

Research on a real-time dynamic monitoring method for silent aspiration after stroke based on semisupervised deep learning: A protocol study

DIGITAL HEALTH
Volume 9: 1–12
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076231183548
journals.sagepub.com/home/dhj



Jia Qiao¹ , Yuan-tong Jiang², Yong Dai³, Yan-bin Gong⁴ , Meng Dai¹, Yan-xia Liu² and Zu-lin Dou¹

Abstract

Objective: This study aims to establish a real-time dynamic monitoring system for silent aspiration (SA) to provide evidence for the early diagnosis of and precise intervention for SA after stroke.

Methods: Multisource signals, including sound, nasal airflow, electromyographic, pressure and acceleration signals, will be obtained by multisource sensors during swallowing events. The extracted signals will be labeled according to videofluoroscopic swallowing studies (VFSSs) and input into a special dataset. Then, a real-time dynamic monitoring model for SA will be built and trained based on semisupervised deep learning. Model optimization will be performed based on the mapping relationship between multisource signals and insula-centered cerebral cortex–brainstem functional connectivity through resting-state functional magnetic resonance imaging. Finally, a real-time dynamic monitoring system for SA will be established, of which the sensitivity and specificity will be improved by clinical application.

Results: Multisource signals will be stably extracted by multisource sensors. Data from a total of 3200 swallows will be obtained from patients with SA, including 1200 labeled swallows from the nonaspiration category from VFSSs and 2000 unlabeled swallows. A significant difference in the multisource signals is expected to be found between the SA and nonaspiration groups. The features of labeled and pseudolabeled multisource signals will be extracted through semisupervised deep learning to establish a dynamic monitoring model for SA. Moreover, strong correlations are expected to be found between the Granger causality analysis (GCA) value (from the left middle frontal gyrus to the right anterior insula) and the laryngeal rise time (LRT). Finally, a dynamic monitoring system will be established based on the former model, by which SA can be identified precisely.

Conclusion: The study will establish a real-time dynamic monitoring system for SA with high sensitivity, specificity, accuracy and F1 score.

¹Department of Rehabilitation Medicine, The Third Affiliated Hospital of Sun Yat-Sen University

²School of Software Engineering, South China University of Technology

³Clinical Medical College of Acupuncture-Moxibustion and Rehabilitation, Guangzhou University of Chinese Medicine

⁴Department of Computer Science and Engineering, The Hong Kong University of Science and Technology

Jiao Qiao and Yuan-tong Jiang contributed equally to this work.

Corresponding authors:

Zu-lin Dou, Department of Rehabilitation Medicine, The Third Affiliated Hospital of Sun Yat-Sen University, No. 600, Tianhe Road, Tianhe District, Guangzhou, Guangdong 510630, China.

Email: douzulin@mail.sysu.edu.cn

Yan-xia Liu, School of Software Engineering, South China University of Technology, Guangzhou Higher Education Mega Centre, Panyu District, Guangzhou 510006, China.

Email: cslyx@scut.edu.cn



Keywords

Real-time dynamic monitoring method, silent aspiration (SA), poststroke dysphagia (PSD), insular cortex, semisupervised deep learning, breathing-swallowing pattern

Submission date: 3 November 2022; Acceptance date: 5 June 2023

Introduction

Aspiration, one of the leading clinical manifestations of poststroke dysphagia (PSD), is characterized by the misdirected inflow of oropharyngeal or gastric contents from the larynx or stomach into the respiratory tract.¹ Aspiration can be classified as silent aspiration (SA) or overt aspiration (OA). Among PSD patients, SA is frequently ignored or misdiagnosed due to the absence of overt symptoms,² with an estimated incidence rate between 51% and 73%.³ Aspiration pneumonia (AP) is a serious complication of SA that can have extremely poor prognoses.⁴ It has been reported that the incidence rate of AP among patients with SA is at least three times as high as the incidence among those with OA and 13 times as high as the incidence among nonaspiration patients.^{3,4} Therefore, SA has been considered an independent risk factor for AP and should be worthy of attention.⁴⁻⁶

The current assessment for SA mainly includes bedside screening tools, clinical evaluations and instrument examinations.⁵ However, some notable shortcomings of these methods should not be ignored. First, low screening sensitivity and specificity limit the application of bedside screening tools in the diagnosis of SA. Second, clinical evaluations heavily rely on well-trained physiotherapists and cannot be applied to achieve real-time dynamic monitoring for SA. Third, although a videofluoroscopic swallowing study (VFSS) is recognized as the gold standard for SA assessment, it still carries the unknown cumulative risk of radiation exposure.⁷

Recently, wearable, noninvasive, convenient, and nonradiative sensors have been applied in clinical assessments of swallowing function. For example, flexible miniature high-precision sensors and biaxial accelerometers have been used to identify aspiration by recording acceleration signals during swallowing.^{8,9} High-resolution cervical auscultation (HRCA) combining a triaxial accelerometer with a high-resolution microphone can precisely measure acceleration and sound signals.¹⁰ In addition, some other important signals, such as airflow, acceleration, sound and electromyography (EMG), can be simultaneously collected and analyzed by a set of combined sensors.^{11,12} However, these methods still have some potential disadvantages: vibration signals from flexible miniature high-precision

sensors are still extracted from a single location, which inevitably leads to a lack of multisource information;^{8,9} HRCA cannot accurately identify SA due to its single-source data acquisition approach;¹⁰ and the combined sensors fail to effectively diagnose SA due to the limitations of data analysis and processing.^{11,12}

To improve the accuracy of SA diagnosis, several machine learning-based dynamic monitoring methods have been proposed in some recent studies. Traditional mode recognition methods utilize prior knowledge of swallowing, breathing and handcrafted features, which may lead to bias and cannot achieve optimal efficacy in real applications.¹² In contrast, machine learning-based methods extract practical features from swallowing data samples and then establish models based on machine learning algorithms, which can serve as effective diagnostic methods for SA. The strong association between some HRCA signal characteristics and penetration aspiration has been verified by building SA detection models with features extracted from the time, frequency and time-frequency domains and various classifiers (including support vector machine and naive Bayes classifiers).¹³ In addition, surface EMG improves the performance of automatic classification models for dysphagia detection.¹⁴⁻¹⁷

Dynamic monitoring models based on deep learning (a mainstream method in machine learning) have been used in the evaluation of swallowing function in recent years.¹⁸ For example, a deep belief network was adopted to classify swallowing signals and showed promising results, indicating that this network has outstanding classification capability for this purpose and can be used to evaluate dysphagia.¹⁹ In addition, thanks to the capacity of modeling sequential events and temporal dependencies in biomedical signals, recurrent neural networks (RNNs) have also been applied in automatic detection and diagnosis for dysphagia, such as acceleration signal detection during swallowing, including hyoid movement, upper esophageal sphincter opening and laryngeal vestibular closure.^{20,21} Thus, the machine learning-based dynamic monitoring technique is without doubt an essential method in the assessment of SA.

