UNIVERSIDAD DE ANTIOQUIA

Assessment of oropharyngeal dysphagia using multimodal biosignals

Sebastián Roldán-Vasco

A thesis submitted in conformity with the requirements for the degree of
Doctor in Electronic and Computer Engineering

Director
Juan Rafael Orozco-Arroyave, Ph.D.

Co-Director
Andrés Felipe Orozco-Duque, Ph.D.

Universidad de Antioquia
Faculty of Engineering
Doctorate in Electronic and Computer Engineering
Medellín
2023

Doctorate in Electronic and Computer Engineering, XVIII Cohort.
GITA Research Group

Centro de Documentación de Ingeniería

Assessment of oropharyngeal dysphagia using multi-modal biosignals

A thesis submitted to

Faculty of Engineering
Universidad de Antioquia

to obtain the degree of

DOCTOR OF ELECTRONIC AND COMPUTER ENGINEERING

presented by

Sebastián Roldán-Vasco

Medellín - Colombia
2023

Dissertation in conformity with the requirements of the
Faculty of Engineering
Universidad de Antioquia, Medellín, Colombia

Day of dissertation: July 19, 2023

Directors:

Prof. Dr.-Ing. Juan Rafael Orozco-Arroyave
Ing. Andrés Felipe Orozco-Duque, Ph.D.

Acknowledgment

This thesis was developed over five years, with a pandemic in between. I am very grateful for my life because despite the millions of people dead worldwide, including more than one hundred thousand in my country, I did not lose any relatives or close friends. But we are here, and it makes me happy. Not too many people can survive the worst pandemic in a century with enough strength to continue with their Ph.D. studies.

I am very grateful to my advisors, Prof. Dr.-Ing. Rafael Orozco, and Ph.D. Ing. Andrés Orozco, who always trusted in me despite my - sometimes - out-of-focus ideas, and helped me to seek the research's spotlight. Thanks to them, they always behaved like patient fathers instead of severe doctors. They were inspiring for me. Thanks to Rafa's security and Andrés' serenity, the construction of this work was much more peaceful than the Ph.D. students usually say. I was very lucky for having them as advisors and friends. Their academic and technical contributions were invaluable, and I wish to keep up with their knowledge to continue this rough but pleasant scientific way.

I would also like to thank the interdisciplinary team in the dysphagia research team that welcomed me despite we spoke different technical languages a lot of times. In particular, Juan Camilo Suárez with the endless academic discussions and his valuable contributions to the swallowing-related phenomena and dysphagia. A lot of medical concepts used in this thesis were clarified by him. I also thank the speech-language pathology team, including Elizabeth Gómez for her love and interest for the patients as well as her dedication to the projects, Lillyana Martinez for her selfless disposition to help in the patients' recruitment in OFA IPS despite the discomforts, and Claudia Bedoya with her rigor in the evaluation of patients. The acquisition protocols and patient recruitment would not be possible without their dedication. In this regard, I acknowledge all the volunteers, healthy but especially the patients, who were fundamental in this process. I recognize their availability, kindness, and willingness to go to the different acquisition facilities, many times with motor disabilities and annoying weather, with the only motivation to contribute to the research. I hope that the results obtained in this work, contribute in some way, one day, to improve their quality of life.

I am also very grateful to the master's and bachelor's students that helped me a lot with the results achieved in this thesis. In particular, I wish to thank Sebastián Restrepo for his passion for coding everything susceptible to being coded, Juan Pablo Restrepo for his valuable technical help and his willingness to support me in everything, Estefanía Pérez with her responsibility for the volunteers' recruitment and for picking me up anywhere, Yorhagy Valencia for his patience with my - sometimes - bad suggestions and to push me to reconsider some strategies in speech analysis, Jessy Moreno with her dedication and simple but focused suggestions to improve the quality of results, and Andrés Flórez with his quiet but important contributions in the speech analysis. Although I was their advisor, I feel that they teach me a lot of things that contributed to making this way easier.

doors to do my studies, and the INSTITUTO TECNOLÓGICO METROPOLITANO, who gave me peace of mind to work in all of the dysphagia-related research projects

And last but not least, I would like thank to my family. The patient love of my mom, the intellectual curiosity of my father, and the impulse of my brother for me to be better every day. And yes, to Fabi, who motivated me to do this Ph.D., and her love despite those several hours that she had to sacrifice so I could do this thesis.

Thanks to them all.

Medellín, January 2023
Sebastián Roldán-Vasco

To the memory of Yorhagy, wherever you are.
And to those patients who died during the research.

Abstract

Dysphagia is a swallowing impairment that affects the food, liquid, or saliva transit from the mouth to the stomach. Dysphagia leads to malnutrition, dehydration, and aspiration of the bolus into the respiratory system, which can lead to pneumonia with subsequent death. Dysphagia is produced by a set of neurogenic and neuromuscular conditions with variable incidence and prevalence. This condition is under-recognized and under-diagnosed. However, physical, economic, social, and psychological burdens have been clearly identified.

The clinically accepted methods for dysphagia diagnosis and follow-up are invasive, uncomfortable, expensive, and experience-dependent. Furthermore, the reliability of some methods is still discussed. In this way, biosignals-based approaches that try to solve the aforementioned problems have been proposed, but no conclusive and hardly reproducible results have been achieved. Otherwise, such strategies generally ignore some physical aspects of the swallowing process.

Therefore, this work explored non-invasive strategies to objectively assess dysphagia. To evaluate different physical aspects of the swallowing process, a multi-modal asynchronous analysis was performed with three biosignals: surface electromyography, accelerometry-based cervical auscultation, and speech. Such biosignals contributed to analyzing the swallowing-related phenomena in electrophysiological, mechanical, and acoustic dimensions. This thesis was focused on understanding oral and pharyngeal phases of the swallowing process by the use of the aforementioned signals. The following methodological steps were proposed to develop the dysphagia assessment scheme: 1) design of an acquisition protocol for the three biosignals in patients with dysphagia and healthy controls; 2) characterization of such biosignals in different mathematical domains, leading to the proposal of interpretable biomarkers; 3) construction of representation spaces and modeling of the swallowing patterns; and 4) evaluation of the multi-modal approach as a reliable method for swallowing assessment.

All signals demonstrated their suitability for dysphagia screening by themselves, but bi- and tri-modal scenarios with Support Vector Machines, Extreme Gradient Boosting, k-Nearest Neighbors, and Gated Multimodal Units outperformed the uni-modal classification results. Specific configurations retrieved outstanding results, i.e. all performance measures obtained values $\geq 0{,}95$. This thesis contributes to reducing the knowledge gap about swallowing-related phenomena and alterations from non-invasive and multi-modal points of view, with high potential to transfer and implement in clinical practice. It also contributes to objectively assessing dysphagia in the consulting room, helping with the diagnosis, follow-up, and rehabilitation of patients with dysphagia.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1. Motivation

Dysphagia refers to alterations of at least one of the three swallowing phases: oral, pharyngeal and esophageal [Morg 10]. When an alteration occurs, the control of the process to move a bolus from the mouth to the stomach is reduced, and the risk of aspiration appears (presence of bolus in the airway) [Lope 14]. Dysphagia affects between 8.4% and 16% of people worldwide [Gira 16]. Only in USA, over 16 million people suffer from oropharyngeal dysphagia, whereas in Europe this number overcomes 40 million people [Taki 16]. Between 400.000 and 800.000 new cases of neurological dysphagia appear annually worldwide [Pfei 16]. Underestimated statistics show that around 3% of adult patients older than 45 years are diagnosed with dysphagia in the USA [Pate 17]. However, incidence and prevalence at the local, regional, and national levels are still unknown.

