinear relationships between signal characteristics and physiological events, and improve classification robustness in the presence of noise and inter-subject variability. By learning directly from labelled acoustic data, ML-based methods offer scalable, objective and reproducible analysis of swallowing sounds, making them particularly suitable for wearable and long-term monitoring scenarios where manual interpretation is impractical [19].

Unlike video-fluoroscopy, acoustic methods are non-invasive and can be implemented in diverse settings, including patients' homes. The absence of ionising radiation makes acoustic analysis suitable for elderly individuals who are more vulnerable to dysphagia [20]. Wearable technologies equipped with acoustic sensors support long-term monitoring outside of clinical environments, thereby improving accessibility and compliance [21]. These methods yield objective metrics essential for tracking neuromuscular performance and decline [20], [22]. Comparative analyses have shown that stethoscopes exhibit relatively low diagnostic accuracy (0.60), whereas acoustic microphones and Doppler-based systems demonstrate higher performance with reported accuracies of 0.94 and 0.80, respectively [23].

Wearable auscultation systems significantly enhance early detection and long-term monitoring of swallowing disorders. For instance, wearable devices have demonstrated automatic swallow detection capabilities with a recall of 79.9% and a precision of 67.6% [24], while deep learning algorithms have achieved classification accuracies as high as 95.96% [25]. Moreover, the non-invasive nature of these sensors reduces the risks associated with procedures like endoscopy [26], and their ability to record continuously allows healthcare providers to analyse swallowing patterns for effective dysphagia management [27]. Wearable sensors, such as accelerometers and microphones placed on the neck, allow for continuous and comfortable monitoring of swallowing sounds and vibrations [28]. Digital transducers used in cervical auscultation similarly capture acoustic signals during deglutition [29].

In parallel, ML algorithms are increasingly applied to classify swallowing events, improving diagnostic accuracy [28]. Despite these technological advancements, several methodological challenges persist. A lack of standardised protocols for sensor placement, signal acquisition, and data analysis contributes to variability across studies [29], while the complexity of swallowing biomechanics demands sophisticated interpretive algorithms that may be difficult to implement clinically [30]. Furthermore, widespread integration into healthcare settings is hindered by limited clinical validation and practitioner acceptance [31]. Various algorithms, including decision trees, support vector machines (SVMs), and neural networks, have demonstrated high accuracy in detecting swallowing events from sound recordings [32]. These models can effectively classify swallowing sounds into different categories [33]. Recent studies have explored more advanced architectures like multi-layer perceptron, convolutional neural network, and convolutional recurrent neural network for classifying swallowing activities in older adults [34]. ML approaches have also been applied to predict and assess voice and swallowing dysfunctions in head and neck cancer patients, offering objective and standardised evaluations [35].

The BODYTUNE system is a portable, multi-auscultation platform originally developed for non-invasive monitoring of vascular sounds, particularly for detecting carotid artery stenosis. It consists of a custom-built digital auscultation device equipped with a MEMS microphone, a battery-powered processing unit, and Bluetooth or Wi-Fi connectivity to a mobile or desktop application. The system has demonstrated stable signal acquisition and repeatability across users, and wavelet-based signal analysis methods have been applied to differentiate vascular dynamics and identify signal artefacts caused by coughing or swallowing.

The objective of the study was to evaluate the feasibility of using a multi-auscultation device inspired by the BODYTUNE device for detecting and characterising swallowing sounds.

2 MATERIALS AND METHODS

A graphical overview of the study workflow, which encompasses four core components: device development, data acquisition, signal processing and segmentation and statistical analysis with machine learning, is shown in the Figure 1. The customised auscultation device inspired by the BODYTUNE system was built by integrating MEMS microphones and modular components to facilitate high-fidelity sound acquisition (top-left). During the experimental phase, the device was placed on the participant's neck to capture swallowing events in real time, which were then transferred to a local system for storage and visualisation (top-right). The recorded signals were processed using filtering and normalisation techniques and segmented into physiologically relevant swallowing phases to extract both time- and frequency-domain features (bottom-right). Finally, statistical tests and supervised ML algorithms were applied to distinguish between saliva and water swallows and to explore the discriminative power of extracted acoustic features (bottom-left).

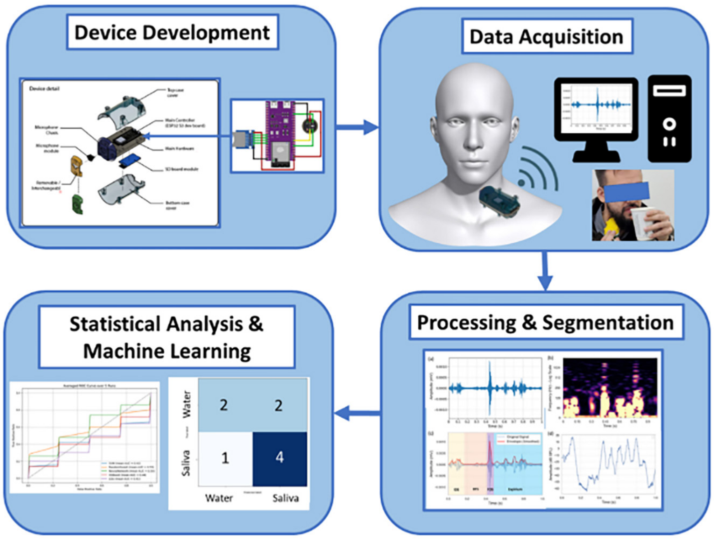

Fig. 1. Overview of the experimental workflow. The study involved four main stages: (1) Device Development – A modular auscultation system was built using MEMS microphones and an ESP32 microcontroller; (2) Data Acquisition – Swallowing sounds were recorded from participants using the wearable system; (3) Processing & Segmentation – Signals were filtered, segmented into swallowing phases, and analysed in time and frequency domains; (4) Statistical Analysis & ML

2.1 Device description

The used device was inspired with respect to and built up on previous publications reported by Suhn et al. and Salvi et al. [36], [37], [38]. Its main function was repurposed from being a vascular detection device to a swallowing sound detection device. The aforementioned auscultation device integrates several hardware components designed for high-fidelity sound acquisition and user-friendly operation. At its core, the initial version employs a single-board computer, the Raspberry Pi Zero W, acting as the host system for data processing, storage, and Wi-Fi connectivity. For mobile use, it includes a lithium-ion polymer battery (LIPO804257, 2000 mAh) manufactured by Shenzhen Pkcell Battery Co., Ltd., paired with a Pimoroni Lipo Shim power management board to regulate power supply. The sensing part comprises a custom-built printed circuit board (PCB) equipped with two bottom-port electret MEMS microphones (Knowles SPH0645LM4HB) capable of digital output via an 18-bit sigma-delta ADC and I2S communication, enabling synchroniSed stereo audio recording at a 16 kHz sampling rate. These microphones are mounted opposite a dual bell-shaped mechanical interface, designed to create a defined air volume and acoustic path between the skin and transducers, with one chamber incorporating a commercially available stethoscope membrane (3M Littmann) to better match acoustic impedance. The microphones are acoustically isolated with double-sided foam tape acting as a gasket and further shielded using acoustic damping materials to reduce environmental noise interference. For user control, the device features a simple interface of two tactile push buttons and two LEDs, balancing ease of operation with compact design. The mechanical housing and transducer interface components were fabricated using 3D printing technology (Formlabs Form2), ensuring precise geometry and fit for the complex acoustic structures. Overall, these integrated hardware elements work together to capture, digitise, and transmit high-quality vascular audio signals for diagnostic purposes.

2.2 Device modifications for swallowing sound acquisition

The modified device for swallowing sound acquisition is shown in the next couple of figures. Figure 2 represents a multisided view of the model taken in Fusion 360 AutoDesk.

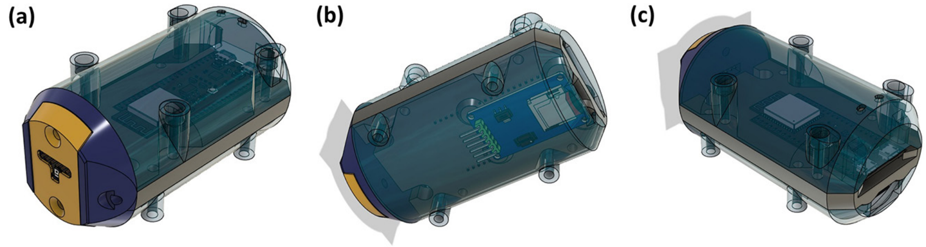

Fig. 2. Fusion 360 AutoDesk model: (a) Frontal view; (b) Bottom view; (c) Top view

As can be seen in Figure 3, the electronic device features several modular components. At its core is the Main Controller, identified as an ESP32 S3 development board, which is mounted on the Main Hardware assembly. The device includes a removable and interchangeable Microphone Module housed in a dedicated Microphone Chassis, allowing for flexibility or upgrades. An SD board module is integrated, suggesting storage or data logging capabilities. The entire assembly is enclosed within a protective casing consisting of a Top Case Cover and a Bottom Case Cover, both designed for secure fitting while maintaining accessibility for component replacement or maintenance. This design highlights modularity and ease of assembly, catering to applications that may require customisable or upgradeable hardware elements.

