IEEE Journal of Translational Engineering in Health and Medicine

Received 28 August 2024; accepted 4 November 2024.
Date of publication 13 November 2024; date of current version 20 November 2024.

Digital Object Identifier 10.1109/JTEHM.2024.3497895

Elevating Patient Care With Deep Learning: High-Resolution Cervical Auscultation Signals for Swallowing Kinematic Analysis in Nasogastric Tube Patients

FARNAZ KHODAMI[1], AMANDA S. MAHONEY[2], JAMES L. COYLE[2], AND ERVIN SEJDIĆ[1,3], (Senior Member, IEEE)

[1]Department of Electrical and Computer Engineering, Faculty of Applied Science and Engineering, University of Toronto, Toronto, ON M5S 1A4, Canada
[2]Department of the Communication Science and Disorders, School of Health and Rehabilitation Sciences, University of Pittsburgh, Pittsburgh, PA 15213, USA
[3]North York General Hospital, Toronto, ON M2K 1E1, Canada

CORRESPONDING AUTHOR: F. KHODAMI (farnaz.khodami@utoronto.ca)

This work was supported in part by the Eunice Kennedy Shriver National Institute of Child Health and Human Development under Grant 1R01HD092239-01 and in part by the National Science Foundation under Grant 1652203.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of the University of Pittsburgh under Application No. 12080498.

ABSTRACT Patients with nasogastric (NG) tubes require careful monitoring due to the potential impact of the tube on their ability to swallow safely. This study aimed to investigate the utility of high-resolution cervical auscultation (HRCA) signals in assessing swallowing functionality of patients using feeding tubes. HRCA, capturing swallowing vibratory and acoustic signals, has been explored as a surrogate for videofluoroscopy image analysis in previous research. In this study, we analyzed HRCA signals recorded from patients with NG tubes to identify swallowing kinematic events within this group of subjects. Machine learning architectures from prior research endeavors, originally designed for participants without NG tubes, were fine-tuned to accomplish three tasks in the target population: estimating the duration of upper esophageal sphincter opening, estimating the duration of laryngeal vestibule closure, and tracking the hyoid bone. The convolutional recurrent neural network proposed for the first task predicted the onset of upper esophageal sphincter opening and closure for 67.61% and 82.95% of patients, respectively, with an error margin of fewer than three frames. The hybrid model employed for the second task successfully predicted the onset of laryngeal vestibule closure and reopening for 79.62% and 75.80% of patients, respectively, with the same error margin. The stacked recurrent neural network identified hyoid bone position in test frames, achieving a 41.27% overlap with ground-truth outputs. By applying established algorithms to an unseen population, we demonstrated the utility of HRCA signals for swallowing assessment in individuals with NG tubes and showcased the generalizability of algorithms developed in our previous studies. Clinical impact: This study highlights the promise of HRCA signals for assessing swallowing in patients with NG tubes, potentially improving diagnosis, management, and care integration in both clinical and home healthcare settings.

INDEX TERMS High-resolution cervical auscultation signals, hyoid bone tracking, laryngeal vestibule closures, upper esophageal opening, videofluoroscopic swallowing study.

I. INTRODUCTION

A NASOGASTRIC (NG) tube is a medical device used to deliver essential nutrition, medication, or fluids to individuals who are unable to consume them orally. These tubes are inserted through the nose and passed down through the esophagus into the stomach, providing a lifeline for patients with various medical conditions [1], [2]. However, the introduction of an NG tube into the aerodigestive tract can significantly affect the patient's ability to swallow safely and efficiently. The altered anatomy and sensation due to the tube's presence can lead to difficulties in coordinating the complex process of swallowing. This disruption to the swallowing mechanism can increase the risk of aspiration, in which ingested material enters the airway instead of the stomach, potentially causing pneumonia and other complications [3]. As a result, patients with NG tubes require vigilant monitoring of their swallowing function to detect any signs of dysphagia or aspiration, ensuring their well-being and overall quality of life [4].

Speech-language pathologists primarily rely on videofluoroscopy swallowing studies (VFSS) to evaluate the swallowing process, identify deficits, trial interventions, and formulate treatment plans. VFSS involves capturing continuous X-ray images of the neck area while participants ingest barium, providing a real-time X-ray video that offers comprehensive insights into both spatial and temporal aspects of swallowing function. Numerous research studies have explored the efficacy of this method and have shown that trained clinicians and automated computer systems can extract valuable clinical information from VFSS videos, in adult and pediatric populations [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16].

Unfortunately, VFSS has inherent limitations for patients and clinical staff including radiation exposure, limited availability, high costs, and subjectivity in interpretation. Patients with complicated swallowing, such as those with NG tubes, often require multiple VFSS. There has been a growing demand for an alternative approach to swallowing assessment that offers greater accessibility, cost-effectiveness, and minimum health risks [17], [18], [19], [20]. In recent years, processing HRCA signals has emerged as an encouraging method that aligns with these desirable attributes. HRCA signals represent swallowing sounds and vibrations, and their non-invasive nature involves the placement of a sensitive microphone and a triaxial accelerometer on the anterior neck. The resulting processed signals offer insight into swallowing anatomy and physiology without the need for VFSS. As of today, this test has established itself as a sought-after technique for swallowing analysis, as evidenced by a collection of studies [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35]. The outcomes reported in these studies generally correspond with the results of the human manual kinematic analysis in videofluoroscopic images, thereby reinforcing the premise that HRCA holds the potential to serve as a viable alternative for X-ray images and other established methods in the field of swallowing research.

While HRCA signals have shown promise in swallowing studies across various patient populations (e.g., stroke, lung transplant) and healthy individuals, further exploration is needed to establish this approach in patients with altered swallowing anatomy and physiology (e.g., those with laryngectomies), before incorporating it into clinical practice. This study aimed to explore the utility of HRCA signals for evaluating swallowing in patients with NG tubes given the altered anatomy and physiology resulting from the introduction of a tube into the aerodigestive tract. The goal was to contribute further evidence supporting the validity, reliability, and clinical applicability of HRCA in healthcare settings. To this end, we investigated several kinematic events in swallowing, including detection of upper esophageal sphincter opening duration, detection of laryngeal vestibule closure duration, and hyoid bone tracking. These kinematics were selected for their critical role in ensuring safe and efficient swallowing, as well as their susceptibility to restricted movement caused by the presence of the NG tube.