Dysphagia is a direct effect of several diseases, disorders and injuries [Lope 14], such as cerebral palsy, traumatic brain injury, stroke, multiple sclerosis, Amyotrophic Lateral Sclerosis (ALS), myasthenia gravis, Parkinson's disease, Alzheimer's disease, and others. Dysphagia reduces the capability of receiving oral feeding and causes malnutrition (weight loss, muscle breakdown, decreased general health) and dehydration [Clav 06, Cich 12]. Silent aspirations or aspirations without adequate cough response enhance the risk of pneumonia and even death [Clav 15, Youm 11]. Besides physical implications in patients, dysphagia has substantial burdens in economic, social, and psychological dimensions [Cich 12, Pate 17]. Dysphagia increases the length of hospital stay to 40% [Chen 16a], and a systematic review reported a mean of 3.98 days longer compared to individuals with no dysphagia [Attr 18]. A rough estimate indicates an economic burden of around 547 million USD per year only in the United States [Chen 16a], and a daily cost per patient of around 2500 USD [Cich 12]. Although there is high variability between reports regarding healthcare costs associated with dysphagia, an approximate cost increase of 40.36% was estimated [Attr 18].

The diagnosis of dysphagia is a experience-dependent, costly and time-consuming process. There is little information regarding the pattern of dysphagia produced by different diseases and medical conditions [Clav 06]. The standard clinical evaluation is the *bedside swallow examination*, which is a clinical checklist made by trained speech-language pathologists (SLP) [Carn 08]. However, most implementations are subjective, have poor validity and rigor, and could ignore the presence of aspirations [Farn 14]. The instrumental

techniques appear to overcome some of these lacks. The most widely used instrumental assessment methods are the videofluoroscopic and fiberendoscopic evaluations of swallowing, intended to detect aspiration and penetration (bolus in the larynx) [Lang 03, Hey 15]. Even though these methods are the gold standard in medical practice, they have several major problems: reduced availability in the local health system; invasiveness; lack of standardized protocols and validated scoring systems [Lang 03]; and the diagnosis depends on the professional expertise [McCu 01a, Farn 14].

With the aim of quantifying and reducing the invasiveness of current procedures to diagnose dysphagia, several techniques based on biosignals have emerged, such as cervical auscultation based on accelerometry [Dudi 15a] and swallowing sounds [Berg 14], surface electromyography [Step 12], mechanomyography [Lee 09b], voice recordings [Ipin 18], nasal flow measurements [Lee 11], and pulse oximetry [Sher 99]. Each biosignal describes different physical properties of swallowing in mechanical, acoustic, bioelectric, or hemodynamic dimensions. Thus, bearing in mind the complexity of the swallowing process, pretending to model and addressing the phenomenon with only one source of information could lead to false or biased conclusions. There is no consensus about which technique best represents clinical variables of swallowing, and those methods are not enough convincing to be regularly used in clinical practice. Furthermore, most of the existing approaches have been addressed descriptively; works that apply robust techniques from the quantification point of view, are mainly focused on the pharyngeal phase, disregarding the sequential patterns of the process. Additionally, most of the studies have considered small databases, which limits their analyses to very specific groups of patients with only one particular disease, and not studying other etiologies.

Dysphagia diagnosis based on reliable non-invasive methods that quantitatively evaluate patterns in the swallowing process considering multiple sources of information is an open research problem.

1.2.   Theoretical Background and State of the art

1.2.1.   Anatomy and physiology of swallowing

Swallowing is a complex neuromuscular process that implies the transit of specific boluses from the mouth to the stomach. The World Health Organization (WHO) defines swallowing as the set of "functions of clearing substances, such as food, drink or saliva through the oral cavity, pharynx, and esophagus into the stomach at an appropriate rate and speed" [ICF]. It requires coordinated voluntary and involuntary movements of more than 30 pairs of muscles in the mouth, tongue, pharynx, larynx, and esophagus bilaterally [Palm 00, Erte 03, Hamm 14, Shaw 13] (see Figure 1.1). Moreover, the swallowing process shares anatomical structures with breathing and phono-articulation functions, leading to co-morbidity when one of them suffers alterations [Farn 17], but the neural processes by which these functions are coordinated are not well understood [Bols 13].

A central pattern generator[1] for human swallowing has been hypothesized in the *medulla oblongata* [Erte 03, Shaw 13]. Normal swallowing responds to stimuli of the oropharynx, larynx, and esophagus via sensory neurons of the cranial nerves V, VII, IX, X, and

---

[1]Self-contained neural circuits

XII [Clav 15]. These cranial nerves have specific functions during the swallow [Shaw 13], summarized as follows [Berg 14]:

- Trigeminal (V): jaw opening.

- Facial (VII): lips retraction and protrusion.

- Glossopharyngeal (IX) and vagus (X): voice production (phonation, articulation, and prosody [Erma 09]) and palatal elevation.

- Hypoglossal (XII): tongue movements such as protrusion, lateralization, circular, and strength against resistance.

Even though the accessory nerve (XI) is not usually considered to participate in the neural control of swallowing, it could have motor innervation of the striated portions of the esophagus, larynx, and pharynx [Cost 18]. However, the mechanism that would involve such cranial nerve in the swallowing process remains unclear.
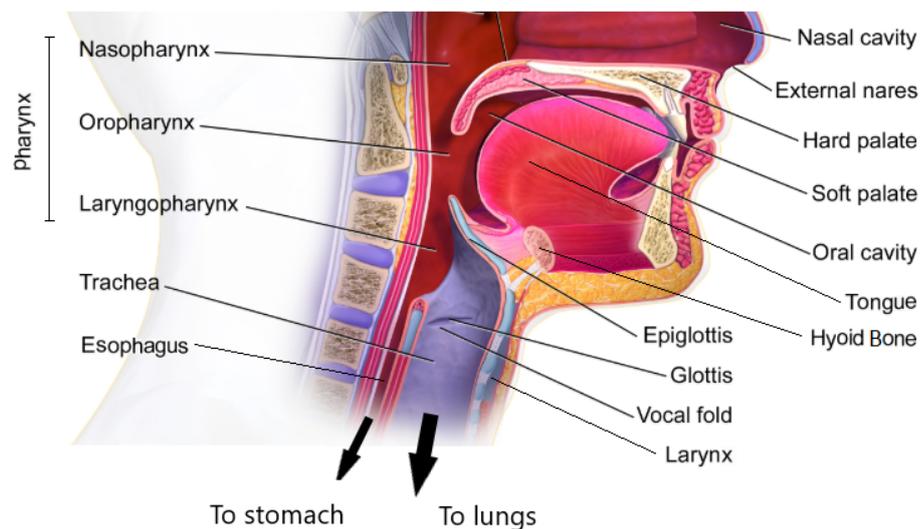
Figure 1.1: Sagittal view of anatomical structures related to swallowing. Adapted from [Blau 13].

Swallowing has characteristics of bilaterality and there is interhemispheric asymmetry, i.e., lateralization of the right hemisphere tends to be greater than of the left one, irrespective of handedness [Erte 03].

The entire process is highly dependent on the bolus size, bolus consistency, and age but independent of the gender [Bolz 13, Youm 11, Aydo 15, Card 10]. Although there are no well-defined patterns of swallowing, this process involves three phases, well-established since 1813 with Magendie's publications [Vaim 07]: oral, pharyngeal, and esophageal [Gupt 14] -see Figure 1.2-. Swallowing phases are described below.