Most importantly, the regulatory mechanism of SA can be further explored with optimized real-time dynamic monitoring models. The occurrence of SA is thought to be associated with swallowing-breathing incoordination,²² which

is regulated by the swallowing–breathing coordination center located in the medulla oblongata of the brainstem and other cortical/subcortical neural networks, including the insular cortex, sensorimotor cortex, prefrontal lobe and thalamus.^{23–25} Among these brain areas, the insular cortex may play an important role since it has been observed to be significantly activated during swallowing tasks.^{26–28} The insular cortex may exert its modulating effect on swallowing function by accepting projection fibers from the thalamic nucleus and integrating sensory–motor information.²⁹ Furthermore, the insular cortex is involved in respiratory function regulation as well. For example, Trevizan-Baú et al. once adopted cholera toxin subunit B (CT-B) in the retrograde tracing of neural regulation for breathing. Finally, the authors found plenty of CT-B neurons in the insular cortex.³⁰ In another study, the authors inoculated pseudorabies virus (PRV) in the thyroarytenoid muscle, which participates in swallowing–breathing coordination, and found that the PRV was transferred from the periphery to the swallowing central pattern generator, respiratory swallowing central pattern generator, hypothalamus and finally to the insular cortex.³¹ Therefore, insula-centered cerebral cortex–brainstem functional connectivity may be involved in the occurrence of SA through the regulation of swallowing–breathing coordination.³² However, more powerful evidence is still urgently needed.

Currently, little research has been conducted on the mapping relationship between insula-centered cerebral cortex–brainstem functional connectivity and multisource signals (such as airflow, acceleration, sound and EMG signals) related to SA. In this protocol, we plan to develop a novel real-time dynamic monitoring system for SA based on deep learning that can elucidate this mapping relationship and increase the weight of the signals with the strongest correlation. Some issues associated with the regulatory mechanism of swallowing function may also be solved with the use of the optimized models from this system.

Objective

This study aims to establish a real-time dynamic SA monitoring system based on the mapping relationship for SA between insula-centered cerebral cortex–brainstem functional connectivity and multisource signals. First, multisource signals, including sound, nasal airflow, EMG, pressure and acceleration signals, will be extracted by multisource sensors, including EMG, nasal airflow, microphone, pressure and triaxial accelerometer sensors, during swallowing events. The extracted signals will then be labeled according to a VFSS to form a dataset. Next, a real-time dynamic SA monitoring model will be constructed and trained based on semisupervised deep learning. Model optimization will be performed based on the mapping

relationship between the multisource signals and insula-centered cerebral cortex–brainstem functional connectivity through resting-state functional magnetic resonance imaging (rsfMRI). Finally, a real-time dynamic monitoring system of SA will be established, and its improved sensitivity and specificity will be demonstrated through clinical applications, thus providing more evidence for precise intervention in SA patients after stroke and improving prevention efficacy in clinical settings.

Method

Trial design

The trial is a single-center, nonrandomized, open-label cohort study. The study was approved by the Medical Ethics Committee of the Third Affiliated Hospital of Sun Yat-Sen University (No. 02-192-01) and registered at ClinicalTrials.gov (No. ChiCTR2300068908). The first patient will be included on 1 December 2023. The planned end date of the study is 1 December 2025.

Patients

Individuals with PSD will be voluntarily enrolled after providing informed consent. Then, the individuals will undergo VFSS examinations, multisource signal acquisition, and rsfMRI. Patients will be recruited from the Rehabilitation Department of the Third Affiliated Hospital of Sun Yat-Sen University.

Inclusion and exclusion criteria

The inclusion criteria are defined as follows: (a) age > 18 years; (b) stroke confirmed by MRI/CT; (c) PSD first evaluated by the volume–viscosity swallow test (V-VST) and manifesting as impaired efficacy and/or safety during oropharyngeal swallowing; (d) no cognitive impairment (Montreal Cognitive Assessment Scale ≥ 26 and minimal state examination ≥ 27),^{33,34} no pathological changes in the head and neck, metallic implants in the neck or tracheotomy; (e) the ability to cooperate with the multisource signal extraction and clinical evaluation procedures; and (f) agreement with the research terms and signed informed consent. The exclusion criteria are defined as follows: (a) cognitive impairment; (b) tracheotomy; (c) hypoxemia; (d) serious illness, including pneumonia, active infection and intractable epilepsy; and (f) inability to cooperate with the signal extraction and clinical evaluation procedures.

Outcomes

Sound signals. Sound signals, including swallowing and cough sounds, will be collected by a microphone