Recent years have witnessed a transformative influence of machine learning methodologies across the domains of medicine and healthcare. These advanced techniques have also found applications in the diagnosis of swallowing and neck-related disorders, offering encouraging avenues for enhanced healthcare outcomes. A notable study conducted by a group of oral surgeons and oncologists [36] reviewed a diverse set of clinical studies employing artificial intelligence for neck cancer detection. They demonstrated that investing in machine learning modeling is a worthwhile endeavor in this field. Another pioneering investigation explored the diagnosis of cervical spondylosis using deep learning techniques [37]. Various models, including convolutional neural networks, K-nearest neighbors, and real-time object detection algorithms, were used to analyze swallowing videos. Among these architectures, the convolutional network emerged as the top performer, achieving an impressive 88% accuracy and 92% precision. It is worth noting that some of our previous studies have also focused on image-based analyses, including bolus detection through mask-RCNN object detection method [5], as well as the estimation of cervical vertebrae length and angle [6].

Research in the field of HRCA signal analysis and machine learning has revealed connections between swallowing vibratory and acoustic signals and key swallowing events, such as upper esophageal sphincter opening, laryngeal vestibule closure, and hyoid bone tracking [38], [39], [40]. These research endeavors have introduced novel neural network architectures to analyze swallowing patterns in patients without NG tubes and healthy participants. To gain insights into HRCA signals recorded from NG tube patients, we fine-tuned deep learning models developed in previous research efforts. This deliberate choice sought to explore the generalizability of these architectures to an unseen data category encompassing patients with altered anatomy and physiology due to the use of the tube.

To date, there has been no research investigating how effective HRCA signals are for assessing swallowing in individuals using feeding tubes. This study filled this gap, providing a foundation for future exploration and potential advancements in patient care. We present possibilities in facilitating early diagnosis of swallowing impairments in this population, contributing to significant enhancements in their overall quality of life [41].

II. METHODOLOGY

A. PARTICIPANTS, STUDY PROCEDURES, AND EQUIPMENT

We employed machine learning architectures proposed in [38], [39], and [40] to investigate upper esophageal

IEEE Journal of Translational Engineering in Health and Medicine

TABLE 1. Demographic data of participants included in the target swallowing kinematic analyses.

Dataset	# of swallows	# of Participants	Age Range	Sex
Upper Esophageal Sphincter Opening Duration				
Training	1340	144	19-94 62.86 $\pm$ 15.42	82 Male 62 Female
Test	188	18	41-84 61.72 $\pm$ 10.86	15 Male 3 Female
Laryngeal Vestibule Closure Duration				
Training	588	102	19-94 63.23 $\pm$ 16.07	60 Male 42 Female
Test	159	20	41-84 62.45 $\pm$ 10.53	16 Male 4 Female
Hyoid Bone Tracking				
Training	1168	157	19-94 63.97 $\pm$ 14.49	98 Male 59 Female
Test	83	21	44-84 63.95 $\pm$ 9.83	16 Male 5 Female

sphincter opening duration, laryngeal vestibule closure duration, and hyoid bone tracking in patients using NG tubes. In essence, our objective was to assess the performance of these established structures on a different category of swallowing data. This exploration becomes crucial because, in real-world applications, we may encounter new categories of participants that were not part of the initial training dataset. Given the practical constraints of allocating resources for frequent model retraining, we aim to understand the extent to which our model can generalize to unseen data and evaluate its performance in handling variations that may arise in different patient categories. This investigation is pivotal for determining the robustness and applicability of our model in diverse clinical scenarios.

To achieve this aim, samples from participants without NG tubes were considered for the training, and the trained models were then applied to samples collected from people with NG tubes. Protocols of these studies were duly reviewed and approved by the Institutional Review Board of the University of Pittsburgh (IRB number [12080498]), ensuring compliance with ethical standards. Additionally, all participants provided informed consent before taking part in the data collection process including the concurrent collection of their HRCA data during the VFSS procedure, as illustrated in Figure 1. Further details on the demographic information of participants can be found in Table 1.

Participants who were referred for advanced examination due to swallowing concerns underwent VFSS in the lateral plane. During the test, they swallowed Varibar thin liquid (Bracco Diagnostics Inc., Monroe Township, NJ) in the course of routine clinical care. Videos were captured using an Ultimax system (Toshiba, Tustin, CA) with a pulse rate of 30 pulses per second and a frame resolution of H1008 * W792. VFSS output stream was captured via an AccuStream Express HD video card (Foresight Imaging, Chelmsford, MA) and digitized with a sampling rate higher than or equal to 60 frames per second [21], then down-sampled to 30 frames per second to remove duplicate frames. The digital video stream was saved to a hard disk using LabView's Signal Express (National Instruments, Austin, Texas). Additionally, participants consented to the concurrent collection of their HRCA data during the VFSS procedure, as illustrated in Figure 1.

Ensuring the robustness and generalization of our HRCA system is crucial for demonstrating its performance across a range of challenging clinical scenarios, including those involving variations in data quality, noise, and missing data. In environments where VFSS may face challenges such as poor video signal quality, HRCA's ability to capture essential physiological swallowing signals proves invaluable. HRCA leverages signal-based technology that detects muscle activity and swallowing sounds, which are less susceptible to visual obstructions such as poor lighting or rapid patient movements that commonly impair VFSS.

The strategic placement of HRCA sensors ensures accurate capture of critical physiological signals related to swallowing. These sensors are not only sensitive to physiological parameters but also equipped with advanced signal processing techniques like filtering and noise cancellation. This setup minimizes the impact of extraneous noise, such as patient movement or coughing, which are prevalent during assessments and can obscure important data in video-based systems.

To mitigate the impact of missing data, we capture multiple swallows from each participant. This redundancy ensures that even if data from one or two swallows are missing, we can reliably assess the participant's health status using the information gathered from the remaining swallows. Consequently, this method allows us to overlook the incomplete swallows without compromising the overall assessment, thereby preserving the integrity and utility of our dataset. This strategy enhances the robustness and reliability of our findings, ensuring that our conclusions about each participant's health are well-supported by a comprehensive collection of data points.