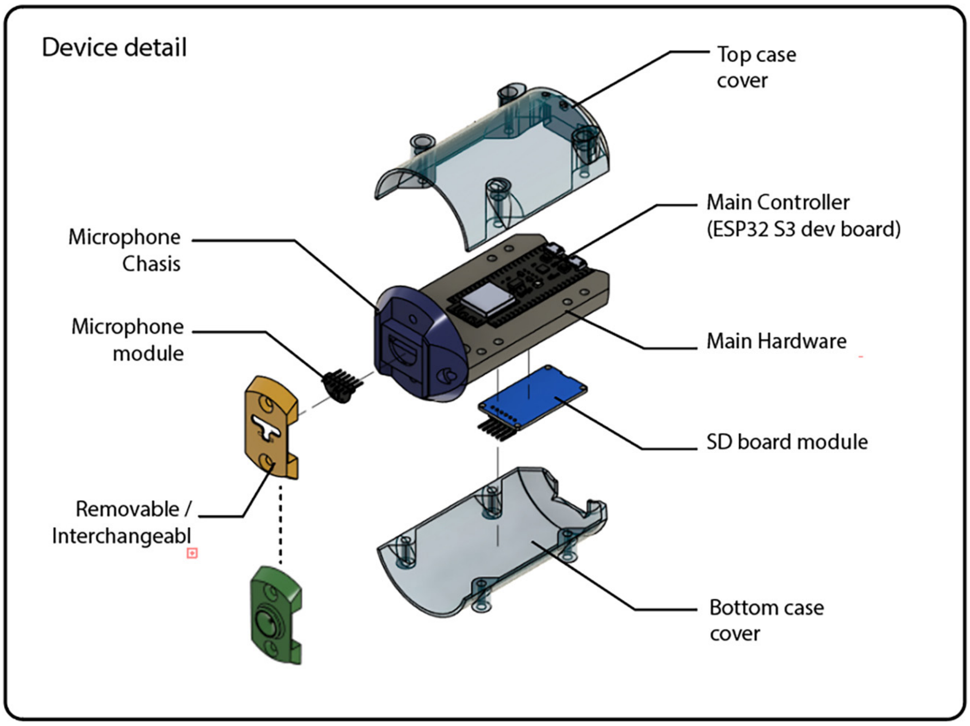

Fig. 3. Exploded view diagram of the auscultation device

Figure 4 depicts a wiring diagram showing how the internal hardware of the auscultation device is connected. An ESP32 development board interfaces with an SD card module and an external microphone module. The SD card module is connected via SPI, with four green wires routed from its pins to the ESP32's SPI pins (labelled 35, 36, 37, and 38), while the red and black wires supply 3.3 V power and ground, respectively, from the ESP32 to the SD card reader. Simultaneously, a circular microphone module, MSM261S4030H0, is connected below the ESP32. It has four connections: VCC (power), GND (ground), WS (Word Select), and SCK (Serial Clock). In the diagram, the red wire connects the microphone's VCC to the ESP32's 3V3 pin, the black wire goes to GND, the green wire links WS to pin 5, and the yellow wire connects SCK to pin 18 on the ESP32. This layout suggests the microphone operates over a digital protocol like I2S, allowing high-quality audio data capture, which can be stored on the SD card for logging or processing. The neat colour-coded wiring and labelling clearly indicate power, ground, and data connections, making it easier to replicate the setup in hardware prototyping.

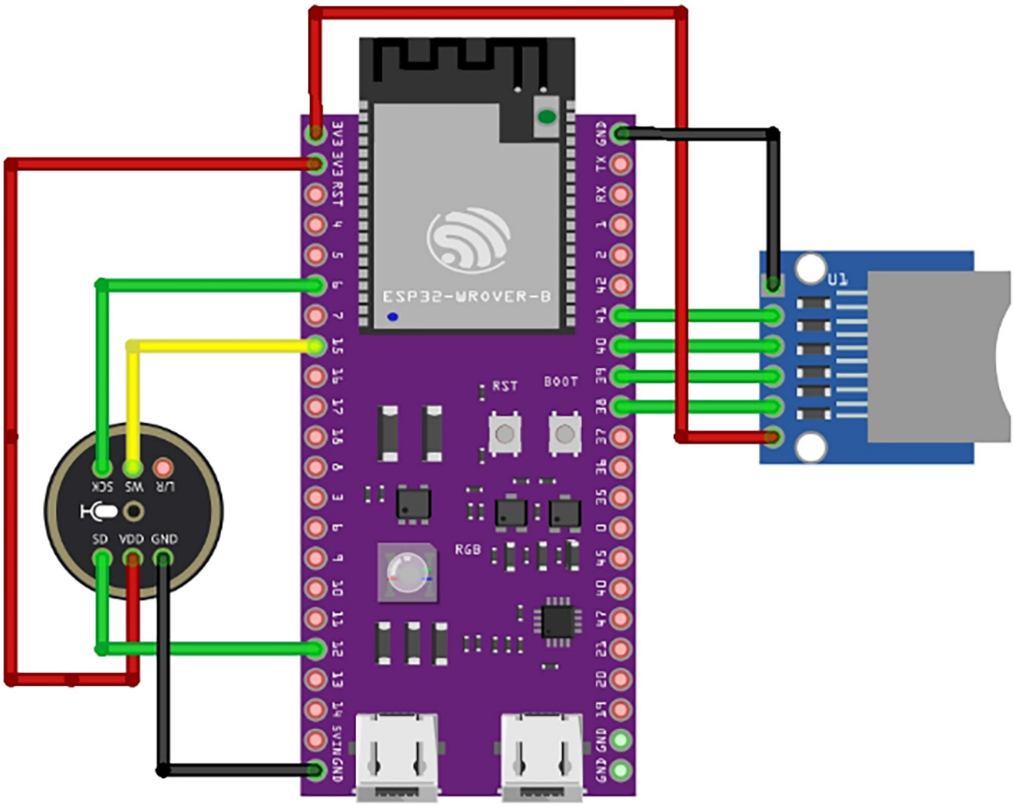

Fig. 4. Circuit diagram of the internal hardware of the auscultation device

2.3 Study population

This study included a convenience sample of eight healthy adult volunteers (six males and two females) aged 18 to 60 years. Participants were recruited from the general community. Inclusion criteria were as follows: no self-reported history of swallowing difficulties, no diagnosed neurological, respiratory, or otolaryngological conditions, and no previous surgery in the head or neck region. Exclusion criteria included the presence of acute respiratory infection at the time of testing, any history of dysphagia, or ongoing medication that could influence neuromuscular function related to swallowing. All participants provided written informed consent prior to participation. The study was approved by the Institutional Ethics Committee Dental Clinic of Vojvodina [01-22/13-2025].

2.4 Signal acquisition protocol

All measurements were conducted in a dedicated laboratory room with closed doors and controlled ambient conditions to minimise environmental noise. Participants were seated in an upright position with their feet flat on the floor and their heads supported in a neutral, midline posture. Before data collection, each participant was instructed to breathe normally and remain still during recordings. Swallowing sounds were recorded using the described auscultatory device, positioned lateral to the trachea, just inferior to the cricoid cartilage, as recommended in previous cervical auscultation literature [39], [40]. Each session consisted of two parts: (1) saliva swallowing and (2) water swallowing. For the water swallowing task, participants were instructed to consume 5 mL of thin liquid (IDDSI-0 [41]) from a plastic cup in a single, unassisted bolus. Each swallowing condition (saliva and water) was repeated three times, with at least 30 seconds of rest between each swallow to allow respiratory and muscular recovery. Before and after each swallowing task, two to three natural respiratory cycles (inhalation and exhalation) were recorded to capture the full swallow-respiratory sequence. The total duration of each recording sequence was approximately 20 to 30 seconds, as per established CA procedures. To address potential environmental noise and motion artefacts, the auscultatory device was equipped with dual MEMS microphones enclosed in acoustically shielded chambers, and the raw signal was band-pass filtered in real time between 20–2000 Hz. Additionally, participants were asked to avoid speaking, coughing, or excessive movement during the protocol. All recordings were saved in uncompressed format (WAV) at 44.1 kHz and securely stored for post-processing and signal analysis. During each swallow, the modified auscultation device was positioned laterally at the neck, between the cricoid cartilage and the trachea, following established cervical auscultation guidelines [39], [42]. To minimise the impact of environmental and mechanical noise, the device's dual-chamber mechanical interface included foam gaskets and acoustic insulation materials.

2.5 Signal processing

Preceding feature extraction, the signal was normalised and the surrounding noise was filtered out with the initial subtraction of the baseline sound signal. The normalisation was done per signal to ensure comparability of data and uniformity. Noise filtering was done digitally as well, administering a bandpass filter in the range 20–2,000 Hz.

The swallowing sound signal consists of three stages: Initial Discrete Sound (IDS), Bolus Transit Sound (BTS) and Final Discrete Sound (FDS) [42]. Corresponding to this, the acquired sound signal was segmented 1000 samples after and before each swallowing action. Following that, each of the three stages was segmented, and from them, the same time-domain and frequency-domain features were extracted.