Oral phase: It is voluntary and depends on the viscosity of the bolus [Clav 15]. The oral phase is controlled by cranial nerves V, VII, and XII [Walt 18]. In this phase, the soft palate is sealed by the tongue, avoiding the pass of the bolus to the pharynx -see Figure 1.2-. This phase is characterized by chewing, containment, control, swallow initiation, propulsion,
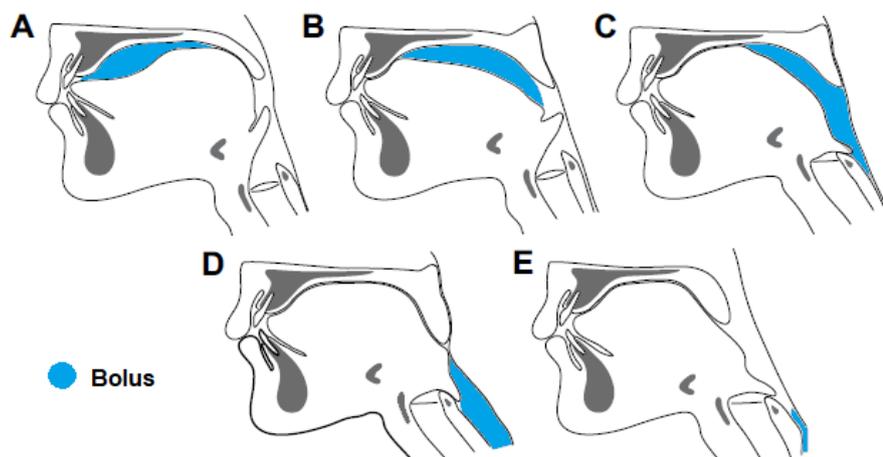
Figure 1.2: Diagram of the normal swallowing process. A) End of the oral preparatory phase. B) Bolus propelled from the mouth to the pharynx (end of oral phase). C) Beginning of the pharyngeal phase: elevation of the soft palate, nasopharynx closure, and upward-forward movement of the larynx. D) End of pharyngeal phase: upper esophageal sphincter opens, tongue contacts the pharyngeal wall. E) Esophageal phase: Soft palate descends, pharynx and larynx open again. From [Mats 08b] with written permission.

and transport of the bolus [Cich 12]. Impaired control of the tongue and dental problems produce disorders that affect this phase [Palm 00]. Some authors divide the oral stage into oral initial (labial sealing) and oral final (tongue squeezing against the hard palate) [Vaim 04a], whilst others divide it into oral preparatory and oral transit (or propulsive) phase [Shaw 13, Walt 18]. This division based on non-overlapped phases is known as the *four sequence model*. However, for solid food, a refinement of the oral phase in the four-sequence model has been proposed, in which there is a gradual bolus aggregation in the oropharynx simultaneously with the food processing in the oral cavity. This is known as the *process model* [Mats 09].

Pharyngeal phase: It is involuntary, although it may be initiated consciously [Dudi 15a], and it involves also the clearing of residues [Cich 12]. This phase is initiated with the propulsion of the bolus by the tongue and the opening of the upper-esophagic sphincter [Clav 15] -see Figure 1.2-. Although cranial nerves IX and X participate in the triggering mechanism mediated by the medulla [Walt 18], the whole process is not well-known [Erte 03]. The suprahyoid muscles -see Figure 1.5- pull the hyoid bone toward the mandible [Gurg 13, Poor 14], the larynx rises and the vocal cords close, locking the air-flow through the glottic plane [Mats 08b, Poor 14]. It has been documented that the suprahyoid muscles activate before the motion of the anterior and posterior tongue [Li 17]. Subsequently, the infrahyoid muscles -see Figure 1.5- descend the hyolaryngeal complex toward the sternum [Poor 14]. Such displacement moves the larynx under the base of the tongue and closes the laryngeal vestibule before opening the upper esophageal sphincter [Mats 08b]. Thus, there is a sealing of the unwanted pathways including the oral cavity by the tongue, the nasopharynx by the soft palate and the larynx by the epiglottis[2], observed as an apneic

---

[2]Fibrocartilage anterior to the laryngeal inlet [Walt 18]

period and it is followed by exhalation [Dudi 15a] -see Figure 1.2-. The airway protection depends on the integrity of the pharyngeal phase [Youm 11, Bols 13]. Since oral and pharyngeal phases are highly correlated, they are commonly known as "oropharyngeal". This integrated phase has an extraordinary complexity and velocity [Erte 03]: the activation of the muscles implied in this phase are ballistic-like (less than 600 ms) [Bols 13]. This phase is also the most critical from a health preservation point of view [Mart 08].

Esophageal phase: It is completely involuntary [Dudi 15a], and it is simpler and slower than the oropharyngeal one [Erte 03]. This phase lasts 5-6 s approximately [Walt 18]. It involves the relaxation of the upper esophagic sphincter, its opening by the movement of the hyoid bone and the larynx, the contact between the posterior pharyngeal wall and the posterior surface of the tongue (clearing the pharynx of residues) -see Figure 1.2-, and consequently peristaltic propulsion of the bolus from the lower esophagic sphincter to the stomach, mainly by the action of smooth muscles [Chen 16a, Palm 00, Erte 03]. During liquid intake, the gravity force helps to the bolus passage if the person lies in an upright position [Walt 18]. The lower esophagic sphincter closes after the passage of the bolus in order to prevent gastroesophageal reflux [Palm 00]. When weakness or incoordination of esophageal musculature appears, the esophageal propulsive force may be affected. By contrast, overactivity of the esophageal musculature may lead to loss of effectiveness in the transport of food through the esophagus due to spasms [Palm 00]. The esophageal phase is separated from the oropharyngeal one, because they are controlled by distinct neuroanatomic circuits [Brou 00].

1.2.2. Dysphagia

The term *dysphagia* refers to abnormalities in the swallowing process. Dysphagia can be classified as oropharyngeal or esophageal, depending on the affected swallowing phase. Oropharyngeal dysphagia is related to difficulties in swallowing initiation, coughing, choking, repeated swallows, voice changes, or nasal regurgitation [Chen 16a], as well as to malfunction of the airway protection mechanism, that leads to penetration -meal in the laryngeal vestibule and above true vocal folds- or aspiration -meal pass through the trachea- [Palm 00]. On the other hand, sensations of food "getting stuck" at the chest level, heartburn, chest pain, or odynophagia (swallowing-related pain) could indicate esophageal dysphagia [Chen 16a]. An increase in the number of swallows per standard boluses is reported also as a sign of dysphagia [Berg 14]. However, distinguishing between both types of dysphagia is difficult because they share signs and symptoms, such as slow swallowing speed, piecemeal deglutition, wet voice after swallowing, etc [Kang 17, Palm 00].

Although the prevalence of dysphagia in hospitalized patients is estimated in 12–13% (up to 30% in elderly people) [Chen 16a], in patients under intensive care or at-home nursing care the prevalence can increase up to 60% [Pfei 16, Ekbe 02].