For HRCA data collection, a contact microphone (model C411L, AKG, Vienna, Austria) and a triaxial accelerometer (ADXL 327, Analog Devices, Norwood, Massachusetts) were positioned on the anterior aspect of the larynx, ensuring

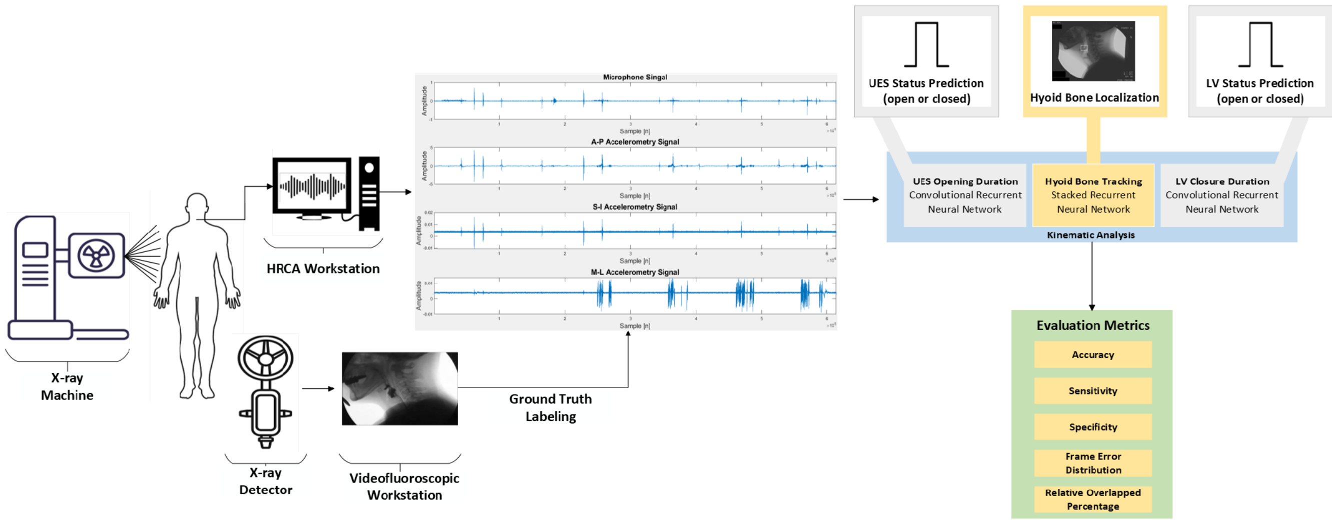

FIGURE 1. The concurrent collection of videofluoroscopic images and HRCA signals from study participants, along with the proposed pipeline for kinematic analysis on swallowing signals.

optimal alignment with the participant's neck. This positioning was carefully chosen to capture unobstructed videos and high-quality signals, as validated in prior studies [21]. The accelerometer effectively recorded vibratory signals at a frequency of 20 KHz across three accelerometry directions: anterior-posterior, superior-inferior, and medial-lateral.

To predict the duration of upper esophageal sphincter opening, 188 thin liquid swallows from 18 participants using NG tubes were considered as the test set. Each patient with an NG tube contributed an average of 10.44 swallows ($\pm$ 3.33), ranging from 5 to 18. The training data for this study encompassed 1,340 swallows from 144 subjects without an NG tube. Transitioning to the prediction of laryngeal vestibule closure duration, we used a test dataset consisting of 159 thin liquid swallows from 20 patients with NG tubes. Each patient with an NG tube contributed an average of 7.95 swallows ($\pm$ 4.25), with a range from 1 to 18. The training dataset consisted of 588 swallows collected from 102 participants without an NG tube. For the hyoid bone tracking study, the training dataset included 1,168 swallows collected from 157 subjects without an NG tube. For this study, we analyzed 83 thin liquid swallows from 21 patients using NG tubes as the test data. On average, each patient contributed approximately 3.95 swallows ($\pm$ 2.92), with counts ranging from 1 to 9. The differences among the training datasets, as well as among the test datasets for these three tasks, stem from various factors such as head movement, technical limitations, and subjective judgment. These factors contribute to human error during the annotation of kinematic events, leading to variability and discrepancies within each set of data across the tasks.

B. SWALLOW KINEMATIC ANALYSIS

Videofluoroscopic image analysis was conducted by trained raters specialized in swallowing kinematic analysis, following established standards and protocols [42]. Intra and inter-rater reliability was achieved a priori with intraclass correlation coefficients exceeding .9 and percent exact agreement rates surpassing 80%, demonstrating the consistency and accuracy of the measurements. Inter and intra-rater reliability was maintained throughout the swallowing kinematic event annotation process on 10% of the swallows, with ICCs over .9, indicating a high level of agreement. While conducting swallowing kinematic measurements, each of the three trained raters randomly selected one swallow every 10 for another rater to reevaluate. This process aimed to ensure consistent inter-rater reliability within an acceptable margin of human error (3-frame tolerance) and achieve an intraclass correlation coefficient of 1.00 for each event. Additionally, each rater re-rated a random sample of 10% of the swallows they had already labeled, maintaining an intraclass correlation coefficient of 1 to establish inter-rater reliability for swallowing kinematic measurements.

To identify the kinematic events of interest, raters pinpointed the following events for each swallow:

- The initial event for hyoid tracking is the first noticeable superior and/or anterior motion of the hyoid bone at the start of the swallow. During the data annotation phase, raters marked two specified landmarks on the hyoid bone in each frame of the swallow.
- Upper esophageal sphincter opening is denoted by the moment in which the upper esophageal sphincter initially opens. The opening can be detected just before the bolus enters, indicated by white airspace, or when the bolus of barium contrast pushes the anterior and poster walls of the upper esophageal sphincter apart and enters the space.
- Upper esophageal sphincter closing is denoted by the initial frame in which there is no visible column of air or barium contrast between the posterior and anterior walls of the upper esophageal sphincter, signifying its closure.

- Laryngeal vestibule closure is defined as the moment in which the laryngeal vestibule is initially sealed, characterized by the absence of barium or airspace within the laryngeal inlet.
- Laryngeal vestibule reopening corresponds to the initial frame in which the laryngeal vestibule reopens, and airspace becomes detectable as the epiglottis returns to its resting position [43].

C. PREPROCESSING OF HRCA SIGNALS

HRCA signals captured by the microphone and the triaxial accelerometer were susceptible to environmental and patient-related noises, potentially causing distortions. To ensure the reliability of these signals as inputs for machine learning systems, it was essential to eliminate unwanted effects. This process began by applying a bandpass filter to attenuate both low-frequency and high-frequency noise components, followed by signal amplification (model P55, Grass Technologies, Warwick, Rhode Island). Signals were then digitized using a data acquisition device (National Instruments 6210 DAQ) facilitated by the SignalExpress program in LabVIEW (Signal Express, National Instruments, Austin, Texas), operating at a sampling rate of 20 kHz. Subsequently, data was down-sampled from its original rate of 20 kHz to 4 kHz, a strategic adjustment aimed at mitigating transient noise components.