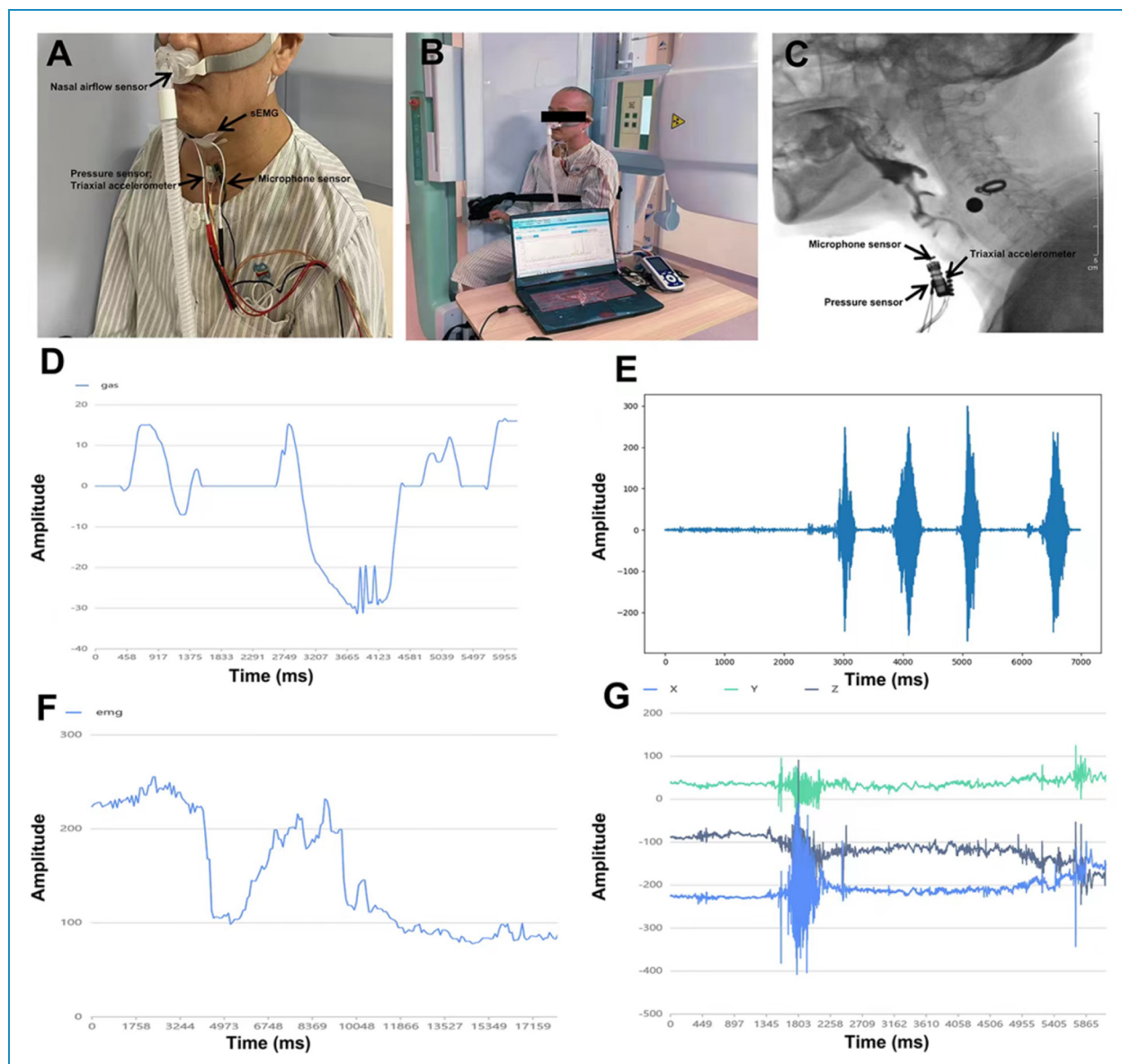


Figure 1. Extraction of multisource signals synchronized with VFSS data. (a–c) Individual with multisource sensors, including those from EMG, nasal airflow, the microphone, pressure and triaxial accelerometer sensors, to extract multisource signals. Example multisource signals: airflow (d), sound (e), EMG signals (f) and acceleration signals (g).

(ECM-LZ1UBMP, Sony, Tokyo, Japan, sample rate: 48 kHz, sensitivity: $-31.0 \text{ dB} \pm 3.0 \text{ dB}$; frequency response: 60 Hz–8 kHz; noise ratio: 68 dB) (Fig. 1A). The microphone will be placed on a point that is on the left of the intersection point between the anterior arch of the cricoid cartilage and midline (the distance is 1 cm).¹⁰ To prevent noisy or mixed signals, the data will be collected in the room used to conduct the VFSS with no murmurs except for the sound of the mechanical operation (Fig. 1B). The extracted sound signals will be saved as *wav* audio files (sampling rate of 48 kHz, approximately 16 bits per sample, monovocal tract). Then, the audio

files $A = \pi r^2$ will be input into MATLAB software, and a preset script will be used for noise reduction, feature extraction and data analysis (Fig. 1E).

Nasal airflow signals. A nasal airflow sensor (AFM3000-200, Guangzhou Aosong Electronics Co. Ltd., China, sample rate: 100 Hz) with a dead-space nasal mask over the nose will be applied to collect the nasal airflow signals, which reflect the respiratory waveform, frequency, rhythm and apnea duration (Fig. 1D, E). Baseline data extraction will be performed prior to the swallowing tasks and consisted of respiratory cycle information, including

exhale duration, inhale duration, apnea duration and breathing–swallowing patterns. The same data will be extracted from patients with SA.

EMG signals. An EMG sensor (LE3100, Shenzhen Dongdixin Technology Co. Ltd., China, sample rate: 8 kHz, resolving power: $0.1 \mu\text{V}$, accuracy: $\pm 2 \mu\text{V}$) will be used to extract EMG signals from the submental muscles (Fig. 1A, F), and the surface electrodes will be fixed around the center of the surface of the submental muscle to detect contractions during swallowing.¹¹

Pressure and acceleration signals. A pressure sensor (PVDF, Analog Devices Inc., MA, USA, sample rate: 8 kHz) will be utilized to collect pressure signals from the thyroid cartilage, which is located at the anterior midline between the thyroid and cricoid cartilage.³⁵ The thyroid cartilage retracts during each swallowing movement and returns to its original position after swallowing. These movements cause the pressure on the pressure sensors to change (Fig. 1A–C).³⁵ A triaxial accelerometer sensor (ADXL, Analog Devices Inc., MA, USA, sample rate: 100 Hz, resolving power: $78 \mu\text{g}/\text{LSB}$) will be utilized to extract acceleration signals from the cricoid cartilage and will be positioned at the anterior midline overlying the arch of the cricoid cartilage (Fig. 1A–C, G).¹⁰

VFSS. The VFSS will be accomplished with a Lanmage dynamic digital radiography machine (Athena Plus 7500; Shenzhen Lanmage Medical Technology Co. Ltd.; Shenzhen, China) during swallowing image acquisition, and the multisource signals will be collected at the same time (Fig. 1C). Each subject will receive 3 ml, 5 ml, or 10 ml of thickened and diluted barium liquid (contrast media: 60% w/v barium sulfate suspension). The viscosity of the liquid will be divided into four levels: lower than 50 mPa.s, in the range of 51–350 mPa.s, in the range of 351–1750 mPa.s, and above 1750 mPa.s.³⁶ A 10-ml injector will be used for bolus administration. The patients will be required to swallow all different volumes and viscosities of boluses, and the specific steps will be performed according to the modified Logemann protocol.³⁷ The VFSS data will be digitally recorded as a video in .avi format using a VFSS acquisition and analysis system (Longest Inc., Guangzhou, China) at 30 frames/s.

rsfMRI. rsfMRI will be adopted to evaluate the mapping relationship between insula-centered functional brain network properties and the different signals. rsfMRI will be conducted with a Siemens Verio 3.0 T scanner (Siemens, Erlangen, Germany) using a 64-channel coil. The T1-weighted magnetization-prepared rapid gradient-echo (MP-RAGE) scan for structural image acquisition will use the following parameters: repetition time (TR) = 2530 ms, echo time (TE) = 2.27 ms, flip angle (FA) = 90° , field of view (FOV) = 256

$\times 256$ mm, matrix = 256×256 , and slice thickness = 1 mm. Functional echo planar images (EPIs) will be acquired in interval sections spanning the entire brain with a gradient–echo sequence with the following parameters: TR = 2000 ms, TE = 30 ms, FOV = 256×256 mm, slice thickness = 3 mm, voxel size = 3 mm \times 3 mm, and FA = 90° .

Sample size

In this prospective study, the sample size refers to the number of swallows. According to previous studies, a total of 3144 swallows from 248 patients with suspected dysphagia are needed.²¹ Each individual repeatedly swallows during the VFSS examination. Early termination of the VFSS will be performed to prevent AP if aspiration (penetration–aspiration scale > 5) occurs, which may result in the loss of 20% to 30% of swallowing samples. Therefore, the study is designed to recruit 350 patients, with 200 subjects in the SA group and 150 subjects in the nonaspiration group.

The sample distribution is defined as follows: (1) A total of 150 individuals are needed in the dataset establishment phase to provide high-quality labeled signals. (2) A total of 50 SA patients are needed in the model optimization phase. (3) A total of 150 individuals are needed in the clinical application phase, with 100 individuals in the SA group and 50 individuals in the nonaspiration group.