Some muscular, peripheral and central nervous structures malfunctioning may cause neurogenic dysphagia, which is manifested by neurological[3] or neuromuscular[4] conditions [Pfei 16]. Both conditions are known as *functional* dysphagia (see Figure 1.3). Other

---

[3]Compromise in peripheral or central nervous system
[4]Alterations in neuromuscular junctions or muscles of pharynx and esophagus

causes such as iatrogenic, infectious, mechanical, and structural may trigger dysphagia too [Sase 17]; however, these origins of dysphagia are out of the scope of this Ph.D. thesis.
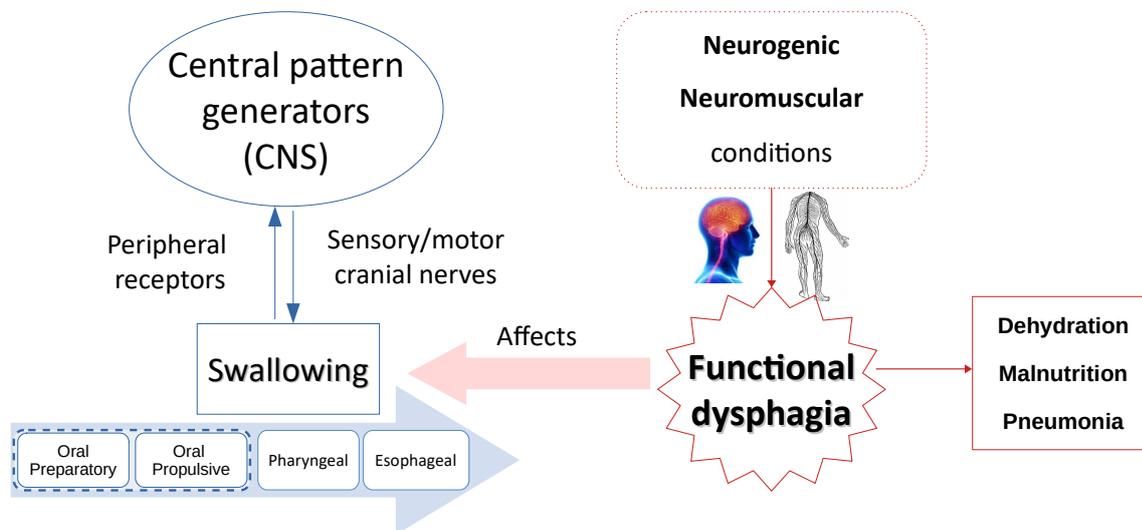
Figure 1.3: Scheme of the physical burden and causes of functional dysphagia.

Neurological conditions are the origin of 75% of oropharyngeal dysphagia manifestations [Erte 03]. It includes brainstem tumors, traumatic brain injury (27-50% of dysphagia's incidence), stroke (between 8.1% and 80% of incidence, 20% of them with aspiration pneumonia), cerebral palsy, Guillain-Barré syndrome, Huntington's disease, multiple sclerosis (44% of incidence), amyotrophic lateral sclerosis (60% of incidence), Parkinson's disease (prevalence between 18.5% and 100%), dementia (45% of institutionalized patients with clinical evaluation, and up to 95% with instrumental evaluation in advanced states), Alzheimer's disease (limited evidence about incidence and prevalence), among others [Cook 99, Ekbe 02, Clav 06, Chen 16a, Baij 09, Taki 16, Ozsu 20]. On the other hand, myopathic conditions include dermatomyositis, myasthenia gravis, myotonic dystrophy, oculopharyngeal dystrophy, polymiositis, among others [Erte 03, Cook 99, Ekbe 02, Clav 06, Chen 16a, Baij 09]. Furthermore, studies conducted in different countries with no consistent age of the population, number of individuals and evaluation methods, show differences in dysphagia prevalence: from 1.7% in China (between 18 and 70 years old), to up to 55% in Spain (>70 years old in patients with pneumonia) [Rode 13]. Disparate or missing epidemiological data associated to some mentioned conditions are an indicator of the gap in swallowing research [Taki 16].

Functional dysphagia can lead to malnutrition, dehydration, recurrent upper respiratory infections, and possible pneumonia (if aspiration occurs) with subsequent death [Pfei 04, Clav 15]. The mortality due to aspiration episodes increases when it is associated with dysphagia (27% in patients with safe swallowing vs. 55% in patients with dysphagia [Cich 12]). Although physical and to a lesser extent economic effects of dysphagia have been studied, psychological and social burdens have been neglected [Farr 07]. Dysphagia can produce important effects on quality of life from both social and emotional points of view [Farr 07], destroying the social opportunities -and pleasure- of mealtimes because the patients feel embarrassed, affecting the quality of relationships between them and caregivers, isolating the patients and generating anxiety, with subsequent affectations on the in-

dividual's dignity [Ekbe 02, Farr 07]. Dysphagia is related to the reduction of self-esteem, exercise, security, work capacity, and leisure time [Gust 91]. Figure 1.4 summarizes the burden of dysphagia.
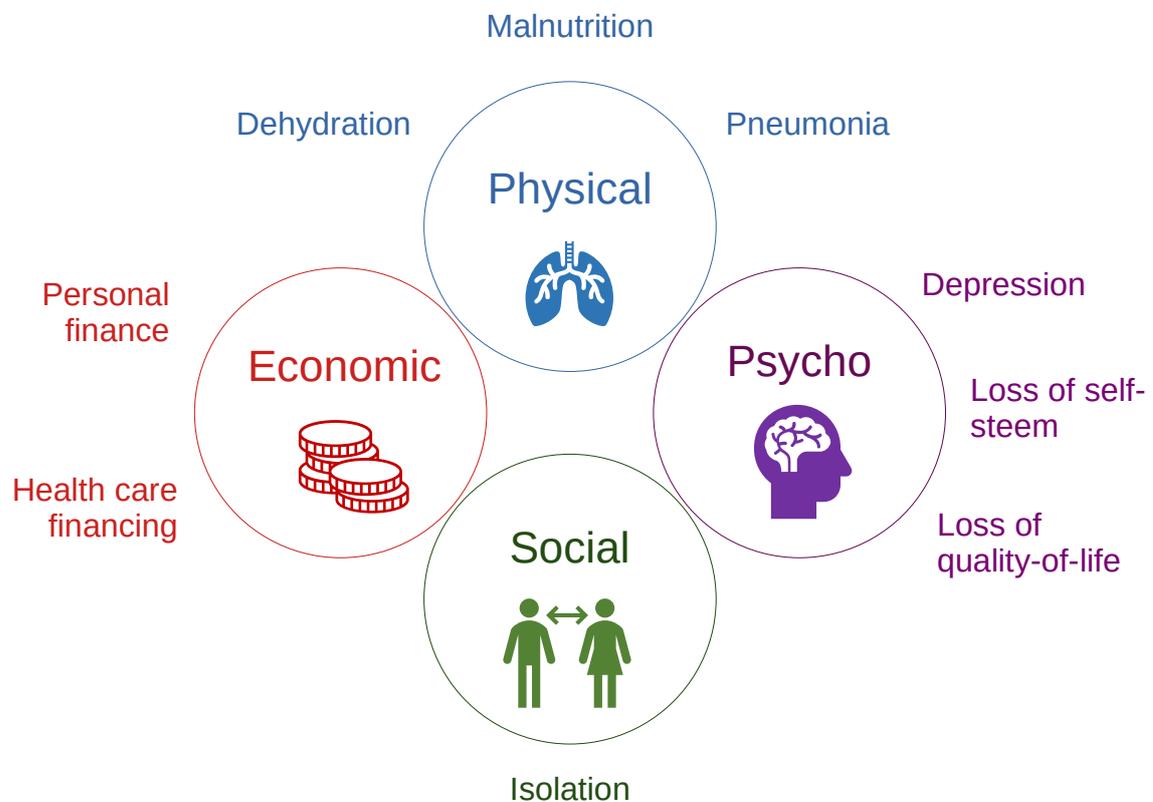
Figure 1.4: Dysphagia-related burden.