In the following step, we modeled the baseline output generated by the attached sensors using an autoregressive process. Sensor-specific finite impulse response filters were developed to target and remove this noise from both acoustic and vibratory inputs. Addressing motion artifacts, that could arise due to voluntary or involuntary participant movement during data collection and contribute to low-frequency noise, a fourth-order least squares spline approximation was applied. As a final step in the preprocessing pipeline, wavelet denoising was employed to further enhance the overall quality of the signals, resulting in a refined dataset ready for subsequent analysis. To segment continuous swallowing signals into individual swallows, we then used an automatic segmentation algorithm introduced in [23]. This algorithm operates directly on the raw swallowing signals in an online fashion.

D. ESTIMATION OF UPPER ESOPHAGEAL SPHINCTER OPENING DURATION

Employing the model presented in [38], we estimated the upper esophageal sphincter opening duration in HRCA signals collected from patients with NG tubes. The model, trained on HRCA signals from patients without NG tubes and healthy participants, was applied to evaluate its performance on the test data. The output was encoded into binary sequences, where "1" represented the upper esophageal sphincter's open status, and "0" represented its closed status. The desired output, representing the duration of the upper esophageal sphincter opening, was calculated considering the period during which the output was "1".

Swallowing events occur sequentially, with temporal dependencies evident in swallowing data [44]. Given our objective to identify temporal features within each single swallow, the design of the proposed neural network should be tailored to handle time-series data. Recurrent neural networks are state-of-the-art for sequential data analysis, although they have some complications. These networks can be computationally slow and challenging to train when dealing with long sequences. Given the typical duration of a single swallow (1-2 seconds) and the high sampling frequency for data recording, each swallow generated more than 4000 sampling points, making it a lengthy sequence for recurrent neural networks. Training a network with such long sequences can be computationally expensive. To address this, one potential solution was to extract important features from the input data to reduce sequence length and then feed these shortened sequences to the network. However, to date, there has been no study focusing on feature importance identification within HRCA signals for predicting kinematic events. Instead of manual feature extraction, convolutional neural networks were employed for this purpose. Convolutional neural networks are considered a popular choice for feature extraction in both signal and image data. A schematic diagram illustrating the different components of the proposed model is shown in Figure 2.

E. ESTIMATION OF LARYNGEAL VESTIBULE CLOSURE DURATION

The convolutional recurrent neural network introduced in [39] was fine-tuned to estimate laryngeal vestibule closure duration in HRCA signals collected from patients using feeding tubes. The model was trained on HRCA signals from cases without NG tubes to assess its performance on NG tube swallows.

In estimating the duration of laryngeal vestibule closure, HRCA signals served as the model's input. The output data were encoded into binary sequences, where an open and closed laryngeal vestibule were represented as "0" and "1", respectively. The duration of laryngeal vestibule closure was calculated according to the window in which the output was "1". The proposed model comprised a hybrid neural network consisting of a 2-layer convolutional neural network as its first component, followed by a 2-layer bidirectional gated recurrent unit to access both past and future data in each time step. Fully connected layers served as the last component of this hybrid model. The dense layers predicted the laryngeal vestibule closure duration by determining its status in each data frame.

While processing swallowing signals using the implemented model, a window slid through each sample, and the outputs from the windows were aggregated to make sequential predictions. Each window had a duration of 0.33 seconds and could either overlap with the next window or not.

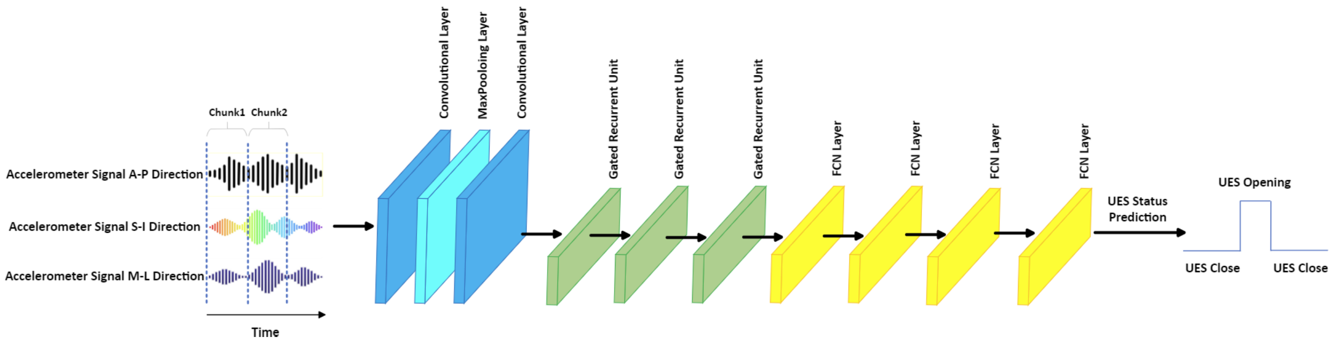

FIGURE 2. Structure of the convolutional recurrent neural network designed for predicting upper esophageal sphincter opening duration.

Overlapping windows had an overlap ratio greater than 0, while non-overlapping windows had an overlap ratio of 0. To determine the optimal approach, the overlap ratio was varied from 0 to 50%, and the model's performance was evaluated in each scenario. Once the model learned the common patterns associated with laryngeal vestibule closure and reopening frames, its performance was assessed using signals recorded from patients with NG tubes.

F. HYOID BONE TRACKING

For hyoid bone tracking, we employed the stacked recurrent neural network proposed in [40]. This network consisted of four hidden layers, each containing 64 neurons. The model's input comprised feature vectors extracted from sensor signals, and its output indicated the position of the hyoid bone at time t. To prepare input data, each swallowing signal was segmented into slices, with the mean and variance calculated for each slice. The corresponding input vector of each swallow was defined based on sensor movements and extracted features. The ground truth outputs of the network were generated according to the videofluoroscopic images. This process included determining the hyoid bone position marked on the image frames using the C2-C4 axis.

The performance of the model in hyoid bone tracking was evaluated using a metric called relative overlapped percentage. This metric allows a comparison between the model and human rater annotation. A higher value for relative overlapped percentage is more desirable, with 100% indicating that the model's tracking output matches that of the human rater. Figure 4 illustrates both human-labeled and predicted hyoid body positions for various overlapping percentages.