Study procedures

Multisource signal extraction and synchronization. Multisource sensors, including microphone, nasal airflow, EMG, pressure and triaxial accelerometer sensors, are flexible and integrable with high sensitivity and high resolving power. These sensors will be applied to collect sound, nasal airflow, EMG, pressure, and acceleration signals during swallowing events. The signal extraction methods are described in Section 4.

To synchronize the multimodal signals, these multisource sensors will be connected to a Raspberry Pi device (4B, Broadcom Corporation). The Raspberry Pi device is a single-board computer based on the Advanced RISC Machine (ARM) architecture. The multiple GPIO (general purpose input/output) pins of the Raspberry Pi are programmed to read the signals from the connected sensors and record the timestamps. During the signal collection, the Raspberry Pi sends the recorded data and timestamps to a computer in real time over the local area network.

High-quality labeling of multisource signals synchronized with the VFSS. First, the interval of effective multisource signals will be determined by the VFSS. The signals will then be labeled by two well-trained speech–language therapists with at least 5 years of clinical experience and imaging technicians according to the VFSS results. The time

points of the oral phase and pharyngeal phase will be labeled in the VFSS video, which will be played back frame by frame at a speed of 30 frames/s by ImageJ open-source software (version 1.42, National Institutes of Health, Bethesda, Maryland). The oral phase refers to the time when the bolus is being processed in the oral cavity to when the bolus is passing the ramus of the mandible, and the pharyngeal phase refers to the time when the bolus is passing the ramus of the mandible to when upper esophageal sphincter relaxation occurs.³⁸ Finally, the multisource signals will be segmented according to the labeled time points. The established dataset includes labeled and unlabeled multisource signals.

Establishment of a real-time dynamic Sa monitoring model based on semisupervised deep learning

Input data preprocessing. Multisource signals are vulnerable to extensive noise interference from internal and external environments, including head motions and large artery pulsations. Therefore, the first step in the signal preprocessing procedure will be to whiten the signals with a finite impulse response (FIR) filter to reinforce decorrelations in the data. The plan is to model the noise for each sensor using an autoregressive model with an order of 10. The order of the autoregressive model will be determined using the Bayesian information criterion. Based on the coefficients of the autoregressive model, a FIR filter will be created. The inverse of the FIR filter will be applied to the collected data to eliminate noise. Next, the acceleration and EMG signals will be denoised. To remove unwanted noise from the acceleration signal, a bandpass filter with a frequency range of 0.1 Hz to 3 kHz will be applied. The acceleration signal will consist of three channels corresponding to the x -, y - and z -axes, with a sampling rate of 200 Hz for the x - and y -axes and a sampling rate of 1600 Hz for the z -axis accelerations.³⁹

The sampling rate of EMG signals will be 2 kHz, and a bandpass filter will be implemented with a cutoff frequency between 20 and 500 Hz. The optimal parameters for the EMG will be as follows: mother wavelet db5, five decomposition levels, soft thresholding and minimax rule for threshold selection.¹⁷

Considering the influence of concurrent events, the input signal will be downsampled to 4 kHz. Furthermore, the amplitudes and frequencies of all the signals will be normalized to eliminate the influence of the signal magnitude.

Location and identification of the swallowing cycle based on an attention mechanism. After the input data are preprocessed, the weight of each source signal will be dynamically learned with an attention mechanism. The acquired timing signal will be divided into frames to obtain the initial signal sequence $\{X_1, X_2, \dots, X_T\}$, where T is the number of frames. The data X_t in the t -th frame contain N source signals $\{x_t^1, x_t^2, \dots, x_t^N\}$, where x_t^s represents the waveform

of the s -th source signal in the t -th frame. By feeding the signal into an RNN, the feature vector h_t^s corresponding to each x_t^s will be obtained, and the feature weights of the source signals can be computed using the attention mechanism, as follows:

$$\alpha_t^s = \frac{\exp(W \cdot x_t^s)}{\sum_{s=1}^5 \exp(W \cdot x_t^s)} \quad (1)$$

where W is a feature fusion parameter, S represents the number of signals and the upper limit of the signal source is 5. Signals that exhibit better performance on the target task will be assigned larger weights, resulting in the feature fusion of multisource timing signals. The feature weights of different source signals will be summed to obtain the fusion feature vector \tilde{X} :

$$\tilde{X}_t = \sum_{s=1}^5 \alpha_t^s h_t^s \quad (2)$$

Finally, a swallowing event classifier will be constructed to classify the multivariate sequence features \tilde{X} and obtain a class probability matrix G .

The temporal boundaries of swallowing events will be determined with a multiscale sliding window. Based on the swallowing recognition result G , a window of length d will be selected, with the t -th frame chosen as the starting point in the signal sequence; thus, the class probability corresponding to the window is $G[t: t + d]$. The swallowing cycle location will be determined when more than 90% of the scores in the window are greater than the threshold σ . To evaluate the localization results, the scores of d frames in the window will be averaged to obtain the confidence score c , and the localization result will be expressed as $(t, t + d, c)$. According to previous results, the window size will be designed with multiple scales, and the window slides on the signal time axis with step size δ . After the signal timing has been determined, the sliding windows of different scales output the target markers detected by various windows. If two windows intersect, the nonmaximum suppression (NMS) method will be applied to select an appropriate window. Finally, the beginning and end positions of each window correspond to the swallowing cycle.

Training and testing of a real-time dynamic SA monitoring model based on semisupervised deep learning. The establishment of a real-time dynamic SA monitoring model is based on the feature fusion of multisource signals and the swallowing cycle location. To address the problems caused by the limited labeled data and class imbalances, a semisupervised deep learning method with pseudolabel generation will be applied to develop the real-time dynamic SA monitoring model.

First, a small number of labeled samples will be sent to a high-performance classifier for training. The classifier is then used to extract SA features to improve the signal recognition accuracy. If unlabeled samples exist, the samples are predicted by the trained high-performance classifier and sorted according to the predicted probability. High-probability samples will then be considered to be reliably labeled and added to the training set. This procedure will be repeated until the number of labeled samples meets the iterative requirements. The new integrated dataset will be used to train a convolutional RNN with stronger classification capability, which finally outputs the SA detection results.

To validate the performance of the proposed method, a 10-fold subject cross-validation will be performed on the collected swallowing samples. The subjects will be first randomly divided into 10 groups, each consisting of approximately 10% of the total samples. In each fold of the cross-validation, one of the 10 groups will be selected as the testing set, while the remaining samples are utilized for training. This process will be repeated 10 times, with each group serving as the validation set once (Fig. 2).

Model optimization based on mapping relationships in functional brain networks

rsfMRI analyses. MATLAB (MathWorks, version 7.14) and RESTPLUS in SPM 12.0 will be used for rsfMRI data preprocessing.

Amplitude of low-frequency fluctuation (ALFF): After the preprocessed data are bandpass filtered (0.01-0.08 Hz), RESTPLUS software (<http://www.restfmri.net/forum/REST-GCA>) will be used to calculate the ALFF of each voxel in the whole brain, and anatomical automatic labeling (AAL) will be used to mark the bilateral insula on the template as seed points. Then, analyses of differences in the ALFF values in the insular region of interest (ROI) between the SA and non-aspiration groups will be performed to determine the relationship between this brain region and SA.