The extracted features in each stage are as follows [42]–[44]:

Time-domain: duration of the acoustic signal (DAS), peak intensity (PI), duration of the peak intensity (DPI) and root-mean square (RMS);

Frequency domain: PF at peak intensity (FPI), PF, average spectral power over different frequency bands (Plow, Pmed, Phigh), mean power frequency (MPF). Where FPI represents a distributed parameter, analysing parts of the signal spectra through a window of size 256 samples, while PF represents the frequency at which the signal peaks in amplitude across the whole time scale.

The SNR was computed for each swallowing recording using a power-based definition (in dB), where the signal power was estimated from the central portion of the trimmed swallowing event and the noise power from the leading and trailing segments of the recording, representing background noise and motion artefacts. For subjects with multiple repetitions, SNR values were averaged to obtain a single representative value per subject.

Before statistical analysis and machine learning, the data was organised in two groups, swallow sound signals acquired when swallowing saliva and swallow sound signals acquired when swallowing 5 ml of water [42].

2.6 Statistical analysis

Descriptive statistical analysis was conducted on the extracted acoustic features across different swallowing segments (IDS, BTS, FDS), including metrics such as mean, standard deviation, and interquartile ranges for both saliva and water swallow groups. To visualise signal characteristics and distribution patterns, multiple representations were generated: time-domain signal traces, box plots of key features, and wavelet scalograms highlighting the temporal-spectral content of representative swallows. Statistical comparisons between swallow and non-swallow events, as well as between saliva and water swallows, were performed using non-parametric Wilcoxon rank-sum tests due to the small sample size, with p-values and effect sizes reported to assess significance and magnitude of differences. In addition, preliminary ML classification was implemented using supervised models including SVM, random forests and LDA, employing cross-validation and PCA-based dimensionality reduction to evaluate the discriminative potential of the extracted features and explore feasibility for future automated dysphagia screening (as shown in Figure 5). Due to the small dataset, all ML analyses were subject-independent to avoid information leakage. This cross-validation strategy ensures that generalisation is evaluated across participants rather than across repeated samples from the same individual.

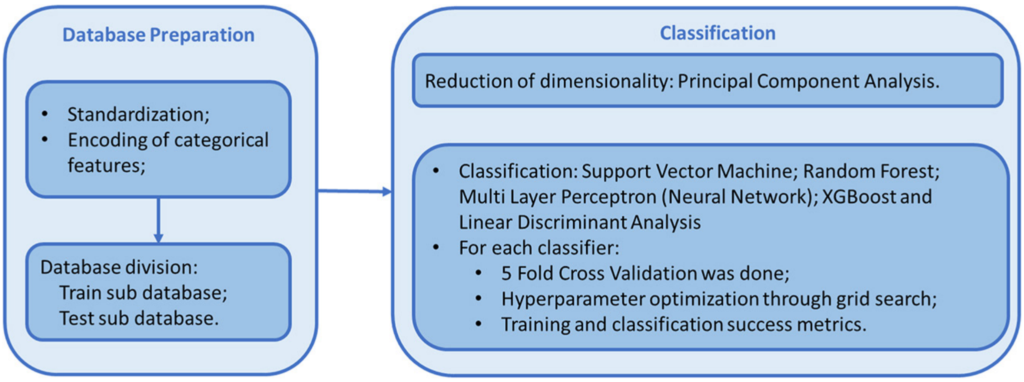

Fig. 5. Workflow of ML-based classification of swallowing events

3 RESULTS

The feasibility of the multi-auscultation device was confirmed through successful data acquisition in all participants. This study included a convenience sample of eight healthy adult volunteers (six males and two females) aged 18 to 60 years. In all cases, the device demonstrated stable signal acquisition with clear identification of swallowing events and minimal ambient interference. The MEMS-based dual-microphone setup consistently captured high-fidelity acoustic signals that were readily processed using the predefined filtering and segmentation pipeline. Participants reported no discomfort during the recordings, and adherence to the swallowing protocol was achieved without deviation. No technical difficulties or procedural inconsistencies were encountered, indicating strong system usability and participant compliance under laboratory conditions.

Figure 6 illustrates the full signal acquisition and preprocessing workflow performed on a representative swallowing sound recording. Panel (a) presents the raw waveform as captured by the modified auscultation device, encompassing various physiological and ambient components, including the swallowing sound, normal breathing cycles, and potential motion or environmental noise. To enhance signal quality, a band-pass filter (20–2000 Hz) was applied, as shown in panel (b), effectively attenuating low-frequency drift and high-frequency artefacts while preserving relevant swallowing acoustics. The time–frequency spectrogram in panel (c) reveals concentrated energy during the swallowing phase, predominantly within the 300–1200 Hz band, which aligns with acoustic characteristics of swallowing events. In panel (d), the processed waveform is annotated into three distinct segments: the initial swallowing sound (blue region), followed by the expirium phase (green region), which captures post-deglutition airflow, and the remainder of the signal, which includes low-amplitude baseline activity and periodic acoustic pulses (red boxes), likely attributable to vascular or muscular noise. This segmentation facilitated the subsequent identification of relevant swallowing epochs for feature extraction and classification.

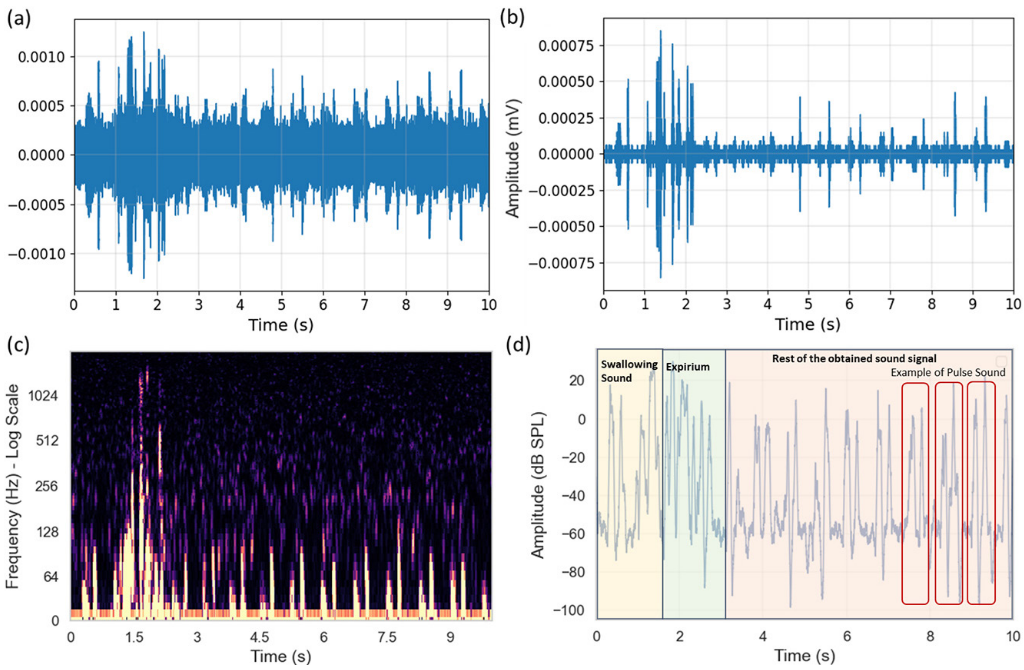

Fig. 6. Representative overview of the swallowing sound data acquisition and preprocessing pipeline: (a) Raw signal acquired during a single swallowing session; (b) Denoised and normalised signal after band-pass filtering (20–2000 Hz); (c) Time–frequency representation (spectrogram) of the processed signal; (d) Annotated signal trace with segment labels: the swallowing sound (blue zone), followed by expirium (green zone). Regions of repeated pulse-like artefacts are highlighted in red boxes

To characterise the acoustic structure of the swallowing event, the trimmed segment was analysed in both the time and frequency domains (see Figure 7). The raw waveform (see Figure 7a) revealed a complex signal morphology with a prominent peak near 0.45 seconds, marking the high-energy phase associated with bolus propulsion. Spectral decomposition (see Figure 7b) indicated that most of the energy was confined to the 200–1000 Hz range, consistent with swallowing signatures. The signal was further segmented into canonical swallowing phases—IDS, BTS, FDS, and Expirium—based on amplitude envelope dynamics (see Figure 7c). These phase labels enabled finer-grained temporal alignment and targeted feature extraction. Lastly, the sound pressure profile in dB SPL (see Figure 7d) showed intensity fluctuations aligning with the segmented stages, with the peak dB value corresponding to the FDS. This segmentation framework supports structured analysis of the swallowing process and informs subsequent feature-based classification tasks.