III. RESULTS

A. PREDICTION OF UPPER ESOPHAGEAL SPHINCTER STATUS

The proposed model underwent training using 1,340 non-NG tube swallows for over 100 epochs and was subsequently tested on 188 previously unseen NG tube signals. Table 2 presents details on the model's performance in predicting upper esophageal sphincter opening duration for both training and test datasets.

TABLE 2. Evaluation of the proposed convolutional recurrent neural network for upper esophageal sphincter status prediction.

Metric	Training	Test
Accuracy	97.12%	88.91%
Sensitivity	96.92%	88.96%
Specificity	97.48%	91.20%

Figure 5 illustrates the frame error distribution for the upper esophageal sphincter opening and closure. The model predicted the onsets of upper esophageal sphincter opening and closure in 67.61% and 82.95% of NG tube swallows, respectively, falling within an acceptable range of human error in the analysis of swallowing kinematics (3-frame tolerance).

B. PREDICTION OF LARYNGEAL VESTIBULE STATUS

The model underwent training using 588 non-NG swallows for over 100 epochs and was subsequently evaluated on 159 unseen NG swallows. The model's performance is reflected in Table 3 for varying overlapping ratios ranging from 0% to 50%.

Variation in the overlap ratio proved beneficial in model performance evaluation, allowing for the selection of the optimal value for this hyperparameter. When the overlap ratio was configured at 50%, all evaluation metrics for both training and test datasets exceeded 80%. With an overlapping ratio set at 50%, Figure 6 illustrates the distribution of frame errors in predicting the onset of laryngeal vestibule closure and its reopening. The model succeeded in predicting the onset of laryngeal vestibule closure and its reopening within a tolerable range of human error in swallowing kinematic analysis for 79.62% and 75.80% of NG tube swallows.

C. EVALUATION OF HYOID BONE TRACKING ALGORITHM

The model suggested for hyoid bone tracking underwent training on 1,168 HRCA signals collected from patients without NG tubes and healthy participants for 10^6 epochs and was subsequently tested on 83 unseen NG tube signals. For each sample in the test set, we computed the relative overlapped percentage as the evaluation metric. The average relative overlapped percentage achieved by the model was

IEEE Journal of Translational Engineering in Health and Medicine

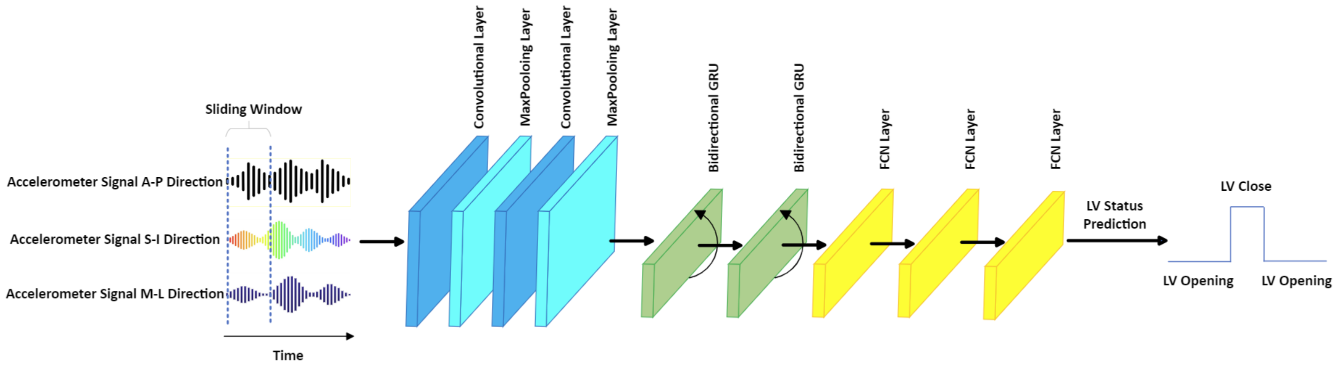

FIGURE 3. Structure of the proposed convolutional recurrent neural network for predicting laryngeal vestibule closure duration.

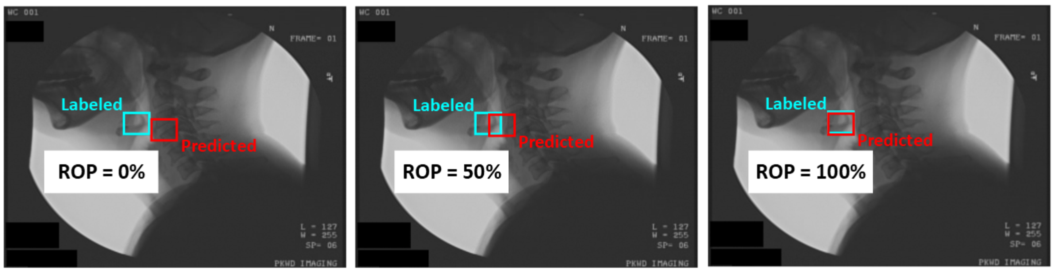

FIGURE 4. An illustration depicting the hyoid body as labeled by humans and predicted by the model at relative overlap percentages of 0%, 50%, and 100%.

TABLE 3. Evaluation of the proposed convolutional recurrent neural network for laryngeal vestibule status prediction.

Metric	Training	Test
0R = 0%		
Accuracy	82.22%	81.01%
Sensitivity	79.89%	77.14%
Specificity	83.49%	82.77%
OR = 10%		
Accuracy	82.78%	80.49%
Sensitivity	79.66%	78.78%
Specificity	84.42%	81.27%
OR = 20%		
Accuracy	82.38%	81.10%
Sensitivity	78.55%	75.66%
Specificity	84.46%	83.58%
OR = 30%		
Accuracy	83.05%	81.08%
Sensitivity	80.51%	79.40%
Specificity	84.49%	81.84%
OR = 40%		
Accuracy	83.35%	81.13%
Sensitivity	79.39%	74.02%
Specificity	85.47%	84.36%
OR = 50%		
Accuracy	83.68%	81.35%
Sensitivity	81.83%	81.12%
Specificity	84.81%	81.45%

41.27%, indicating that the stacked recurrent neural network detected at least 41.27% of the hyoid bone's body in the test frames. When comparing this performance to the average relative overlapped percentage reported in [40], which stood at 51.61%, it is important to note that this study employed swallowing data from distinct groups of participants with different health conditions for the training and evaluation processes. Despite this complexity, the model's relative overlapped percentage of 41.27% represents a promising and noteworthy achievement in this study.