Granger causality analysis (GCA): GCA will be used to calculate the direction and strength of functional connections between cortical regions and the insular cortex. Using the bilateral insula in the AAL template as seed points for GCA, voxel-level bidirectional GCA analyses of the whole brain will be implemented via RESTPLUS software. The signed path coefficients will be used to characterize the strength and direction of connections between voxels in the whole brain and those in the insular cortex. In addition, other cortical or subcortical regions, including the sensorimotor cortex, prefrontal lobe, thalamus, laryngeal motor cortex and sensory cortex, will be investigated.

Mapping relationship analyses between the insula-centered cerebral cortex-brainstem functional connectivity and multisource timing signals will be performed as follows. First, the point coordinates of brain regions with strong functional connections will be obtained according

to the previous steps and used as centers to draw spherical ROIs. RESTPLUS software was used to extract the GCA values of all samples in the $y-x$ direction for correlation analyses. Then, continuous variables, including multisource timing signals and GCA values, will be evaluated for normality with Kolmogorov-Smirnov tests. Pearson chi-square tests or Spearman correlation analyses (as appropriate) will be applied to clarify the relationship between the multisource timing signals and brain functional connectivity, which provides prior knowledge for model optimization.

Model optimization. The mapping relationship between the multisource signals and insula-centered cerebral cortex-brainstem functional connectivity explores the weight of different multisource signals and will be used as prior knowledge for model optimization.

The association between various physiological signals and swallowing disorders can be reflected by the actual clinical manifestation. This encompasses the significance of different physiological signals and the characteristic parameters that are pivotal in monitoring malabsorption. All these elements can serve as prior knowledge to guide the optimization of the model. Incorporating prior knowledge in the signal feature fusion stage further improves the feature fusion performance and optimizes the proposed aspiration monitoring model.

The first step will be to assign different initial weights to different signals according to the a priori distribution before feature fusion. Specifically, based on the importance weights of the a priori five meta-signals $\{\beta^c\}_{s=1}^5$, a learnable weight α will be obtained based on the feature fusion based on the attention mechanism, and the online learned α and the a priori β work together to determine the fused feature vector:

$$\tilde{X}_t = \sum_{s=1}^5 \alpha_s^t \beta^s h_t^s \quad (3)$$

Second, an additional prior branch will be introduced. Prior knowledge will be obtained by encoding the prior features, and the prior features will be further fused with \tilde{X}_t . At this point, the attention weight γ will be introduced to identify the importance of the prior features in the model. It will be assumed that the manually set prior is x_{prior} , and the feature $X_{\text{prior}} = \text{encoder}(x_{\text{prior}})$ will be obtained after encoding. The fused features are as follows:

$$\hat{X}_t = \gamma \tilde{X}_t + (1 - \gamma) X_{\text{prior}} \quad (4)$$

Selecting appropriate prior knowledge helps to further optimize the fused features and enhance the model effect. After obtaining the base model and the clinical trial analysis, prior knowledge can be further updated, and the model can continue to be optimized.

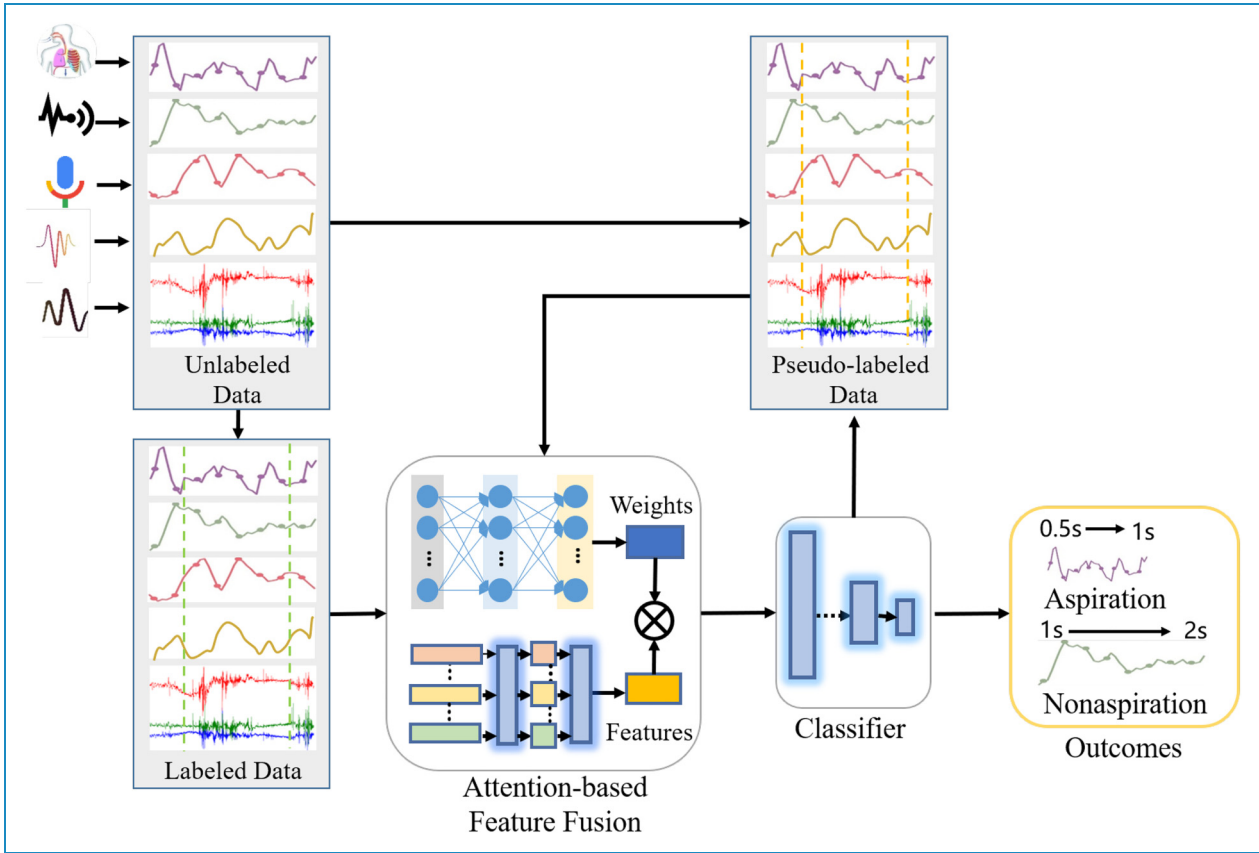


Figure 2. Framework of the SA monitoring model based on semisupervised learning.

Clinical application. The real-time dynamic monitoring system for SA will have been established when the accuracy, precision and recall rate meet the standard criteria. Then, the improved sensitivity and specificity of the system will be demonstrated through clinical applications. Precise patient interventions can be adopted according to the real-time dynamic monitoring results, and a VFSS can be applied simultaneously to test the accuracy of the system.