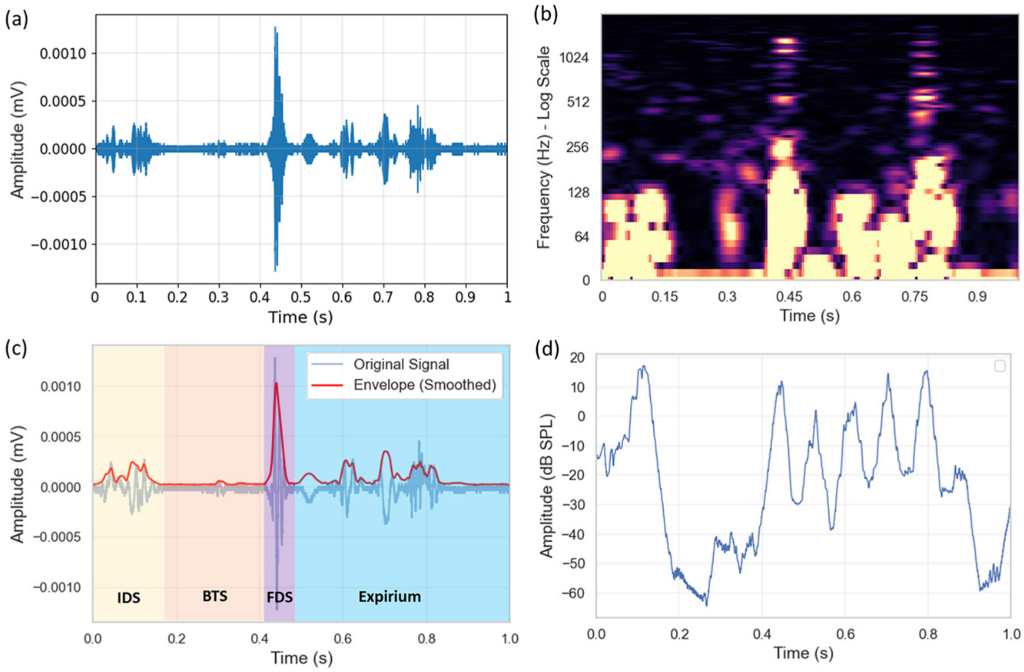

Fig. 7. Signal segmentation and feature representation of a single swallowing event. (a) Raw acoustic waveform of the extracted swallowing segment; (b) Spectrogram (log-scale); (c) Annotated time-domain signal overlaid with smoothed amplitude envelope; (d) Amplitude profile in decibels (dB SPL)

The cumulative representation of normalised swallowing envelopes revealed clear temporal clustering of acoustic activity in both saliva (see Figure 8a) and water (see Figure 8b) swallows. Water swallows tended to exhibit sharper envelope peaks with more consistent alignment across participants, while saliva swallows showed slightly more dispersed onset and offset profiles.

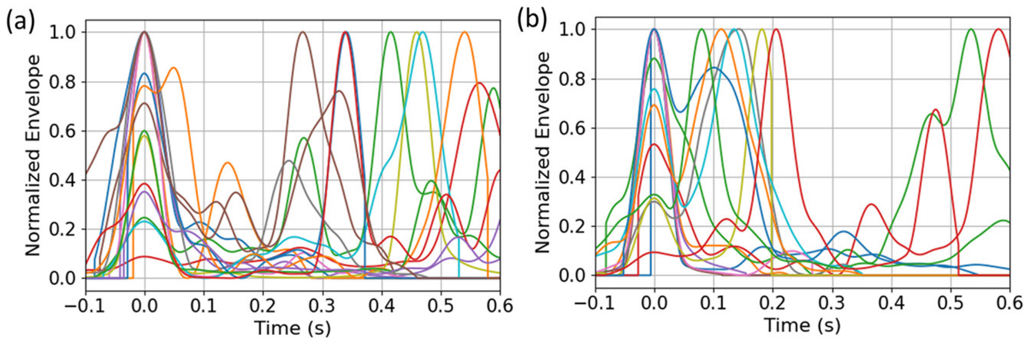

Fig. 8. Cumulative representation of swallowing envelopes: (a) Saliva; (b) Water

Acoustic feature analysis revealed generally consistent signal profiles across saliva and water swallows, with minor differences in both time and frequency domains. Notably, the PF during the FDS segment was significantly higher in water swallows compared to saliva ($p = 0.007$), suggesting increased spectral activity during final bolus propulsion. Additionally, the MPF also differed significantly in the FDS segment ($p = 0.049$), indicating altered frequency-energy distribution, possibly due to bolus volume effects. This is shown in Table 1.

Table 1. Comparison of time-domain and frequency-domain acoustic features across swallowing segments (IDS, BTS, and FDS) for saliva and water swallows

Segment	IDS		BTS		FDS	
Liquid	Saliva	Water	Saliva	Water	Saliva	Water
DAS	0.363 (0.404)	0.292 (0.214)	0.200 (0.172)	0.147 (0.111)	0.195 (0.083)	0.126 (0.114)
PI	0.916 (0.152)	0.970 (0.092)	0.610 (0.101)	0.636 (0.093)	0.757 (0.215)	0.698 (0.181)
DPI	0.023 (0.066)	0.032 (0.043)	0.143 (0.163)	0.115 (0.116)	0.086 (0.118)	0.053 (0.061)
RMS	0.524 (0.106)	0.574 (0.064)	0.522 (0.108)	0.568 (0.064)	0.526 (0.108)	0.569 (0.065)
FPI	1819.556 (4737.76)	233.790 (128.32)	2680.88(5459.1)	246.094 (161.51)	1001.294 (2779.99)	258.399 (175.54)
PF	8.959 (14.499)	4.471 (3.654)	10.515 (11.241)	9.902 (11.513)	**5.804 (5.641)**	**24.706 (44.76)**
Plow	1.853 (2.186)	1.64 (1.458)	0.982 (0.873)	0.746 (0.507)	0.887 (0.492)	0.687 (0.622)
Pmed	33.847 (43.3)	72.971 (65.22)	15.455 (4.884)	17.306 (4.473)	35.318 (52.868)	17.012 (5.235)
Phigh	1.407 (1.839)	0.928 (0.224)	0.895 (0.427)	0.814 (0.177)	1.040 (1.110)	0.823 (0.203)
MPF	7.314 (7.384)	6.348 (4.060)	6.194 (4.703)	5.856 (3.138)	**5.715 (2.991)**	**9.248 (8.035)**

Notes: Data are presented as mean (standard deviation). P-values indicate group differences between saliva and water within each segment.

For liquid (water) swallowing, SNR values ranged from –14.2 dB to 25.2 dB, reflecting substantial inter-subject variability in recording quality. Higher SNR values indicate clearer acoustic swallowing events, while negative SNR values correspond to low-energy swallows in which background noise dominates. For saliva swallowing, SNR values were systematically lower, ranging from –5.6 dB to 10.3 dB, which is physiologically expected due to the smaller bolus volume and reduced acoustic energy associated with saliva swallows.

GridSearchCV, from the module Sklearn, was used to do the grid search for optimal hyperparameters. The results yielded optimal hyperparameters for each classification model after a comprehensive search over their respective parameter spaces.

For SVM, the best performance was achieved using a polynomial kernel with degree 4, a regularisation parameter C of 0.1, and coef 0 of 0.5, along with PCA retaining 95% of variance. The random forest model performed best with 50 trees, a maximum depth of 10, and using log2 as the feature subset strategy, with PCA set to retain 99% variance. For the Neural Network, the top configuration used the SGD solver with a single hidden layer of 50 neurones, ReLU activation, adaptive learning rate, and early stopping enabled, also with PCA set to 95%. XGBoost achieved its optimal results with 100 estimators, a learning rate of 0.1, a max depth of 3, and regularisation parameters (reg_alpha = 0.01, reg_lambda = 1), along with an 80% column sampling and 70% subsampling strategy, with PCA retaining 99%. Lastly, LDA performed best using the least squares (lsqr) solver with automatic shrinkage and PCA set to 95%. These tuned parameters reflect model configurations that balance complexity and generalisation for the classification task. Following that all models were trained and tested with the optimal hyperparameters, and the performance metrics are shown in Table 2. Figure 9 shows the receiver operating characteristic curve and the confusion matrix of the model with the best performance (linear discriminant analysis – LDA).

Table 2. Performance metrics of ML classifiers for swallowing sound classification

Model	Accuracy	Precision	Recall	F1 Score
SVM	0.56	0.58	0.80	0.67
Random Forest	0.34	0.43	0.60	0.50
Neural Network	0.45	0.50	0.60	0.55
XGBoost	0.45	0.50	0.60	0.55
LDA	**0.67**	**0.67**	**0.80**	**0.73**

To evaluate the discriminative power of acoustic features in classifying swallow types, we trained five classification models using dimensionality-reduced datasets. Given the limited sample size and modest AUC values (0.41–0.59), these results should be interpreted cautiously. The classification experiment serves primarily as an exploratory assessment of feature separability rather than as evidence of clinically reliable performance. As shown in Figure 9a, the neural network and XGBoost achieved the highest mean AUC (0.53), followed by the random forest classifier (0.51), indicating higher threshold-independent discrimination between swallow types. In contrast, LDA had the AUC value of 0.47, suggesting limited global ranking performance. Finally, SVM had the lowest performance output of 0.45. However, when evaluated at a fixed decision threshold, LDA achieved the highest accuracy (0.67) and F1 score (0.73), as summarised in Table 2, and demonstrated the most balanced classification between saliva and water swallows in the confusion matrix (see Figure 9b). This highlights the importance of jointly considering both threshold-independent (AUC) and threshold-dependent (accuracy and F1 score) metrics when interpreting classifier performance in small and imbalanced biomedical datasets. Figure 9a shows the averaged ROC curves with 95% confidence bands obtained from repeated stratified cross-validation. All classifiers exhibited mean AUC values close to 0.5, with wide confidence intervals, indicating limited and highly variable discriminative performance between water and saliva swallows. The substantial overlap of confidence bands across models suggests that none of the evaluated classifiers consistently outperformed chance level. These findings highlight the inherent difficulty of distinguishing subtle swallowing conditions using hand-crafted acoustic features in healthy volunteers and motivate the exploration of richer features and multimodal approaches.