IV. DISCUSSION

In the realm of modern healthcare, the significance of machine learning in processing medical data is undeniable. Expanding upon established neural network models from previous research on patients without NG tubes and healthy individuals, our study focused on analyzing swallowing kinematics using HRCA signals from patients with NG tubes. This approach accounts for the significant alterations in swallowing anatomy and physiology caused by the presence of the tube. This endeavor sought to assess the applicability of existing architectures to an uncharted data category within this domain, thereby demonstrating the potential for this technology to be extended to patients with NG tubes. By training models on HRCA signals from individuals without NG tubes and subsequently evaluating them on NG tube samples, we achieved notable outcomes. Our predictions for key swallowing events, such as upper esophageal sphincter opening and closure, demonstrated accuracy of 67.61% and 82.95%, respectively, within a 3-frame error tolerance. Similarly, laryngeal vestibule closure and reopening were identified in 79.62% and 75.80% of the test samples, respectively, while

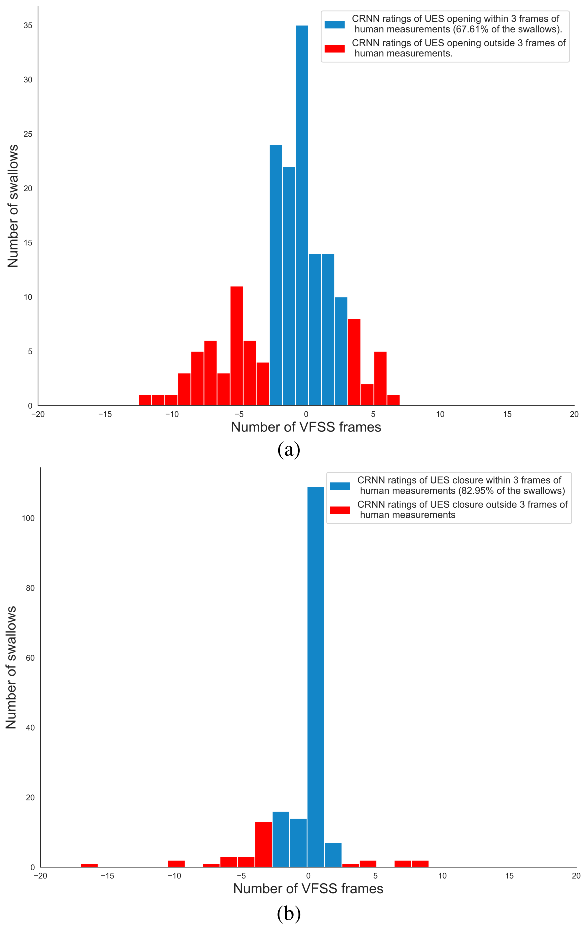

FIGURE 5. The frame error distribution for prediction of upper esophageal sphincter opening duration in the test set is illustrated in (a) and (b), showcasing the error distribution for predicting the onset of upper esophageal sphincter opening and its closure. The blue bars correspond to errors of no more than 3 frames.

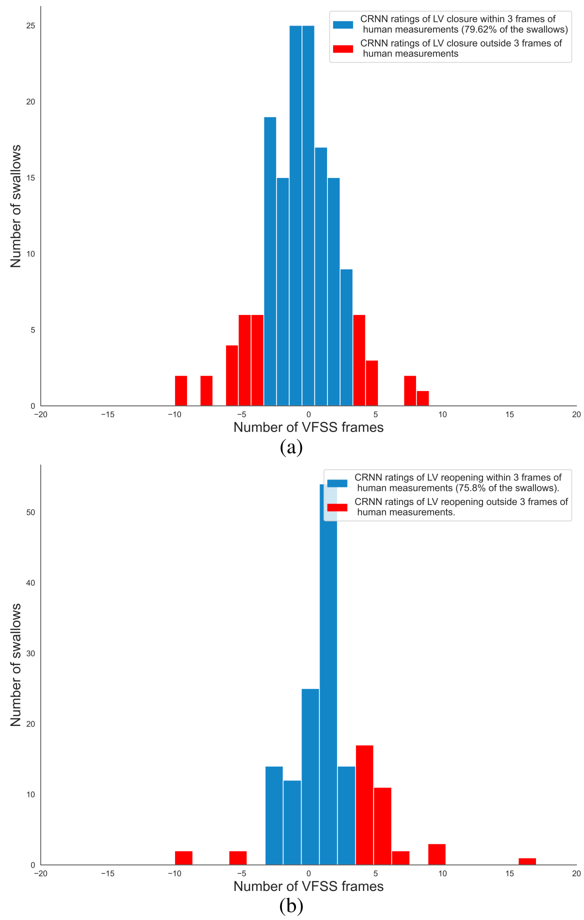

FIGURE 6. The frame error distribution for prediction of laryngeal vestibule closure duration in the test set is illustrated in (a) and (b), showcasing the error distribution for predicting the onset of laryngeal vestibule closure and its reopening. The blue bars correspond to errors of no more than 3 frames.

hyoid bone tracking yielded a relative overlapped percentage of 41.27% in the target population.

Our previous work has shown that HRCA is feasible for a diverse range of patient populations. The results from this study are a promising step toward using HRCA signals for swallowing kinematic analysis in patients with NG tubes. The potential to use HRCA signals instead of VFSS in the future offers patient with varied anatomy and comorbidities an accessible, non-invasive, cost-effective, and objective method for assessing swallowing. By extending the application of machine learning to the novel cohort of patients with NG tubes, we not only facilitate wider access to dysphagia assessment but also advocate for a paradigmatic reevaluation of diagnostic methodologies. Although our findings propel us closer to adopting HRCA signal analysis in clinical settings, further investigations are needed. Larger sample sizes and a broader range of patient populations are warranted to fully comprehend the complexities of swallowing mechanics, and we must address interpretability challenges posed by machine learning models before HRCA can be widely adopted.

Future research endeavors should prioritize extending HRCA signal analysis to include larger, diverse patient cohorts, thereby improving the applicability of findings across a wide range of health conditions associated with dysphagia. While our research team has already explored various groups, including patients with neurodegenerative diseases, stroke survivors, and individuals with lung transplants, we have yet to investigate populations with anatomical differences, such as those who have undergone radiation for head and neck cancer, patients with laryngectomies, or individuals with tracheostomies.