Statistical analyses

Multisource signal analyses. *Apnea time:* The nasal airflow signals will be labeled by the VFSS, with points i and j marked in one respiratory cycle. k is an arbitrary integer ($k \in (i, j)$) that satisfies the condition $x(k) = 0$. f_s refers to the sampling rate of the nasal airflow signal. The apnea time is calculated as follows:

$$\text{Apnea}(x) = \frac{j - i}{f_s} \quad (5)$$

Aspiration sound: Adobe Audition (Adobe Systems) will be used for sound signal analyses. The morphological characteristic parameters will be set as follows: (1) The variance represents the degree of dispersion in the signals. (2) The skewness represents the third-order matrix of the signals

and reflects the degree of asymmetry in the signal probability density function. (3) Kurtosis is a fourth-order matrix that reflects the degree of steepness in the signal probability density function. (4) The jitter values are a measure of the variation in the period of the sound signal. It can be quantified using the local, absolute and relative average perturbation (RAP) and the five point-period perturbation quotient (PPQ5). (5) The shimmer value measures the variation in the amplitude of the sound signal. It is calculated by localdbShimmer, amplitude perturbation quotient (APQ)3, APQ5 and APQ11.

The frequency domain sound signal features (including the bandwidth and spectral centroid) will be extracted by short-time Fourier transforms. The bandwidth of a signal refers to the range of frequencies that the signal occupies. It is calculated as the difference between the upper and lower frequencies of the signal. The spectral centroid is the mean of the frequencies weighted according to the energy and is an important parameter that reflects the frequency and energy distribution of sound signals.

EMG signals: After filtering, the resulting EMG signals will be split into several smaller data segments as data windows. It will be assumed that the total number of windows of interest is N . The EMG signals will be analyzed in the time and frequency domains. For the time domain,

the root mean square (RMS) will be calculated by taking the mean of the squared EMG samples over a window of interest and then taking the square root of the resulting value. The RMS can be written as follows:

$$\text{RMS} = \sqrt{\frac{\sum_{i=1}^N \text{EMG}_i^2}{N}} \quad (6)$$

For the frequency domain, the EMG spectra and power spectra will be obtained by short-time Fourier transforms. $P(f_k)$ refers to the power at the k -th frequency bin of the EMG spectra, and P denotes the power spectral density function. The mean power frequency (MPF) is defined as the frequency at which half of the power is above this frequency and half is below. MPF can be calculated using the following equation:

$$\text{MPF} = \frac{\sum_k (f_k \times P(f_k))}{\sum_k P(f_k)} \quad (7)$$

The median frequency (MD) is calculated as follows:

$$\text{MD} = f_{m-1} + \Delta F \times \frac{0.5 \sum_k P(f_k) - \sum_{k=1}^{m-1} P(f_k)}{\sum_{k=1}^m P(f_k)} \quad (8)$$

where m is the index of the current frequency bin at or above the median frequency, f_{m-1} is the frequency bin below the median frequency, and ΔF is the frequency resolution.

Pressure signals: Pressure signals are used to evaluate laryngeal motion (LM). The original LM signal will be transformed into the integrated LM (iLM) involving the calculation of the area under the receiver-operating characteristic (ROC) curve (AUC) of the LM waveform. The AUC value can be obtained by integrating the absolute value of the LM signal over time. The equation for calculating the integrated LM can be written as follows:

$$\text{iLM} = \int \text{LM}(t) dt \quad (9)$$

To match the pressure signals with the thyroid cartilage motion, the laryngeal rise time (LRT) and laryngeal activation duration (LAD) will be calculated based on the iLM. The calculation steps for these two parameters are as follows: (1) Within each determined swallowing cycle, the time point (P) where the sensor's raw output (LM) reaches its peak is searched for first. Due to the different characteristics of the pressure impedance sensor, this corresponds to the situation where the speed of the laryngeal elevation reaches its maximum. (2) Then, the data are searched backward from P for the zero-crossing point. If iLM has a positive value at this zero-crossing point, the search is continued forward until the local minimum of iLM with a negative value is found, which indicates the starting point (T1) of LRT. (3) A search is performed forward from P to the first zero-crossing point M . M is the time point when the

larynx reaches its maximum height. (4) A search is performed backward from M for the zero-crossing point. If iLM has a positive value at this zero-crossing point, the search is continued forward until the local minimum of iLM with a negative value is found, which indicates the ending point (T2) of LAD. In conclusion, LRT represents the duration between time T1 and M , while LAD represents the duration between time P and T2 (Fig. 3).

Accelerometry signals: The feature set in the time domain comprises statistical metrics, such as the mean and standard deviation. In addition, other features, such as the difference between the minimum and maximum values, zero-crossings, peak to average ratio (PAR), signal magnitude area (SMA), signal vector magnitude and differential signal vector magnitude (DSVM), will be computed.⁴⁰

The feature set in the frequency domain will be calculated by the fast Fourier transform (FFT) algorithm. The FFT algorithm will be used to convert the time series data of each component into the frequency domain. The resulting features in the frequency domain include energy and entropy metrics.⁴¹ Spectral energy in certain frequencies refers to the amount of energy or power that is present in a specific frequency range within the signal's frequency spectrum. The Fourier transform decomposes the signal into its component frequencies. Entropy is a statistical measure of the amount of uncertainty or randomness in the signal. It can be used to analyze the complexity or irregularity of the signal's frequency distribution.

Ethics and dissemination

The trial protocol was approved by the Medical Ethics Committee of the Third Hospital of Sun Yat-Sen University (2022-02-192-01). The study will be conducted according to good clinical practice guidelines suggested by researchers in the neurorehabilitation, artificial intelligence, and neuroimaging fields. The investigators will obtain written informed consent from all individuals by the investigators. The study's findings will be published in high-ranking peer-reviewed international journals and communicated to the (inter)national scientific community.

Serious adverse events

Safety risks for participants are minimal, as real-time dynamic monitoring is noninvasive. Moreover, VFSSs are conventional examinations in clinical applications and are conducted by well-trained physiotherapists to reduce complications. Furthermore, fMRI does not include the unknown cumulative risk of radiation exposure.