Diagnosis of dysphagia

Although dysphagia is under-recognized, under-diagnosed and under-managed [Ekbe 02, Rode 13], its clinical diagnosis can be divided into three components: screening, Clinical Swallow Examination -CSE- and instrumental evaluation [Gira 16, Berg 14].

The screening implies preliminary tests and checklists aiming to determine whether the patient requires a more detailed examination [Dudi 15a], i.e. it aims to identify the presence or risk to develop dysphagia [Berg 14]. Several screening methods are used such as initial assessment of cognitive status, water swallow test, pudding swallow test, gag reflex, voluntary cough, evaluation of voice changes, repetitive swallow saliva test, among others [Hass 14, Carn 08, Yagi 17]. However, only the 3-oz water swallow test has been validated to perform screening [DePi 92, Carn 08]. Most of the available screening methods have questionable reliability due to their subjectivity [Lesl 04] and lack of quantitative analysis [Yagi 17]. In addition, it is common to find hospitals with different protocols for the detection of dysphagia with the aforementioned lack of validation. It is estimated that the number of such protocols increased by around 20 % in the last decade [Fest 16]. Furthermore, most of the available screening methods are intended to evaluate the stroke population, disregarding other diseases and conditions that produce dysphagia [Mozz 17].

Table 1.1: Regularly used scales for CSE

Test	Classification	Characteristics	References
EAT-10	Healthy, oropharyngeal dysphagia without aspiration, and oropharyngeal dysphagia and aspiration	Ten points Easy to apply and interpret Validated for Spanish language (Colombia)	[Sifr 14, Gira 16]
PAS	From 1 (Material does not enter the airway) to 8 (Material enters the airway, passes below the vocal folds, and no effort is made to eject)	Eight points Classification according to the aspiration risk	[Rose 96]
MPAS	Five-points simplification of PAS	Simpler to implement that PAS	[Kang 17]
DOSS	From 7 (normal swallows) to 1 (severe dysphagia)	Evaluate the severity of dysphagia using videofluoroscopy Subjective by far Controversial inter-rater reliability	[Zark 18, ONei 99]
MASA	Mild, moderate and severe dysphagia	5- and 10- rating scale Validity comparable to other clinical scales Not evaluated in non-neurological patients for long-term	[Carn 08, Crar 13a, Anto 10]
SWAL-QOL	From 1 to 5 (or 6, depending on the item)	Ninety-three items No disease specific Time consuming	[McHo 00a, McHo 00b, Guti 15]

Another limitation of most swallowing screening studies is that they tend to correlate dysphagia only to the presence of aspiration. However, it is only one component of dysphagia and not always an obligatory symptom, and it may be intermittent [Carn 08].

The CSE is a clinical assessment procedure oriented to the diagnosis, treatment, and management of dysphagia. It could also include the follow-up and education of the patient [Berg 14, Carn 08]. Even though some authors consider the screening as part of the CSE [Hass 14, Carn 08], there are well-defined examples of CSE-related methods: the 10-item self-perception Eating Assessment Tool (EAT-10); the Penetration-Aspiration Scale (PAS) and its modified version MPAS; the Dysphagia Outcome and Severity Scale (DOSS); the Mann Assessment of Swallowing Ability (MASA); and the self-perception tool for evaluation of quality-of-life and quality-of-care outcomes (SWAL-QOL). Some characteristics of these methods are shown in Table 1.1. Although other scales and scores are available, most of the methods are disease-specific, lack sufficient validation or have reported contradictory sensitivity and specificity [Carn 08, Crar 05, List 90, Hill 89]. The CSE is considered to be insensitive to different forms of pharyngeal dysphagia, especially when silent symptoms are present, i.e. in approximately 50% of the cases [Youm 11]. Although non-instrumental tests have shown high sensitivity for detecting aspiration, specificity is too low [Dudi 18a]. Another limitation of the CSE is the subjectivity and questionable reliability because it depends on the professional's expertise and its variability compared to the *gold standard* instrumental methods [Lesl 04].

Finally, after the screening and CSE, the diagnosis is confirmed by instrumental methods such as the *VideoFluoroscopic Study of Swallowing* -VFSS- and the *Fiberoptic Endoscopic Evaluation of Swallowing* -FEES- [Gira 16]. Although the VFSS is considered the gold standard in swallowing examination, it has shown limited inter- and intra-judge reliability [Lesl 04]. Otherwise, several studies have shown that FEES has a high level of agreement with VFSS, which suggests that both methods provide the same effectiveness for patient management [Aviv 00]. Although VFSS and FEES are the reference methods for instrumental assessment, both share the lack of standardized protocols and scoring systems, which makes them highly subjective and expert-dependent [Lang 03]. Furthermore, when VFSS is used several times for follow-up examinations, the radiation-related risks increase, e.g. induced cancers, tissue reactions, and damage in the lens of the eyes [Earl 19]. In addition, this method is sometimes unavailable, time-consuming, and it increases healthcare costs [Wils 12]. Regarding FEES, it provides limited information in comparison with VFSS, because it investigates only the pharyngeal phase and it does not quantify the bolus

inhaled under aspiration conditions [Nacc 08]. Moreover, FEES is uncomfortable and frequently associated with gagging, vomiting, and more rarely, complications such as laryngospasms [Nacc 08]. These drawbacks motivate the research community to address other less invasive and more objective methods [Zora 10].

1.2.3. Biosignals for swallowing evaluation

Aiming to objectify the assessment of swallowing functions and to overcome the limitations related to invasiveness, objectivity, availability, and reliability, different strategies based on signal acquisition and processing have been proposed. Most of them are based on sounds acquired with stethoscopes or microphones, and laryngeal motion recorded with accelerometers [Yagi 17, Sejd 18]. Furthermore, other authors have explored evaluation techniques based on surface electromyography [Vaim 09, McKe 02, Hsu 13], respiratory flow [Lee 09a, Lee 11], piezoelectric sensors [Kala 15], pulse-oximetry [Brit 18], and even voice recordings [Ipin 18].

The aforementioned techniques require different approaches in terms of processing, characterization, modeling, and analysis because they are produced by particular phenomena related to different swallowing *dimensions*, concepts which will be explained in Chapter 3. In this thesis, three biosignals are considered to expand the understanding of dysphagia: surface electromyography, accelerometry-based cervical auscultation, and speech. A brief description of each kind of biosignal is shown next.