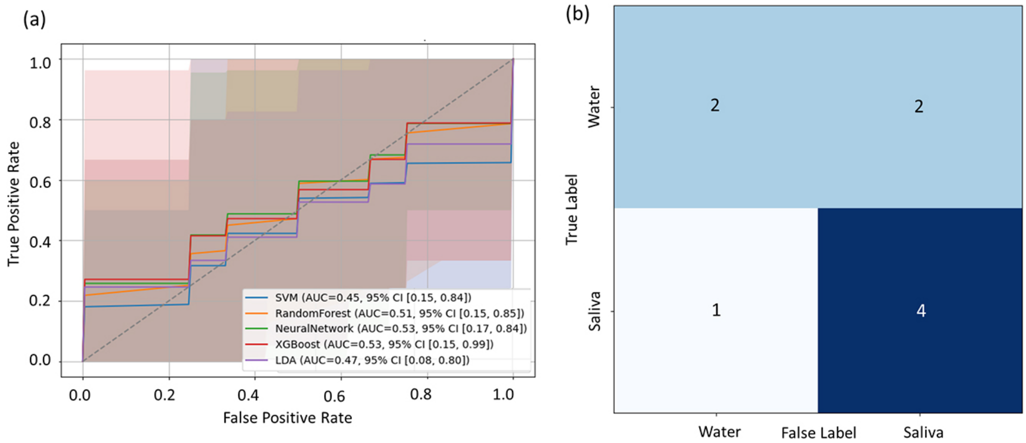

Fig. 9. (a) Receiver operating characteristic (ROC) curves for five classification models; (b) Confusion matrix for the best-performing classifier, LDA

4 DISCUSSION

In the present study, we were inspired by the BODYTUNE system, originally developed as a digital auscultation platform for vascular sound monitoring. This system represents a flexible, research-grade device, not a commercially certified medical product. It combines modular hardware—featuring MEMS microphones, ESP32-based control, and SD card data logging—with customisable software for real-time acoustic acquisition and processing. Prior applications have included carotid pulse detection and home-based vascular risk screening [36], where its multi-sensor layout and ergonomic casing proved effective for neck-based auscultation. We adapted the system for swallowing sound detection due to its proven ability to capture high-fidelity, low-frequency signals in a wearable format and its capacity for standalone data acquisition without reliance on external computers.

In the present investigation, we started with the interpretation and visualisation of raw and filtered acoustic signals, along with time–frequency representations which allowed us to isolate and characterise the swallowing event. Recent research has explored the potential of auscultation for dysphagia screening and swallowing analysis. Cervical auscultation signals have been associated with specific swallowing kinematic events, such as hyoid elevation and laryngeal vestibule closure [45]. Advanced signal processing techniques and ML algorithms have been applied to automatically extract swallowing events from cervical auscultation recordings with high accuracy [46]. These methods have shown promise in characterising swallowing sounds and vibrations, potentially improving the effectiveness of dysphagia screening [28]. Additionally, researchers have also identified a post-swallow glottal release sound that may provide information about the relationship between swallowing and respiration [47]. While these studies, together with the results obtained in our study demonstrate the potential of cervical auscultation in swallowing assessment, further research is needed to standardise segmentation frameworks and fully elucidate the physiological sources of cervical auscultation signals for clinical application.

In the present investigation, the segmentation of the acoustic swallowing signal into IDS, BTS, FDS and Expirium phases provided a structured view of the swallowing process, enabling precise alignment of physiological events with signal features. In particular, the FDS phase exhibited the highest amplitude and spectral energy, consistent with bolus propulsion through the pharynx. This approach is in partial alignment with the methodological approaches used in the similar studies, where researchers on swallowing sound analysis have differently identified distinct phases in the swallowing process. Yadollahi et al. defined three segments: IDS, Bolus transit sounds (BTS) and the entire swallowing sound signal (WHL) [48]. Similarly, Almeida et al. described four components: Discrete initial signal (DIS), main signal (MS), discrete final signal (DFS) and expiratory return (ER) [49]. Honda et al. characterised the swallowing sound into three periods: oral, pharyngeal and repositioning phases, correlating them with specific physiological events [50]. Moussavi used a hidden Markov model to demonstrate that swallowing sounds consistently have three stages across subjects and bolus textures [51].

The cumulative envelope profiles obtained during our analysis confirm that the temporal dynamics of swallowing sounds are more synchronised and intense during water swallowing than saliva swallowing. The earlier and sharper envelope peaks seen in water swallows most certainly derive from greater and coordinated laryngeal and pharyngeal efforts for thin fluid propulsion. This finding stands in line with existing research exploring muscle activation and pressure patterns during swallowing activity. The acoustic feature comparison revealed that several frequency-domain parameters differed between saliva and water swallows, supporting the hypothesis that bolus consistency influences swallowing acoustics. In particular, PF and MPF were significantly higher during water swallows, consistent with the need for more forceful and rapid bolus propulsion. These results align with previous studies showing that thin liquids elicit stronger pharyngeal contractions and greater tongue-palate pressures [52], [53], contributing to increased spectral energy during swallowing. Conversely, the slightly elevated low- and mid-band spectral power (Plow, Pmed) observed in saliva swallows may reflect longer or more variable swallowing durations and reduced muscular effort, as also noted by Sejdic et al. [54]. In line with previous reports, research also indicates that swallowing sounds can differentiate between bolus types, with thicker consistencies generally associated with longer durations and lower frequencies [55], [56].

Importantly, the observed reduction in SNR for saliva swallowing compared to liquid swallowing (the range of SNR for saliva was –5.6 dB to 10.3 dB, while for liquid it was –14.2 dB to 25.2 dB) was consistent across subjects, supporting the validity of the SNR estimation and reflecting genuine physiological differences rather than computational artefacts. The obtained SNR ranges are in agreement with previously reported values for cervical auscultation recordings in healthy volunteers, where very noisy recordings go below 0 dB, while high quality goes as high as 20 dB [57].

In the present study, we explored the feasibility of using supervised ML classifiers to differentiate between saliva and water swallows based on extracted acoustic features. Among the five models tested—SVM, random forest, neural network, XGBoost and LDA—the LDA model demonstrated the best overall performance, with an accuracy of 0.67 and an F1 score of 0.73, indicating moderate discriminative power. While SVM and neural network models also showed reasonable recall values, overall classification metrics suggest that the current feature set and sample size only partially capture the complexity required for robust class separation. These findings validate the potential for ML-based classification in acoustic swallowing analysis but also highlight the need for further refinement in both data acquisition and model training to achieve clinically relevant performance levels. These findings are in line with the studies reporting that ML models have shown promising results in classifying swallowing disorders, particularly when incorporating a broader range of bolus textures, larger datasets and more diverse participant cohorts. These approaches leverage multimodal data inputs and advanced algorithms to achieve high classification accuracies, positioning them as viable non-invasive alternatives to traditional assessments like video-fluoroscopic swallow studies (VFSS). For instance, a wireless multimodal wearable system combining electromyography and swallowing sounds achieved an accuracy of 89.47% in diagnosing silent aspirations using deep learning techniques [58]. Similarly, a SegNet-based method utilising mel-spectrogram features reached a mean F1-score of 80.13% (±4.62%) under 5-fold cross-validation, demonstrating robustness across various water volumes and swallowing tasks [59]. Another study employing multimodal biosignals—surface electromyography and accelerometry-based cervical auscultation—achieved a classification accuracy of 92% (±7%) using an SVM classifier, underscoring the utility of integrated signal modalities [60]. Furthermore, transformer-based neural networks demonstrated detection accuracies exceeding 90% for classification and 85% for segmentation, outperforming hybrid neural networks in terms of generalisability and precision [61]. Despite these advances, ongoing challenges such as methodological heterogeneity, dataset standardisation and limited population diversity remain barriers to widespread clinical adoption.