Expected results

Multisource signals will be stably extracted by multisource sensors. The events of the swallowing cycle, including the

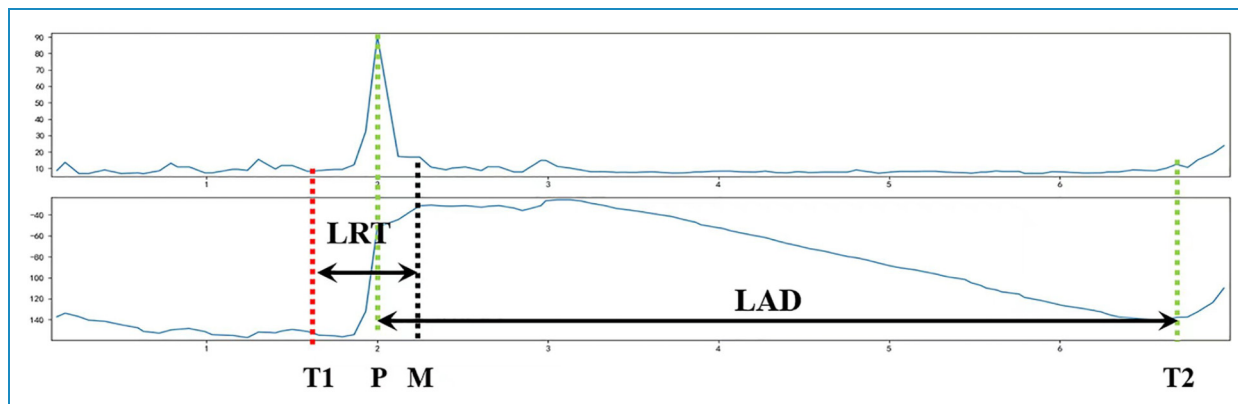


Figure 3. Calculation of the LRT and LAD. LRT, laryngeal rise time; LAD, laryngeal activation duration.

oral phase and pharyngeal phase, will be consistent with the time series VFSS results. A total of 3200 swallows will be obtained from patients with SA, with 1200 swallows with nonaspiration labeled by the VFSS and 2000 unlabeled swallows. The dataset will include swallowing data.

A significant difference in the multisource signals is expected to be found between the SA and nonaspiration groups. The sound signals will show differences in the variance, skewness, kurtosis, bandwidth and spectral centroid. The nasal airflow data will indicate that the apnea time in the SA group is shorter than that in the nonaspiration group, and the breathing–swallowing patterns will differ between the SA and nonaspiration groups (mainly manifesting as inhaling directly after swallowing in the SA group). Moreover, the EMG signals are expected to show differences in the time domain (including in the iEMG and RMS) and the frequency domain (including in the MPF and MF). Furthermore, the pressure signal analyses will show that the LM displacement is shorter and the LRT is longer in the SA group than in the nonaspiration group.

The features of labeled and pseudolabeled multisource signals will be extracted based on semisupervised deep learning to establish a dynamic SA monitoring model. Moreover, strong correlations will be found between the GCA value (from the left middle frontal gyrus to the right anterior insula) and the LRT. Therefore, the dynamic SA recognition model will be optimized by increasing the LRT weight. The results will demonstrate that the accuracy, precision and recall of the model meet the desired requirements. Finally, a dynamic SA monitoring system will be established for clinical applications to assess the occurrence of SA in routine daily practice.

Discussion

The present study establishes a real-time dynamic monitoring system for SA in stroke patients by extracting multisource signals during aspiration and verifies the central regulatory mechanism of SA. This system can be an early

warning system of SA both in inpatient and outpatient rehabilitation. It was reported that AP is closely related to SA^{4–6}; therefore, the system may decrease the occurrence of AP, which may increase the quality of life of SA patients.

Current swallowing aspiration detection models mostly use supervised deep learning methods,²¹ but the effectiveness of the models is limited due to the limited clinical annotated samples. In this study, a semisupervised deep learning method will be used, which integrates labeled and unlabeled samples for feature learning and addresses the problem of an insufficient number of labeled samples by generating pseudolabels. Furthermore, with more signal sources than other studies,^{8–12} the model can learn high-level representations simultaneously from raw signals and have substantial expressive power.

In addition, precise interventions will be applied in SA patients according to our results. The optimal treatment strategy for aspiration includes swallowing exercises (such as the Mendelsohn maneuver, supraglottal swallowing, shaker exercises and effort pitch glide), neuromuscular electrical stimulation and noninvasive brain stimulation (NIBS).^{42–44} The present study will explore the mapping relationship between multisource signals and brain function and find the most relevant brain region of SA, which provides a feasible method for choosing the central target of NIBS. Furthermore, trying to intervene at the peripheral level by evaluating the characteristics of multisource signals of SA provides a precise way to treat SA.



There are some limitations in this protocol. First, the VFSS carries the unknown cumulative risk of radiation exposure. Second, injector use to induce swallowing is not representative of natural swallowing conditions, which may lead to potential bias. Third, we used rsfMRI to explore the mapping relationship between the functional connectivity of neural networks and multisource signals, which is indirect. Fourth, simultaneous multiple-data acquisition from PSD patients using excessive multiple sensors may cause discomfort.

Conclusion

The study will establish a real-time dynamic monitoring system for SA in PSD patients based on semisupervised deep learning by extracting multisource signals during SA, with high sensitivity, specificity, accuracy and F1 score.

Declaration of conflicting interests: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Natural Science Foundation of China, Science and Technology Program of Guangzhou, China (grant numbers 82272617 and 2023B03J1234).

ORCID iDs: Jia Qiao  <https://orcid.org/0000-0002-8470-9580>
Yan-bin Gong  <https://orcid.org/0000-0001-6044-4146>