Accelerometry-based cervical auscultation (Acc)

In principle, the cervical auscultation signals are acoustic waves generated in the pharyngeal region during the swallowing process [Lee 08]. Such signals in swallowing are produced by propagating pressure waves through the aerodigestive tract with its valves (lips, velopharyngeal region, larynx and cricopharyngeal muscle) [Dudi 15a]. They have been used as a clinical assessment tool for dysphagia screening [Lee 08, Dudi 15d, Mova 17b]. These signals provide information about the frequency and duration of swallowing, the number of gulps, cough, and post-swallow breathing [Laga 16]. Most research related to cervical auscultation in swallowing has been focused on the analysis of the sources of signals, the best placement for signal acquisition, denoising and segmentation methods, and classification between normal and abnormal swallows [Mova 17b]. The use of accelerometers, stethoscopes, and microphones become popular to study the swallowing process; however, there is no consensus about the validity of this approach [Sanc 18, Nozu 17],

Cervical auscultation signals have been mainly acquired using microphones [Yagi 17], and to a lesser extent by stethoscopes [Sanc 18, Lesl 07, Haml 94, Zenn 95, Lesl 04]. One of the main drawbacks of the stethoscope is its poor suitability for transmitting frequencies above 1 kHz [Haml 94]; this is a limitation for analysis of patients with dysphagia because they have shown higher spectral components than healthy individuals [Dudi 18b]. Three events are associated with the main swallowing sounds registered by microphones [Bolz 13, Hann 10, Berg 14]: 1) the laryngeal elevation -and lingual propulsion-, 2) the upper esophageal sphincter opening [Hamm 14], and 3) the laryngeal release, also described as glottal release [Laza 04]. When a flushing sound of the bolus is heard prior to the ini-

tiation of the pharyngeal swallow, or when the breathing sound is wet or stridor after the swallow, it is reasonable to suspect aspiration [Shir 14, Dudi 15a].

Notwithstanding, cervical auscultation using single, dual, or tri-axial accelerometers placed at the cricoid cartilage has been used recently also for cervical auscultation. However, this kind of sensor does not measure the acoustic but the mechanical *dimension* of swallowing; it is intended to detect vibrations caused by movements of oropharyngeal structures in superior-inferior (S-I), anterior-posterior (A-P) and medial-lateral (M-L) axes [Lee 08, Mova 17a]; hyolaryngeal excursion is the primary physiological source of information in this case [Zora 10]. Accelerometry signals have highly concentrated spectral information below 300 Hz, much lower than swallowing sounds [Jest 14, Mova 17b], a strong indication that accelerometry is more related to mechanical than acoustic phenomena [Lee 11, Yagi 17]. Although M-L movements could appear in patients with dysphagia, they are not detected in VFSS and this axis has shown pretty similar behavior to A-P and S-I [Mova 17a]. Entropy analysis suggests that A-P axis provides more information due to its bigger amplitude (signal power) [Lee 08], although maximum hyoid excursion in older individuals is reduced in the anterior direction, and patients could have S-I amplitude larger than A-P [Zora 10, Mova 17a]. Several studies have been oriented to describe only A-P and S-I axes through time, information-theoretic, frequency, and time-frequency domain characterizations [Lee 08, Lee 11, Dudi 18b]. However, the works that reported the use of accelerometers for cervical auscultation have limitations in terms of gender and age matching along with a careful protocol design for the evaluation of different volumes. Furthermore, few works have addressed comparisons between healthy individuals and patients with dysphagia. There is a gap in neurological evaluations of the patients recruited in these studies, which hinder the sensitivity when confirming their condition; in fact, several works reported individuals with suspected rather than confirmed dysphagia. Table 1.2 illustrates the works that have addressed swallowing evaluations using accelerometry only or combined with other biosignals.

Surface electromyography (sEMG)

The sEMG is intended to record the electrical activity of muscles [Poor 17], i.e. to evaluate the electrophysiological *dimension* of swallowing. sEMG has been used over the past decades for the assessment of swallowing impairments [Poor 17], and it is an appropriate technique to detect piecemeal deglutition and delays between phases produced by dysphagia [Erte 14]. sEMG allows to analyze the temporal pattern of sequential and orderly contractions, a neuromuscular characteristic of the swallowing process [Erte 03, Palm 89]. However, sEMG has not been considered so far in the list of evaluation techniques of swallowing pathophysiology [Vaim 09]. In contrast to other methods to study the swallowing process, sEMG is non-invasive, time-saving and inexpensive [Vaim 09]. Identification of swallowing events by sEMG has demonstrated acceptable accuracy and reliability for experienced and naïve judges [Crar 07]. However, its reliability depends on the patient's age and type of ingested bolus [Poor 17]. Furthermore, it shows to be unable to accurately diagnose neurologically induced dysphagia [Vaim 09]. Thus, further validation is necessary in order to standardize its use in clinical practice [Poor 17],

The electrode placement is a critical point to report any study conducted with sEMG [Merl 99]. In swallowing, it has to be consistent with the small size of the involved

Table 1.2: State of the art of engineering related contributions in swallowing evaluation

Paper	Brief data description	sEMG	Acc	Speech	Other	Goal
[Gall 98]	12 HC		•			Event detection
[Laza 04]	15 HC, 11 DP		•			Healthy vs. dysphagic
[Afka 07]	1 HC	•	•		•	Swallow detection
[Sejd 09]	408 HC		•			Swallow segmentation
[Lee 09a]	17 HC		•		•	Swallow segmentation
[Sejd 10]	408 HC		•			Vocalization detection & swallow detection
[Nikj 11]	30 DP		•			Safe vs. unsafe swallows
[Lee 11]	24 DP		•		•	Normal vs. abnormal swallows
[Sejd 12]	408 HC		•			Swallow segmentation
[Sejd 13]	40 DP		•			Healthy vs. dysphagic swallows
[Stee 13]	40 DP		•			Aspiration detection
[Hsu 13]	26 P[a]	•			•	Discrimination of dysphagia severity
[Schu 14]	31 HC, 41 DP	•			•	Swallow detection
[Dudi 15d]	23 DP		•		•	Swallow detection
[Kala 15]	30 HC		•		•	Swallow detection
[Papa 16]	14 HC		•		•	Chewing detection
[Cons 18]	10 HC, 10 P[b]	•				Swallow detection
[Dudi 18a]	55 HC, 53 DP		•			Healthy vs. dysphagic swallows
[Ipin 18]	6 P[c]			•		Dysphagia detection
[Mao 19]	114 DP		•			Anatomical structure tracking
[Rieb 19]	41 DP	•			•	Onset detection
[Stee 19]	305 DP		•			Safe efficient swallow detection
[Dono 20]	116 DP, 15 HC		•		•	Event detection
[Suzu 20]	8 HC	•				Swallow detection
[Khal 20a]	248 DP		•		•	Swallow detection
[Khal 20b]	116 DP		•		•	Event detection
[Dono 21b]	51 HC, 20 DP		•		•	Healthy vs. dysphagic
[Dono 21a]	171 HC, 170 DP		•		•	Compare features event detection
[Dono 21c]	114 DP		•		•	Anatomical structure tracking
[Mao 21]	16 HC, 120 DP		•		•	Event detection
[McNu 21]	5 HC, 5 P[d]	•				Event detection
[Dono 22a]	70 HC		•		•	Reference values & healthy vs. dysphagic swallows
[Dono 22b]	36 HC		•		•	Non-effortful vs. effortful swallows
[Park 22]	449 P[e]			•		Mild vs. severe dysphagia & Risk of respiratory complications
[Zhao 22]	83 HC, 143 DP			•		Dysphagia detection