There is also a compelling need to advance research to encompass a wider range of swallowing events. While the opening and closing of the upper esophageal sphincter and laryngeal vestibule, and hyoid tracking are crucial components of swallowing, many significant swallowing kinematics remain unexplored and require further training of HRCA. The development of multitask models capable of predicting multiple kinematic events within a unified framework represents another intriguing and necessary avenue for exploration. Since clinicians evaluate multiple kinematic events simultaneously during a VFSS, a multitask model would better replicate real-world conditions, enhancing both resource efficiency and practical implementation. Furthermore, to build trust among medical professionals in the effectiveness of the HRCA system, it is essential to conduct extensive clinical trials. These trials should compare the HRCA's performance directly with human assessments, demonstrating its reliability and accuracy. Addressing these challenges holds the promise of enhancing the utility of HRCA signal analysis in the realm of swallowing assessment, thereby marking a significant stride toward more inclusive and efficient clinical practices.

V. CONCLUSION

This study underscored the potential of HRCA signal analysis as a practical tool for swallowing assessment and dysphagia diagnosis, offering a viable alternative to VFSS and other well-established swallowing tests. By demonstrating the feasibility of using HRCA signals for swallowing kinematic analysis in patients with an NG tube, we have opened up a new avenue for comprehensive practices in this field. Moreover, fine-tuning of machine learning architectures from prior research illustrated how HRCA data analysis for swallowing evaluation can extend to various cases with different health conditions, paving the way for more robust and adaptable diagnostic tools. We hope our exploration in this study and the call for further investigations into the interpretability of HRCA signals can signify promising directions for future research, reflecting a progressive evolution in dysphagia assessment methodologies.

REFERENCES

[1] K. Mamun and J. Lim, "Role of nasogastric tube in preventing aspiration pneumonia in patients with dysphagia," *Singap. Med. J.*, vol. 46, no. 11, p. 627, 2005.

[2] W. Rao, X. Zhang, J. Zhang, R. Yan, Z. Hu, and Q. Wang, "The role of nasogastric tube in decompression after elective colon and rectum surgery: A meta-analysis," *Int. J. Colorectal Disease*, vol. 26, pp. 423–429, Nov. 2011.

[3] S. Gillson. *How the Upper Esophageal Sphincter Plays a Role in Acid Reflux Pain*. Verywell Health. Accessed: Nov. 12, 2023. [Online]. Available: https://www.verywellhealth.com/upper-esophageal-sphincter-1742526

[4] *Dysphagia (swallowing problems)*. NHS choices, Accessed: Nov. 12, 2023. [Online]. Available: https://www.nhs.uk/conditions/swallowing-problems-dysphagia/

[5] H. Caliskan, A. S. Mahoney, J. L. Coyle, and E. Sejdić, "Automated bolus detection in videofluoroscopic images of swallowing using mask-RCNN," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Montreal, QC, Canada, Jul. 2020, pp. 2173–2177.

[6] Z. Zhang, S. Mao, J. Coyle, and E. Sejdić, "Automatic annotation of cervical vertebrae in videofluoroscopy images via deep learning," *Med. Image Anal.*, vol. 74, Dec. 2021, Art. no. 102218.

[7] A. S. Mahoney, Y. Khalifa, E. Lucatorto, E. Sejdić, and J. L. Coyle, "Cervical vertebral height approximates hyoid displacement in videofluoroscopic images of healthy adults," *Dysphagia*, vol. 37, no. 6, pp. 1689–1696, Dec. 2022.

[8] M. G. Rugiu, "Role of videofluoroscopy in evaluation of neurologic dysphagia," *Acta Otorhinolaryngol. Italica*, vol. 27, no. 6, p. 306, 2007.

[9] M. L. Splaingard, B. Hutchins, L. Sulton, and G. Chaudhuri, "Aspiration in rehabilitation patients: Videofluoroscopy vs bedside clinical assessment," *Arch. Phys. Med. Rehabil.*, vol. 69, no. 8, pp. 637–640, 1988.

[10] G. J. Bours, R. Speyer, J. Lemmens, M. Limburg, and R. De Wit, "Bedside screening tests vs. videofluoroscopy or fibreoptic endoscopic evaluation of swallowing to detect dysphagia in patients with neurological disorders: Systematic review," *J. Adv. Nursing*, vol. 65, no. 3, pp. 477–493, 2009.

[11] A. Ysunza, M. C. Pamplona, J. M. Ortega, and H. Prado, "Video fluoroscopy for evaluating adenoid hypertrophy in children," *Int. J. Pediatric Otorhinolaryngol.*, vol. 72, no. 8, pp. 1159–1165, Aug. 2008.

[12] M. P. Hiorns and M. M. Ryan, "Current practice in paediatric videofluoroscopy," *Pediatric Radiol.*, vol. 36, pp. 911–919, Mar. 2006.

[13] J. A. Logemann, A. W. Rademaker, B. R. Pauloski, Y. Ohmae, and P. J. Kahrilas, "Normal swallowing physiology as viewed by videofluoroscopy and videoendoscopy," *Folia Phoniatrica et Logopaedica*, vol. 50, no. 6, pp. 311–319, 1998.

[14] A. M. Kelly, P. Leslie, T. Beale, C. Payten, and M. Drinnan, "Fibreoptic endoscopic evaluation of swallowing and videofluoroscopy: Does examination type influence perception of pharyngeal residue severity?" *Clin. Otolaryngol.*, vol. 31, no. 5, pp. 425–432, 2006.

[15] D. Kidd, J. Lawson, R. Nesbitt, and J. MacMahon, "Aspiration in acute stroke: A clinical study with videofluoroscopy," *Int. J. Med.*, vol. 86, no. 12, pp. 825–829, 1993.

[16] G. L. Re et al., "Swallowing evaluation with videofluoroscopy in the paediatric population," *Acta Otorhinolaryngol. Italica*, vol. 39, no. 5, p. 279, 2019.

[17] E. Boaden, J. Nightingale, C. Bradbury, L. Hives, and R. Georgiou, "Clinical practice guidelines for videofluoroscopic swallowing studies: A systematic review," *Radiography*, vol. 26, no. 2, pp. 154–162, May 2020.

[18] A. K. Vose, S. Kesneck, K. Sunday, E. Plowman, and I. Humbert, "A survey of clinician decision making when identifying swallowing impairments and determining treatment," *J. Speech, Lang., Hearing Res.*, vol. 61, no. 11, pp. 2735–2756, Nov. 2018.

[19] J.-Y. Hong, N.-K. Hwang, G. Lee, J.-S. Park, and Y.-J. Jung, "Radiation safety in videofluoroscopic swallowing study: Systematic review," *Dysphagia*, vol. 36, pp. 73–82, Apr. 2021.