One of the key strengths of our system lies in its portability, which allows it to be used beyond specialised clinical environments where early dysphagia screening is often needed. The compact design, battery-powered operation and onboard data storage enable independent operation without the need for external computers or tethered infrastructure. Additionally, the system demonstrated high repeatability across all tested participants, with consistent signal quality and reliable detection of swallowing events using the same protocol and sensor positioning. The adaptability of the system is further exemplified by its modular architecture, which permitted a seamless transition from vascular to swallowing sound detection through straightforward mechanical and firmware modifications. This flexibility opens the opportunities for tailoring the device to other physiological sound domains, such as respiratory or gastrointestinal acoustics, and integrating additional sensing modalities like EMG or inertial data for comprehensive functional assessment of the cervical region.

Although the findings from our research show promise, this study also has limitations that need to be taken into consideration. Most prominently, this was a small sample size of just eight healthy adult subjects. Though this convenience sample made it possible to show the technical feasibility and usability of a system, it restricts the generalisability of findings as well as excludes strong statistical inference across demographic or clinical subgroups. Furthermore, lack of a gold standard comparison, for example, simultaneous video-fluoroscopic or endoscopic assessment, makes it impossible to directly validate the inherent acoustic features and swallow-phase segmentation applied in this study. Even though both VFSS and FEES are valuable diagnostic tools for swallowing assessment, neither can be definitively labelled as a singular gold standard due to varying performance across different parameters. Quantitative accuracy varies, as shown by Helliwell et al. [62]. They reported sensitivity and specificity ranges of 0.29–0.33 and 0.96–1.0 for VFSS and 0.37–1.0 and 0.65–0.87 for FEES. Furthermore, Langmore et al. concluded that both techniques are complementary rather than competitive [63]. Critically, Swan et al. revealed significant limitations, noting “insufficient evidence to recommend any individual measure", as a consequence of methodological inconsistencies [64]. Castro et al. similarly found no significant performance differences, suggesting test selection should depend on availability and team expertise [65]. The techniques which appear to be most effective when used together are the previously defined ones, providing comprehensive swallowing biomechanics assessment. In the lack of ground-truth measures of physiology available, interpretation of acoustic markers is inferential yet based on previous literature. In future studies, these weaknesses need to be overcome with larger, more diverse subject groups being included along with alignment with proven instrumental measures of swallowing assessment. Our future work will include validation of ML through VFSS and FEES together as complementary methods of auscultation analysis.

Future research in the field should focus on expanding the ability and clinical utility of the system proposed here by including multimodality fusion, advanced analytics, and clinical validation. A possible direction is to combine data from acoustics with that from surface electromyography (sEMG) to capture both muscular and acoustic aspects of the swallowing process to improve specificity and physiological interpretability of measures. In addition to that, ML-based classifier models need to be created and trained with larger data sets for automatic characterisation of normal vs. abnormal patterns of swallowing, bolus consistencies, or even some dysphagia aetiologies. Time-domain as well as frequency-domain characteristics with envelope dynamics might be incorporated into these models to improve diagnostic yield. Nevertheless, validation in clinically dysphagic populations is necessary in order to deliver sensitivity, specificity, and usability of the system under real-world healthcare scenarios. This would deliver much-needed proof toward moving the platform from feasibility testing to clinical applications.

5 CONCLUSION

In this feasibility study, we demonstrated that a modified version of the BODYTUNE device can reliably capture and analyse swallowing sounds, with the system successfully distinguishing key acoustic features across saliva and water swallows. Our results identified significant differences in frequency-domain features—specifically PF and MPF—during the FDS phase, supporting the device's ability to detect physiological variations related to bolus consistency. Technologically, the system's portability, modular design, and high signal fidelity position it as a viable solution for non-invasive, home-based swallowing assessment. Clinically, such a tool holds promise for early dysphagia detection in at-risk populations, particularly when conventional assessments are inaccessible or impractical. Validation against gold-standard methods like video-fluoroscopy will be essential to establish diagnostic accuracy and clinical relevance, ultimately paving the way toward real-world deployment of this technology for continuous swallowing monitoring and personalised dysphagia care.

6 ACKNOWLEDGMENT

Authors would like to thank the European Union's Horizon Europe Research and Innovation Programme for receiving funding under grant agreement No. 101086348 (Intelligent Wearable System for Enhanced Personalised Gait Rehabilitation – GaitREHub).

7 REFERENCES

[1] L. J. Lazareck and Z. M. Moussavi, "Classification of normal and dysphagic swallows by acoustical means," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 12, pp. 2103–2112, 2004. https://doi.org/10.1109/TBME.2004.836504

[2] C. S. Easterling and E. Robbins, "Dementia and dysphagia," *Geriatr. Nurs.*, vol. 29, no. 4, pp. 275–285, 2008. https://doi.org/10.1016/j.gerinurse.2007.10.015

[3] Y. Suido and S. Teramoto, "Assessment of swallowing function and dysphagia: Is the assessment of swallowing function necessary for the diagnosis of aspiration pneumonia?" in *Aspiration Pneumonia (Respiratory Disease Series: Diagnostic Tools and Disease Managements)*, S. Teramoto and K. Komiya, Eds., Singapore: Springer, 2020, pp. 49–62. https://doi.org/10.1007/978-981-15-4506-1_5

[4] B. C. Villegas, "Clinical swallow evaluation in head and neck cancer," in *Head and Neck Cancer: A Manual and Atlas*, Singapore: Springer, 2018. https://doi.org/10.1007/978-981-10-8282-5_4

[5] H. Wakabayashi, "Assessment of swallowing function and muscle using ultrasonography in general and family medicine," *J. Gen. Fam. Med.*, vol. 21, no. 2, pp. 1–2, 2020. https://doi.org/10.1002/jgf2.298

[6] J. D. Hoffmeister, "Clinical evaluation of swallow," in *Multidisciplinary Management of Pediatric Voice and Swallowing Disorders*, Cham: Springer, 2020. https://doi.org/10.1007/978-3-030-26191-7_16

[7] D. J. Ramsey, D. G. Smithard, and L. Kalra, "Early assessments of dysphagia and aspiration risk in acute stroke patients," *Stroke*, vol. 34, no. 5, pp. 1252–1257, 2003. https://doi.org/10.1161/01.STR.0000066309.06490.B8

[8] J. C. O'Horo, N. Rogus-Pulia, L. Garcia-Arguello, J. Robbins, and N. Safdar, "Bedside diagnosis of dysphagia: A systematic review," *J. Hosp. Med.*, vol. 10, no. 4, pp. 256–265, 2015. https://doi.org/10.1002/jhm.2313

[9] A. Kaindlstorfer and R. Pointner, "An appraisal of current dysphagia diagnosis and treatment strategies," *Expert Rev. Gastroenterol. Hepatol.*, vol. 10, no. 8, pp. 929–942, 2016. https://doi.org/10.1586/17474124.2016.1158098

[10] M. El Halabi, R. Arwani, and H. P. Parkman, "Dysphagia in neurological disorders," *Semin. Neurol.*, vol. 43, no. 4, pp. 530–539, 2023. https://doi.org/10.1055/s-0043-1771458

[11] E. Nunes, L. Menzen, and M. Cardoso, "Swallowing acoustic characteristics of time, intensity, and frequency in healthy adults," *Open J. Otolaryngol.*, vol. 2, no. 2, pp. 10–16, 2019. https://doi.org/10.22259/2639-3603.0202002

[12] S. Morinière, M. Boiron, D. Alison, P. Makris, and P. Beutter, "Origin of the sound components during pharyngeal swallowing in normal subjects," *Dysphagia*, vol. 23, no. 3, pp. 267–273, 2008. https://doi.org/10.1007/s00455-007-9134-z

[13] J. L. Ferrucci, L. D. Mangilli, F. C. Sassi, S. C. Limongi, and C. R. Andrade, "Swallowing sounds in speech therapy practice: A critical analysis of the literature," *Einstein (Sao Paulo)*, vol. 11, no. 4, pp. 535–539, 2013. https://doi.org/10.1590/S1679-45082013000400024

[14] S. R. Youmans and J. A. Stierwalt, "An acoustic profile of normal swallowing," *Dysphagia*, vol. 20, no. 3, pp. 195–209, 2005. https://doi.org/10.1007/s00455-005-0013-1

[15] S. L. Hamlet, R. J. Nelson, and R. L. Patterson, "Interpreting the sounds of swallowing: Fluid flow through the cricopharyngeus," *Ann. Otol. Rhinol. Laryngol.*, vol. 99, pp. 749–752, 1990. https://doi.org/10.1177/000348949009900916

[16] F. S. Soria, R. G. Silva, and A. M. Furkim, "Acoustic analysis of oropharyngeal swallowing using sonar Doppler," *Braz. J. Otorhinolaryngol.*, vol. 82, no. 1, pp. 39–46, 2016. https://doi.org/10.1016/j.bjorl.2015.12.001

[17] O. Brasil, R. Yamasaki, and S. Leão, "Proposta de medição da posição vertical da laringe em repouso," *Rev. Bras. Otorrinolaringol.*, vol. 71, pp. 313–317, 2005. https://doi.org/10.1590/S0034-72992005000300008

[18] K. Oliveira, B. Faria, J. Silva, C. Reis, A. Ghio, and A. C. Gama, "Análise das medidas aerodinâmicas no português brasileiro por meio do método multiparamétrico de avaliação vocal objetiva assistida (EVA)," *Rev. CEFAC*, vol. 15, pp. 119–127, 2013. https://doi.org/10.1590/S1516-18462012005000053