HC: Healthy controls. DP: suspected of confirmed patients with dysphagia. P: patients with specific conditions.
[a]: Myasthenia gravis. [b]: Head and neck cancer. [c]: Parkinson. [d]: Total laryngectomy. [e]: Post-stroke

muscles [Palm 89, Step 12]. However, there is no a standardized protocol to place the electrodes in swallowing studies [Step 12]. Most of them are focused on the pharyngeal phase [Erte 98, Ding 02, Perl 99] and others on the oral phase [Must 17, Dell 18] separately, but few have assessed the muscle activity during the oropharyngeal phase. These works have evaluated four regions mainly [Vaim 07, Zare 17]: masseter muscles (MS), *orbicularis oris* (OR), and the groups of suprahyoid (SH) and infrahyoid muscles (IH). sEMG in SH captures activity from the anterior belly of the digastric, the mylohyoid, and the geniohyoid muscles, whilst in the IH, sEMG captures activity from sternohyoid, omohyoid, and sternothyroid -see Figure 1.5-. Some protocols acquire only four sEMG channels [Vaim 07], disregarding bilateral differences present in patients with hemiparesis or even in healthy subjects, especially in the oral phase [Dell 18].
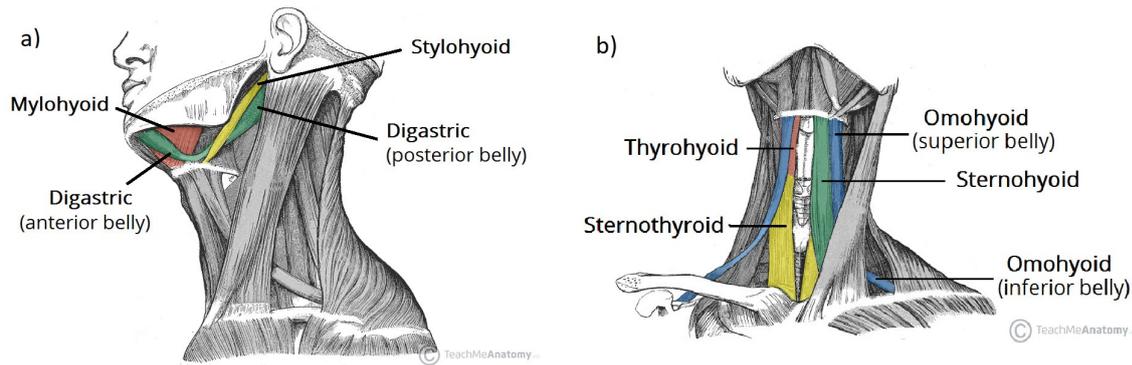
Figure 1.5: Muscles involved in the oropharyngeal phase of swallowing. a) Lateral view of syprahyoid muscles; b) Frontal view of infrahyoid muscles. From [Jone 17b, Jone 17a] with written permission.

The application of sEMG in swallowing has potential as a dysphagia screening clinical tool for optimal patient management [Step 12, Vaim 09] and it is still an open field of research [Dell 18]. However, there are some limitations in the use of sEMG for dysphagia screening: 1) it cannot detect silent aspirations in patients with dysphagia; 2) sEMG can record only the initial part of the esophageal phase [Vaim 09, Vaim 04a]; 3) the gold standard to detect swallowing-related neuromuscular sequences is the visual inspection (VIS) [Vaim 09], which is time-consuming and has moderate reproducibility and repeatability [Cart 15]; and 4) sEMG acquisition in supra- and infra-hyoid muscles is challenging because they have small size and overlapping fibers, which produces cross-talk and low signal-to-noise ratio (SNR) [Step 12, Mona 08]. The latter problem makes the detection of muscle activations (bursts) difficult. Furthermore, the works have investigated only amplitude and temporal related features to characterize and represent the muscle activations during swallowing tasks; there is no consensus about the protocol of acquisition (type of bolus, volumes, repetitions, database), decreasing the outcome of the technique [Vaim 09]; and the databases that have been used are limited in the number of healthy controls and patients, with some exceptions [Aydo 15, Vaim 04b, Vaim 04c, Vaim 04d]. Table 1.2 illustrates the works that have addressed swallowing evaluations using only sEMG or combined with other biosignals, by the application of engineering-related techniques.

Speech

Speech-language pathologists widely use the perceptual evaluation of voice quality during clinical swallow examinations because it gives valuable clues regarding swallowing malfunctioning [Wait 11, Hass 14]. Although Linden et al. suggested that voice abnormalities should alert about aspiration risk [Fest 16], its use as an indicator of laryngeal P/A is controversial [Dudi 15a, Wait 11], and there is a lack of clinical evidence supporting this. Voice-quality assessment is subjective by far, and some authors suggest that evaluation of voice-related perceptual features has variable reliability, especially when the signs are subtle [Dani 15]. Despite the vast literature about dysphagia and voice disorders separately, surprisingly the studies that investigate their association are scarce [Sant 15, Farn 17]. One remarkable exception is the work carried out by Festic et al., who found statistically significant associations between dysphagia and diadochokinetic assessment (rapid repetition

of /puh/,/tuh,/kuh/), especially in patients with confirmed aspiration [Fest 16]. Additionally, few works have addressed the evaluation of speech-related information in patients with dysphagia from an engineering point of view (see Table 1.2). From the list of works, there is one that reported the use of classical machine learning algorithms and voice features to detect the severity of dysphagia in terms of oral feeding and the risk of respiratory complications in post-stroke patients [Park 22]. Another paper implemented feature engineering with classical speech-related descriptors extracted from a throat vibrator instead of microphones to detect dysphagia with machine learning algorithms [Zhao 22].

Two voice-quality alterations (clinical or perceptual features) are detected routinely by speech-language pathologists in dysphagic or aspirated patients [Lede 02, Dani 15][5]: wet voice and dysarthria.

The wet voice, also known as *wet phonation* or *gurgly voice* [Kang 17], describes the bubbling sound produced by stasis of secretions, liquids or food in the laryngeal vestibule [Bass 14]. Wet voice is easy to recognize, and the agreement percentage of detection is high, pre- and post-swallow [Dani 15]. Wetness has shown relatively good post-swallow specificity but poor sensitivity [Wait 11, McCu 01b, Samp 14], but Groher et al. found moderately high sensitivity, low specificity and accuracy dependent on the bolus viscosity [Groh 06]. Again, contradictory results about its correlation with P/A are reported in the literature [Fest 16]. Although its reliability is poor by itself [Sant 15], wet voice helps to increase it considerably when combined with cervical auscultation or 3-oz water test [Cavi 10]. Even though the wet voice has been implemented as a clinical predictor of aspiration in clinical care setting [Cavi 10], it remains unclear and debatable whether the wet voice can be used for assessment of dysphagia or risk of aspiration [Kang 17, Chan 12].

Dysarthria, a speech disorder resulting from impaired neuromuscular control, produces weak, uncoordinated, abnormal tone, slowing, and inaccurate oral and vocal movements [Ende 08]. There is a significant co-morbidity and correlation between dysarthria and dysphagia [Morg 10, Fals 09, Lapa 17], because dysarthria produces weakness or incoordination of the tongue and orofacial muscles (the most important articulators), which play a key role in the swallowing process [Kuma 14]. Although there are no known clinical predictors of the binomial dysphagia-dysarthria [Flow 13], dysarthria has been reported as a moderate predictor of dysphagia severity [Dani 97], especially for oral stage problems rather than pharyngeal impairments [Bahi 16]. Furthermore, the capability of dysarthria to predict P/A in clinical evaluations is controversial [Keag 17, Okub 12], but these conclusions have been made in limited pathologies.

The presence of speech-related alterations in patients with dysphagia is supported by the fact that both processes, i.e. speech production and swallowing, share anatomical structures and some networks at the neurological level; particularly, the glossopharyngeal nerve (co-responsible for the oropharyngeal swallow response through palatal elevation) and the vagus nerve (primary peristalsis of the esophagus), participate in voice production (phonation, articulation and prosody) [Clav 15, Erma 09, Berg 14]. Figure 1.6 shows a flowchart of the functional and anatomical relationship between the swallowing and speech production processes.