[20] V. J. Earl and M. K. Badawy, "Radiation exposure to staff and patient during videofluoroscopic swallowing studies and recommended protection strategies," *Dysphagia*, vol. 34, no. 3, pp. 290–297, Jun. 2019.

[21] J. M. Dudik, J. L. Coyle, and E. Sejdić, "Dysphagia screening: Contributions of cervical auscultation signals and modern signal-processing techniques," *IEEE Trans. Human-Mach. Syst.*, vol. 45, no. 5, pp. 465–477, Aug. 2015.

[22] J. M. Dudik, A. Kurosu, J. L. Coyle, and E. Sejdić, "A statistical analysis of cervical auscultation signals from adults with unsafe airway protection," *J. Neuroeng. Rehabil.*, vol. 13, p. 7, Jan. 2016.

[23] Y. Khalifa, J. L. Coyle, and E. Sejdić, "Non-invasive identification of swallows via deep learning in high resolution cervical auscultation recordings," *Sci. Rep.*, vol. 10, no. 1, p. 8704, May 2020.

[24] C. Rebrion et al., "High-resolution cervical auscultation signal features reflect vertical and horizontal displacements of the hyoid bone during swallowing," *IEEE J. Transl. Eng. Health Med.*, vol. 7, pp. 1–9, 2019.

[25] 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.

[26] Q. He et al., "The association of high resolution cervical auscultation signal features with hyoid bone displacement during swallowing," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 9, pp. 1810–1816, Sep. 2019.

[27] C. Yu, Y. Khalifa, and E. Sejdić, "Silent aspiration detection in high resolution cervical auscultations," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Chicago, IL, USA, May 2019, pp. 1–4.

[28] J. L. Coyle and E. Sejdić, "High-resolution cervical auscultation and data science: New tools to address an old problem," *Amer. J. Speech-Lang. Pathol.*, vol. 29, no. 2S, pp. 992–1000, 2020.

[29] A. Sabry, A. S. Mahoney, S. Mao, Y. Khalifa, E. Sejdić, and J. L. Coyle, "Automatic estimation of laryngeal vestibule closure duration using high-resolution cervical auscultation signals," *Perspect. ASHA Special Interest Groups*, vol. 5, no. 6, pp. 1647–1656, 2020.

[30] K. Shu, J. L. Coyle, S. Perera, Y. Khalifa, A. Sabry, and E. Sejdić, "Anterior–posterior distension of maximal upper esophageal sphincter opening is correlated with high-resolution cervical auscultation signal features," *Physiol. Meas.*, vol. 42, no. 3, 2021, Art. no. 035002.

[31] Y. Khalifa, C. Donohue, J. L. Coyle, and E. Sejdić, "On the robustness of high-resolution cervical auscultation-based detection of upper esophageal sphincter opening duration in diverse populations," in *Big Data III: Learning, Analytics, and Applications*, vol. 11730. Bellingham, WA, USA: SPIE, Apr. 2021, pp. 148–153.

[32] C. Donohue, Y. Khalifa, S. Mao, S. Perera, E. Sejdić, and J. L. Coyle, "Establishing reference values for temporal kinematic swallow events across the lifespan in healthy community dwelling adults using high-resolution cervical auscultation," *Dysphagia*, vol. 37, no. 3, pp. 664–675, 2022.

[33] P. Frizelle, A.-K. Tolonen, J. Tulip, C.-A. Murphy, D. Saldana, and C. McKean, "The impact of intervention dose form on oral language outcomes for children with developmental language disorder," *J. Speech, Lang., Hearing Res.*, vol. 64, no. 8, pp. 3253–3288, Aug. 2021.

[34] C. Donohue, Y. Khalifa, S. Perera, E. Sejdić, and J. L. Coyle, "Characterizing effortful swallows from healthy community dwelling adults across the lifespan using high-resolution cervical auscultation signals and MBSImP scores: A preliminary study," *Dysphagia*, vol. 37, pp. 1103–1111, Sep. 2021.

[35] R. Schwartz, Y. Khalifa, E. Lucatorto, S. Perera, J. Coyle, and E. Sejdić, "A preliminary investigation of similarities of high resolution cervical auscultation signals between thin liquid barium and water swallows," *IEEE J. Transl. Eng. Health Med.*, vol. 10, pp. 1–9, 2022.

[36] S. Volpe et al., "Machine learning for head and neck cancer: A safe bet?—A clinically oriented systematic review for the radiation oncologist," *Frontiers Oncol.*, vol. 11, Nov. 2021, Art. no. 772663.

[37] M. Sreeraj, J. Joy, M. Jose, M. Varghese, and T. Rejoice, "Comparative analysis of machine learning approaches for early stage cervical spondylosis detection," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 6, pp. 3301–3309, 2022.

[38] Y. Khalifa, C. Donohue, J. L. Coyle, and E. Sejdić, "Upper esophageal sphincter opening segmentation with convolutional recurrent neural networks in high resolution cervical auscultation," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 2, pp. 493–503, Feb. 2021.

[39] S. Mao, A. Sabry, Y. Khalifa, J. L. Coyle, and E. Sejdić, "Estimation of laryngeal closure duration during swallowing without invasive X-rays," *Future Gener. Comput. Syst.*, vol. 115, pp. 610–618, Feb. 2021.

[40] S. Mao, Z. Zhang, Y. Khalifa, C. Donohue, J. L. Coyle, and E. Sejdić, "Neck sensor-supported hyoid bone movement tracking during swallowing," *Roy. Soc. Open Sci.*, vol. 6, no. 7, Jul. 2019, Art. no. 181982.

[41] M. Panebianco, R. Marchese-Ragona, S. Masiero, and D. A. Restivo, "Dysphagia in neurological diseases: A literature review," *Neurol. Sci.*, vol. 41, no. 11, pp. 3067–3073, Nov. 2020.

[42] P. E. Shrout and J. L. Fleiss, "Intraclass correlations: Uses in assessing rater reliability," *Psychol. Bull.*, vol. 86, no. 2, p. 420, 1979.

[43] J. L. Young, P. Macrae, C. Anderson, I. Taylor-Kamara, and I. A. Humbert, "The sequence of swallowing events during the chin-down posture," *Amer. J. Speech-Lang. Pathol.*, vol. 24, no. 4, pp. 659–670, Nov. 2015.

[44] W. J. Dodds, E. T. Stewart, and J. A. Logemann, "Physiology and radiology of the normal oral and pharyngeal phases of swallowing," *Amer. J. Roentgenol.*, vol. 154, no. 5, pp. 953–963, 1990.