[19] D. W. Wong *et al.*, "Current technological advances in dysphagia screening: Systematic scoping review," *J. Med. Internet. Res.*, vol. 5, no. 27, p. e65551, 2025. https://doi.org/10.2196/65551

[20] J. Baqué, O. Huret, P. Rayneau, M. Schleich, and S. Morinière, "Acoustic analysis of swallowing of an experimental meal of three food textures: A comparative aging study," *Dysphagia*, vol. 39, no. 3, pp. 452–458, 2024. https://doi.org/10.1007/s00455-023-10629-3

[21] B. P. So *et al.*, "Swallow detection with acoustics and accelerometric-based wearable technology: A scoping review," *Int. J. Environ. Res. Public Health*, vol. 20, no. 1, p. 170, 2022. https://doi.org/10.3390/ijerph20010170

[22] H. Firmin, S. Reilly, and A. Fourcin, "Non-invasive monitoring of reflexive swallowing," *Speech Hear. Lang.*, vol. 10, pp. 171–184, 1997. [Online]. Available: http://laryngograph.com/pdfdocs/paper5.pdf

[23] K. V. M. Taveira *et al.*, "Diagnostic validity of methods for assessment of swallowing sounds: A systematic review [dataset]," *Figshare*, 2018. https://doi.org/10.6084/m9.figshare.7243523.v1

[24] T. Olubanjo and M. Ghovanloo, "Real-time swallowing detection based on tracheal acoustics," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Florence, Italy, 2014, pp. 4384–4388. https://doi.org/10.1109/ICASSP.2014.6854430

[25] Y. Song *et al.*, "Multimodal deep ensemble classification system with wearable vibration sensor for detecting throat-related events," *npj Digit. Med.*, vol. 8, p. 14, 2025. https://doi.org/10.1038/s41746-024-01417-w

[26] T. A. Rafeedi *et al.*, "Wearable, epidermal devices for assessment of swallowing function," *npj Flex. Electron.*, vol. 7, p. 52, 2023. https://doi.org/10.1038/s41528-023-00286-9

[27] L. Gravellier, M. Le Coz, J. Farinas, and J. Pinquier, "Detection of pharyngolaryngeal activities in real-world settings using wearable sensors," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Orlando, FL, USA, 2024, pp. 1–5. https://doi.org/10.1109/EMBC53108.2024.10782007

[28] D. Li *et al.*, "A review on intelligent aid diagnosis for dysphagia using swallowing sounds," *Interdiscip. Nurs. Res.*, vol. 2, no. 4, pp. 250–256, 2023. https://doi.org/10.1097/NR9.0000000000000040

[29] J. M. Dudik, J. L. Coyle, and E. Sejdić, "Dysphagia screening: Contributions of cervical auscultation signals and modern signal-processing techniques," *IEEE Trans. Hum.-Mach. Syst.*, vol. 45, no. 4, pp. 465–477, 2015. https://doi.org/10.1109/THMS.2015.2408615

[30] E. Sejdić, G. A. Malandraki, and J. L. Coyle, "Computational deglutition: Using signal- and image-processing methods to understand swallowing and associated disorders," *IEEE Signal Process. Mag.*, vol. 36, no. 1, pp. 138–146, 2019. https://doi.org/10.1109/MSP.2018.2875863

[31] M.-L. Huckabee, P. Macrae, and K. Lamvik, "Expanding instrumental options for dysphagia diagnosis and research: Ultrasound and manometry," *Folia Phoniatr. Logop.*, vol. 67, no. 6, pp. 269–284, 2015. https://doi.org/10.1159/000444636

[32] L. F. Santoso *et al.*, "Applying machine learning algorithms for automatic detection of swallowing from sound," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Berlin, Germany, 2019, pp. 2584–2588. https://doi.org/10.1109/EMBC.2019.8857937

[33] S. Miyagi, S. Sugiyama, K. Kozawa, S. Moritani, S.-i. Sakamoto, and O. Sakai, "Classifying dysphagic swallowing sounds with support vector machines," *Healthcare (Basel)*, vol. 8, no. 2, p. 103, 2020. https://doi.org/10.3390/healthcare8020103

[34] D. Li *et al.*, "Development of the swallowing activity classification model for old adults based on acoustic analysis," *Innov. Aging*, vol. 8, suppl. 1, p. 927, 2024. https://doi.org/10.1093/geroni/igae098.2991

[35] Y. Srinivasan, A. Liu, and A. Rameau, "Machine learning in the evaluation of voice and swallowing in the head and neck cancer patient," *Curr. Opin. Otolaryngol. Head Neck Surg.*, vol. 32, no. 2, pp. 105–112, 2024. https://doi.org/10.1097/MOO.0000000000000948

[36] T. Sühn *et al.*, "Auscultation system for acquisition of vascular sounds—towards sound-based monitoring of the carotid artery," *Med. Devices (Auckl.)*, vol. 13, pp. 349–364, 2020. https://doi.org/10.2147/MDER.S268057

[37] R. Salvi *et al.*, "BODYTUNE: Multi auscultation device—personal health parameter monitoring at home," *Curr. Dir. Biomed. Eng.*, vol. 7, pp. 5–8, 2021. https://doi.org/10.1515/cdbme-2021-2002

[38] I. Müller, J. Henze, A. Burmann, R. Salvi, and R. Baum, "User-centred design of a patient app for carotid artery monitoring at home," in *Human Systems Engineering and Design (IHSED 2023): Future Trends and Applications*, W. Karwowski, T. Ahram, M. Milicevic, D. Etinger, and K. Zubrinic, Eds., vol. 112, USA: AHFE International, 2023. https://doi.org/10.54941/ahfe1004111

[39] P. Fuentealba *et al.*, "Carotid sound signal artifact detection based on discrete wavelet transform decomposition," *Curr. Dir. Biomed. Eng.*, vol. 7, no. 2, pp. 299–302, 2021. https://doi.org/10.1515/cdbme-2021-2076

[40] M. Jaghbeer, A. L. Sutt, and L. Bergström, "Dysphagia management and cervical auscultation: Reliability and validity against FEES," *Dysphagia*, vol. 38, no. 1, pp. 305–314, 2023. https://doi.org/10.1007/s00455-022-10468-8

[41] K. Takahashi, M. E. Groher, and K. Michi, "Methodology for detecting swallowing sounds," *Dysphagia*, vol. 9, pp. 54–62, 1994. https://doi.org/10.1007/BF00262760

[42] International Dysphagia Diet Standardisation Initiative (IDDSI), "The IDDSI framework and descriptors," *ver. 2.0*, 2019. [Online]. Available: https://creativecommons.org/licenses/by-sa/4.0/ [Accessed: Sep. 28, 2025].

[43] M. Aboofazeli and Z. Moussavi, "Automated classification of swallowing and breath sounds," in *Proc. IEEE Eng. Med. Biol. Soc. (EMBC)*, San Francisco, CA, USA, 2004, pp. 3816–3819. https://doi.org/10.1109/IEMBS.2004.1404069

[44] A. Santamato *et al.*, "Acoustic analysis of swallowing sounds: A new technique for assessing dysphagia," *J. Rehabil. Med.*, vol. 41, no. 8, pp. 639–645, 2009. https://doi.org/10.2340/16501977-0384

[45] S. R. Youmans and J. A. G. Stierwalt, "Normal swallowing acoustics across age, gender, bolus viscosity, and bolus volume," *Dysphagia*, vol. 26, no. 4, pp. 374–384, 2011. https://doi.org/10.1007/s00455-010-9323-z

[46] A. Kurosu, J. L. Coyle, J. M. Dudik, and E. Sejdić, "Detection of swallow kinematic events from acoustic high-resolution cervical auscultation signals in patients with stroke," *Arch. Phys. Med. Rehabil.*, vol. 100, no. 3, pp. 501–508, 2019. https://doi.org/10.1016/j.apmr.2018.05.038

[47] A. Y. Cichero and B. E. Murdoch, "What happens after the swallow? Introducing the glottal release sound," *J. Med. Speech Lang. Pathol.*, vol. 11, no. 1, pp. 31–41, 2003.