References

- Ghebre YT and Raghu G. Idiopathic pulmonary fibrosis: novel concepts of proton pump inhibitors as antifibrotic drugs. *Am J Respir Crit Care Med* 2016; 193: 1345–1352.
- Zhang M, Li C, Zhang F, et al. Prevalence of dysphagia in China: an epidemiological survey of 5943 participants. *Dysphagia* 2021; 36: 339–350.
- Patel DA, Krishnaswami S, Steger E, et al. Economic and survival burden of dysphagia among inpatients in the United States. *Dis Esophagus* 2018; 31: –7.
- Chinese Expert Consensus Group of Dysphagia and Nutrition Management. Chinese expert consensus on food and nutrition management for dysphagia (2019 version). *Asia Pac J Clin Nutr* 2020; 29: 434–444.
- Dziewas R, Michou E, Trapl-Grundschober M, et al. European stroke organisation and European society for swallowing disorders guideline for the diagnosis and treatment of post-stroke dysphagia. *Eur Stroke J* 2021; 6: Lxxxix–Lxxcxv.
- Mandell LA and Niederman MS. Aspiration pneumonia. *N Engl J Med* 2019; 380: 651–663.
- Giraldo-Cadavid LF, Leal-Leaño LR, Leon-Basantes GA, et al. Accuracy of endoscopic and videofluoroscopic evaluations of swallowing for oropharyngeal dysphagia. *Laryngoscope* 2017; 127: 2002–2010.
- Lee K, Ni X, Lee JY, et al. Mechano-acoustic sensing of physiological processes and body motions via a soft wireless device placed at the suprasternal notch. *Nat Biomed Eng* 2020; 4: 148–158.
- Steele CM, Mukherjee R, Kortelainen JM, et al. Development of a non-invasive device for swallow screening in patients at risk of oropharyngeal dysphagia: results from a prospective exploratory study. *Dysphagia* 2019; 34: 698–707.
- Coyle JL and Sejdić E. High-resolution cervical auscultation and data science: new tools to address an old problem. *Am J Speech Lang Pathol* 2020; 29: 992–1000.
- Shieh WY, Wang CM, Cheng HK, et al. Using wearable and non-invasive sensors to measure swallowing function: detection, verification, and clinical application. *Sensors (Basel)* 2019; 19: 2624.
- Yagi N, Nagami S, Lin MK, et al. A noninvasive swallowing measurement system using a combination of respiratory flow, swallowing sound, and laryngeal motion. *Med Biol Eng Comput* 2017; 55: 1001–1017.
- Donohue C, Khalifa Y, Mao S, et al. Characterizing swallows from people with neurodegenerative diseases using high-resolution cervical auscultation signals and temporal and spatial swallow kinematic measurements. *J Speech Lang Hear Res* 2021; 64: 3416–3431.
- McNulty J, de Jager K, Lancashire HT, et al. Prediction of larynx function using multichannel surface emg classification. *IEEE Trans Med Robot Bionics* 2021; 3: 1032–1039.
- Park HY, Park D, Kang HS, et al. Post-stroke respiratory complications using machine learning with voice features from mobile devices. *Sci Rep* 2022; 12: 16682.
- Roldan-Vasco S, Orozco-Duque A and Orozco-Arroyave JR. Swallowing disorders analysis using surface emg biomarkers and classification models. *Digit Signal Process* 2023; 133: 103815.
- Roldan-Vasco S, Restrepo-Urbe JP, Orozco-Duque A, et al. Analysis of electrophysiological and mechanical dimensions of swallowing by non-invasive biosignals. *Biomed Signal Process Control* 2023; 82: 104533.
- Avanzo M, Wei L, Stancanello J, et al. Machine and deep learning methods for radiomics. *Med Phys* 2020; 47: e185–e202.
- Dudik JM, Coyle JL, El-Jaroudi A, et al. Deep learning for classification of normal swallows in adults. *Neurocomputing* 2018; 285: 1–9.
- Goulas A, Damicelli F and Hilgetag CC. Bio-instantiated recurrent neural networks: integrating neurobiology-based network topology in artificial networks. *Neural Netw* 2021; 142: 608–618.
- Khalifa Y, Mandić D and Sejdić E. A review of hidden Markov models and recurrent neural networks for event detection and localization in biomedical signals. *Inf Fusion* 2021; 69: 52–72.
- Saad M, Afsah O, Baz H, et al. Clinical and videofluoroscopic evaluation of feeding and swallowing in infants with oropharyngeal dysphagia. *Int J Pediatr Otorhinolaryngol* 2021; 150: 110900.
- Horton KK, Segers LS, Nuding SC, et al. Central respiration and mechanical ventilation in the gating of swallow with breathing. *Front Physiol* 2018; 9: 785.
- Smith JC, Abdala APL, Borgmann A, et al. Brainstem respiratory networks: building blocks and microcircuits. *Trends Neurosci* 2013; 36: 152–162.
- Steuer I and Guertin PA. Central pattern generators in the brainstem and spinal cord: an overview of basic principles, similarities and differences. *Rev Neurosci* 2019; 30: 107–164.
- Babaei A, Ward BD, Siwec RM, et al. Functional connectivity of the cortical swallowing network in humans. *Neuroimage* 2013; 76: 33–44.
- Dionisio S, Mayoglou L, Cho SM, et al. Connectivity of the human insula: a cortico-cortical evoked potential (ccep) study. *Cortex* 2019; 120: 419–442.
- Mihai PG, Otto M, Platz T, et al. Sequential evolution of cortical activity and effective connectivity of swallowing using fmri. *Hum Brain Mapp* 2014; 35: 5962–5973.

29. Wilmskoetter J, Daniels SK and Miller AJ. Cortical and sub-cortical control of swallowing-can we use information from lesion locations to improve diagnosis and treatment for patients with stroke? *Am J Speech Lang Pathol* 2020; 29: 1030–1043.
 30. Trevizan-Baú P, Dhingra RR, Furuya WI, et al. Forebrain projection neurons target functionally diverse respiratory control areas in the midbrain, pons, and medulla oblongata. *J Comp Neurol* 2021; 529: 2243–2264.
 31. Van Daele DJ and Cassell MD. Multiple forebrain systems converge on motor neurons innervating the thyroarytenoid muscle. *Neuroscience* 2009; 162: 501–524.
 32. Wilmskoetter J, Marebwa B, Basilakos A, et al. Long-range fibre damage in small vessel brain disease affects aphasia severity. *Brain* 2019; 142: 3190–3201.
 33. Ciesielska N, Sokołowski R, Mazur E, et al. Is the Montreal cognitive assessment (moca) test better suited than the mini-mental state examination (mmse) in mild cognitive impairment (mci) detection among people aged over 60? Meta-analysis. *Psychiatr Pol* 2016; 50: 1039–1052.
 34. Monchi O, Petrides M, Petre V, et al. Wisconsin card sorting revisited: distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. *J Neurosci* 2001; 21: 7733–7741.
 35. Shieh WY, Wang CM, Cheng HK, et al. Noninvasive measurement of tongue pressure and its correlation with swallowing and respiration. *Sensors (Basel)* 2021; 21.
 36. Calmarza-Chueca F, Cristina-Sánchez-Gimeno A, Perez-Nogueras J, et al. Institutionalized elderly are able to detect small viscosity variations in thickened water with gum-based thickeners: should texture classifications be reviewed? *BMC Geriatr* 2021; 21: 647.
 37. Logemann JA. *Manual for the videofluorographic study of swallowing* (2nd ed). Austin, TX: Pro-Ed, 1993.
 38. Qiao J, Wu ZM, Ye QP, et al. Relationship between post-stroke cognitive impairment and severe dysphagia: a retrospective cohort study. *Brain Sci* 2022; 12.
 39. Kang YJ, Arafa HM, Yoo J-Y, et al. Soft skin-interfaced mechano-acoustic sensors for real-time monitoring and patient feedback on respiratory and swallowing biomechanics. *NPJ digital Medicine* 2022; 5: 147.
 40. Le H-L, Nguyen D-N and Nguyen H-N. The novel method of pedestrian fall detection based on pso and rf using accelerometer data. *2021 International Conference on System Science and Engineering (ICSSE)* 2021: 111–115.
 41. Lu D-N, Nguyen D-N, Nguyen T-H, et al. Vehicle mode and driving activity detection based on analyzing sensor data of smartphones. *Sensors* 2018; 18: 1036.
 42. Balou M, Herzberg EG, Kamelhar D, et al. An intensive swallowing exercise protocol for improving swallowing physiology in older adults with radiographically confirmed dysphagia. *Clin Interv Aging* 2019; 14: 283–288.
 43. Qiao J, Ye QP, Wu ZM, et al. The effect and optimal parameters of repetitive transcranial magnetic stimulation on post-stroke dysphagia: a meta-analysis of randomized controlled trials. *Front Neurosci* 2022; 16: 845737.
 44. Zhong L, Rao J, Wang J, et al. Repetitive transcranial magnetic stimulation at different sites for dysphagia after stroke: a randomized, observer-blind clinical trial. *Front Neurol* 2021; 12: 625683.
-