[48] A. L. Perlman, P. M. Palmer, T. M. McCulloch, and D. J. Vandaele, "Electromyographic activity from human laryngeal, pharyngeal, and submental muscles during swallowing," *J. Appl. Physiol.*, vol. 86, no. 5, pp. 1663–1669, 1999. https://doi.org/10.1152/jappl.1999.86.5.1663

[49] S. T. Almeida, E. L. Ferlin, M. A. M. P. Parente, and H. A. S. Goldani, “Assessment of swallowing sounds by digital cervical auscultation in children,” *Ann. Otol. Rhinol. Laryngol.*, vol. 117, no. 4, pp. 253–258, 2008. https://doi.org/10.1177/000348940811700403

[50] T. Honda *et al.*, “Characterization of swallowing sound: Preliminary investigation of normal subjects,” *PLoS ONE*, vol. 11, no. 12, p. e0168187, 2016. https://doi.org/10.1371/journal.pone.0168187

[51] A. Yadollahi and Z. Moussavi, “Feature selection for swallowing sounds classification,” in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2007, pp. 3172–3175. https://doi.org/10.1109/IEMBS.2007.4353003

[52] D. Kennedy *et al.*, “Tongue pressure patterns during water swallowing,” *Dysphagia*, vol. 25, no. 1, pp. 11–19, 2010. https://doi.org/10.1007/s00455-009-9223-2

[53] I. J. Cook *et al.*, “Timing of videofluoroscopic, manometric events, and bolus transit during the oral and pharyngeal phases of swallowing,” *Dysphagia*, vol. 4, no. 1, pp. 8–15, 1989. https://doi.org/10.1007/BF02407397

[54] E. Sejdić, C. M. Steele, and T. Chau, “Segmentation of dual-axis swallowing accelerometry signals in healthy subjects with analysis of anthropometric effects on duration of swallowing activities,” *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1090–1097, 2009. https://doi.org/10.1109/TBME.2008.2010504

[55] M. L. Ng and A. Ki, “Acoustic characteristics associated with liquid swallowing sounds of different bolus consistencies in young healthy adults,” *Int. J. Lang. Commun. Disord.*, vol. 57, pp. 552–564, 2022. https://doi.org/10.1111/1460-6984.12706

[56] C. Feng, K. Volkman, C. Wagoner, and K.-C. Siu, “The effects of different viscous liquids and solid foods on swallowing speeds and sounds among healthy adults,” *Int. J. Lang. Commun. Disord.*, vol. 57, pp. 78–89, 2022. https://doi.org/10.1111/1460-6984.12675

[57] Y. H. Wu *et al.*, “Characteristics of real-world signal to noise ratios and speech listening situations of older adults with mild to moderate hearing loss,” *Ear Hear.*, vol. 39, no. 2, pp. 293–304, 2018. https://doi.org/10.1097/AUD.0000000000000486

[58] B. Shin *et al.*, “Automatic clinical assessment of swallowing behavior and diagnosis of silent aspiration using wireless multimodal wearable electronics,” *Adv. Sci.*, vol. 11, p. 2404211, 2024. https://doi.org/10.1002/advs.202404211

[59] S. Subramani, M. V. A. Rao, A. Roy, and P. S. Hegde, “SegNet-based deep representation learning for dysphagia classification,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Singapore, 2022, pp. 1141–1145. https://doi.org/10.1109/ICASSP43922.2022.9747732

[60] J. P. Restrepo-Uribe, S. Roldan-Vasco, E. Perez-Giraldo, J. R. Orozco-Arroyave, and A. Orozco-Duque, “Electrophysiological and mechanical approaches to the swallowing analysis,” in *Applied Computer Sciences in Engineering, Communications in Computer and Information Science*, J. C. Figueroa-García, F. S. Garay-Rairán, G. J. Hernández-Pérez, and Y. Díaz-Gutierrez, Eds., vol. 1274, Cham: Springer, 2020, pp. 281–290. https://doi.org/10.1007/978-3-030-61834-6_24

[61] A. Anwar, Y. Khalifa, E. Lucatorto, J. L. Coyle, and E. Sejdić, “Swallowing assessment using high-resolution cervical auscultations and transformer-based neural networks,” in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Orlando, FL, USA, 2024, pp. 1–5. https://doi.org/10.1109/EMBC53108.2024.10782280

[62] K. Helliwell, V. J. Hughes, C. M. Bennion, and A. Manning-Stanley, “The use of videofluoroscopy (VFS) and fibreoptic endoscopic evaluation of swallowing (FEES) in the investigation of oropharyngeal dysphagia in stroke patients: A narrative review,” *Radiography (Lond)*, vol. 29, no. 2, pp. 284–290, 2023. https://doi.org/10.1016/j.radi.2022.12.007

[63] S. E. Langmore, “Evaluation of oropharyngeal dysphagia: Which diagnostic tool is superior?” *Curr. Opin. Otolaryngol. Head Neck Surg.*, vol. 11, no. 6, pp. 485–489, 2003. https://doi.org/10.1097/00020840-200312000-00014

[64] K. Swan, R. Cordier, T. Brown, and R. Speyer, “Psychometric properties of visuoperceptual measures of videofluoroscopic and fibre-endoscopic evaluations of swallowing: A systematic review,” *Dysphagia*, vol. 34, no. 1, pp. 2–33, 2019. https://doi.org/10.1007/s00455-018-9918-3

[65] M. A. Ferrari de Castro *et al*., “Endoscopic and videofluoroscopic evaluations of swallowing for dysphagia: A systematic review,” *Braz. J. Otorhinolaryngol*., vol. 91, no. Suppl 1, p. 101598, 2025. https://doi.org/10.1016/j.bjorl.2025.101598

8 AUTHORS

Pedro Galdames works at the Faculty of Engineering and LeufuLab, Universidad de Austral Chile, where he finished his BSc degree in civil engineering and is pursuing his academic career. His interests lie in development and rapid prototyping of medical and biomedical systems.

Bojan Petrović enrolled at the Faculty of Dentistry in Belgrade in 1994 and graduated in 2000. He enrolled in master’s studies in 2004 and specialist studies in preventive and paediatric dentistry at the Medical Faculty in Novi Sad in 2005. He defended his master’s thesis entitled “Influence of occlusal forces and morphology of the fissure system of the tooth on the effectiveness of prophylactic measures in dentistry” in April 2007. In May 2010, defended a doctoral dissertation entitled “Use of ICF classification in the choice of treatment of persons with mental retardation”. He is the author of more than 100 scientific papers published in national, international and foreign journals. As a Lecturer, Dr. Petrović participated in 11 programmes of continuous medical education.

David Valencia works at the Faculty of Engineering and LeufuLab, Universidad de Austral Chile. Over the years, he has dedicated himself to perfecting the design and use of different materials for 3D printing, with an emphasis on wearable medical devices and biocompatible parts for use both outside and inside the human body. All these iterations are accompanied by extensive use of parametric modelling in order to make micro-adjustments according to the different properties of the materials studied.

Lazar Milić is a researcher at the Department of Power, Electronics and Telecommunication, Faculty of Technical Sciences, University of Novi Sad. He is currently pursuing a doctorate in electronics, under the mentorship of Professor Goran M. Stojanović. Under his mentorship, he has done his bachelors thesis in the field of textile electronics, applied in biomedical engineering, as well as his master thesis in the field of edible electronics, again applied in biomedical engineering. Aside from the aforementioned fields, his interests lie in micro and nanoelectronics, as well as bioelectronics and biosensors. He has published 37 journal and conference papers in the past 5 years, gaining 144 citations and an h-index of 6, in that time (E-mail: lmilic@uns.ac.rs).

Sanja Kojić is an Assistant Professor at the Department of Power, Electronics and Telecommunications, Faculty of Technical Sciences (FTS), University of Novi Sad (UNS), Serbia. She received an MSc degree in 2010 and a PhD degree in 2022 from FTS-UNS, in electrical engineering. Author of over 104 articles, including over 60 in leading peer-reviewed journals with impact factors and one patent. Her scientific papers have been cited 649 times, and her h-factor is 14. Her areas of research cover microfluidics, flexible, printed and green electronics, mechanical and microscopic characterisation, and archaeology. Mrs. Kojić has more than 10 years of work experience on national and international projects (H2020, FP7, ERASMUS, TEMPUS, EUREKA, and COST). She runs a volunteering programme at the Engineering Creativity Centre of the University of Novi Sad, where students can get access and training to the newest fabrication equipment for the textile, flexible and printed electronics as well as microfluidic chips.

Guillaume Serandour has been teaching technological innovation and digital manufacturing for Engineering undergraduate students at the Universidad Austral de Chile since 2009 using a Problem and Project-Based Learning approach. He established LeufüLAB, a makerspace which draws students and academics into activities that make use of modern digital manufacturing technologies. An important focus of my work is the use of those technologies applied to health care. His disciplines are mechanical engineering, biomedical engineering and engineering education. His h-index is 7, with over 600 citations.

Goran M. Stojanović is a full Professor at the Department of Power, Electronics and Telecommunications, Faculty of Technical Sciences (FTS), University of Novi Sad (UNS), Serbia. He received a BSc, an MSc and a PhD degree in 1996, 2003 and 2005, respectively, from FTS-UNS, all in electrical engineering. He has 27 years of experience in R&D. His research interests include sensors, flexible electronics, textile electronics, edible electronics and microfluidics. He is an author/co-author of 391 articles, 5 books, 3 patents, and 1 chapter in a monograph. He was a keynote speaker at 12 international conferences, as he was a supervisor of 13 PhD students, 40 MSc students and 60 diploma students at the FTS-UNS. He has more than 17 years' experience in coordination of EU-funded projects (Horizon Europe, H2020, EUREKA, ERASMUS, CEI), with a total budget exceeding 18.86 MEUR. Currently, he coordinates 5 Horizon projects in the field of green electronics and textile electronics (https://www.stojanovicgoran.com).