---

[5]The other clinical identifiers of aspiration risk are dysphonia, abnormal volitional cough, abnormal gag reflex, and cough after swallow [Lede 02].
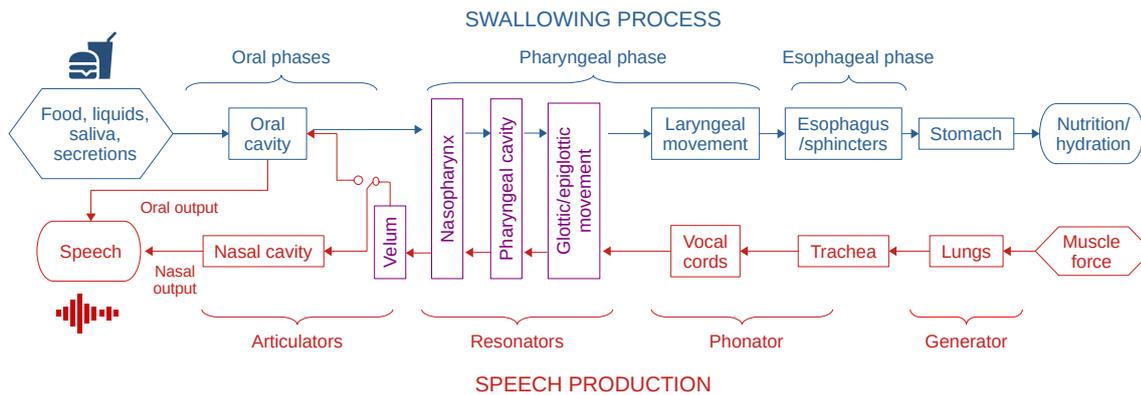
Figure 1.6: Swallowing (blue) and speech (red) processes in a counter way. The purple structures are shared by both processes. Relevant structures involved in each phase of the swallowing and speech production processes are included.

Open questions in swallowing-related biosignals

In general, the aforementioned signals have shown limitations that have impeded their use in the consulting room for the evaluation of dysphagia. One of these limitations is the contradictory reliability when assessing dysphagia. Another restraint is the number of individuals recruited for each experiment published in the literature. Most of the limited studies that assessed both healthy controls and patients with dysphagia, either lack age and gender matching, or report descriptive results only, i.e. such works do not provide performance measures such as sensitivity or specificity. Systematic scientific evaluation and validation are required, also the number of participants needs to be increased to report stronger conclusions. On the other hand, the swallowing process lacks of standardized characteristics (features or descriptors) to quantitatively describe each swallowing *dimension*. Works that intended to characterize different non-invasive biosignals report contradictions about which descriptors represent well swallowing-related phenomena under different conditions such as gender, age, swallowing task, or dysphagia severity. Additionally, there are not confident models based on non-invasive biosignals with reliability comparable to the reference methods, which limits their use for swallowing evaluation.

1.3. Contributions to the research on swallowing evaluation

This thesis proposes different machine learning-based methodologies for the evaluation of functional oropharyngeal dysphagia. Two acquisition protocols were designed, one for the evaluation of electromyography and accelerometry during different swallowing tasks, and another for the evaluation of speech recordings before and after such tasks. In age and gender matched databases of healthy individuals and patients with confirmed functional oropharyngeal dysphagia, different feature domains were extracted from each biosignal, so various biomarkers of normal and abnormal swallowing processes are proposed. Additionally, machine learning and deep learning-based models were implemented and evaluated to figure out how each biosignal is capable to represent each swallowing *dimension* under

healthy and dysphagic status. Furthermore, a multimodal approach of the three sources of information, i.e. sEMG, accelerometry-based cervical auscultation, and speech recordings, was performed in order to increase the generalization capability of the implemented models. All of the proposed biomarkers and models were evaluated quantitatively by the computation of clinically accepted performance measures intended to evaluate diagnostic tests.

This work proposes an approach for dysphagia evaluation with non-invasive, low cost and objective strategies, with a combination of signals not previously proposed. Since this approach is exploratory, its results are preliminary and it is not intended to replace invasive reference methods for swallowing examination. In fact, this work is in the first stage in terms of the development of diagnostic tests (an exploratory study in descriptive discrimination stage [Zhou 09]), in which an exploration of the behavior of the signals in healthy and dysphagic populations is performed (see Figure 1.7). However, the advantage of this non-invasive strategy is its possibility to perform multiple evaluations in cross-sectional studies and long-term follow-up examinations without risk of radiation and other risks related to the existing methods. This work has potential application in the screening phase of the dysphagia evaluation, previous to the clinical bedside swallow examination and instrumental evaluation by VFSS (see Figure 1.7), which is typically used to identify patients with functional oropharyngeal dysphagia associated with neurological and neuromuscular disorders.
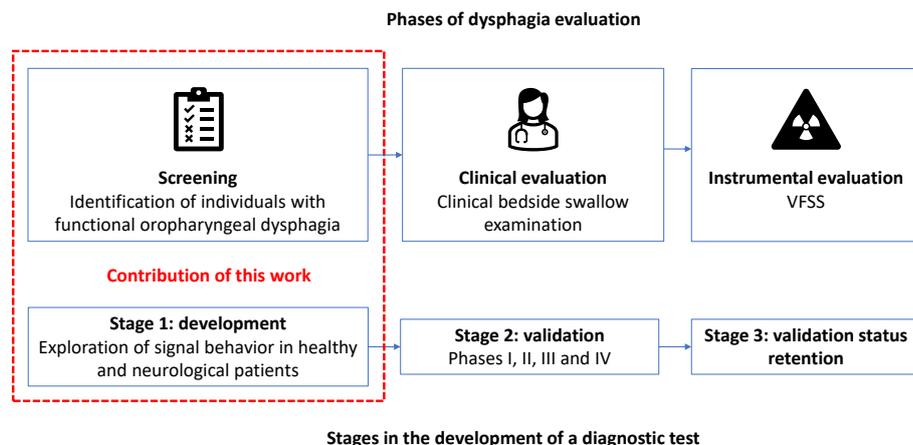
Figure 1.7: Location of the contributions of this work within the phases involved in the dysphagia evaluation and the stages in the development of diagnostic tests.

1.4. Structure of this work

Chapter 2 describes the clinical and epidemiological data of the volunteers recruited for each experiment. Inclusion and exclusion criteria are provided. A detailed description of the acquisition protocol for each biosignal is also provided.

Chapter 3 introduces the concept of swallowing dimensions and explains how this thesis contributes to their analysis. This chapter presents a review of the features applied to sEMG, Acc, and speech-related signals in the context of swallowing evaluation.

Furthermore, a detailed mathematical description of the feature domains and feature selection methods implemented in this work is provided.

Chapter 4 provides a detailed description of the experiments performed related to the selection of features and contribution of each sEMG acquisition channel, Acc axis, and speech dimension. The description of the results is followed by an extensive discussion of the proposed swallowing biomarkers.

Chapter 5 includes the detailed mathematical formulation of the discrimination models of machine learning and deep learning implemented in this work, oriented to the classification of healthy and dysphagic states. It also describes the validation methods to obtain generalized results.

Chapter 6 provides a detailed description of the experiments performed regarding classification/detection performance. The description of the results is followed by an extensive discussion of algorithms and models.

Chapter 7 presents a general description of lessons learned during the protocol design and data collection, as well as an outlook on future research in swallowing evaluation driven by objective methods from engineering-related fields.

Chapter 8 summarizes the main contributions of this thesis in the context of dysphagia screening and the repercussion of the experimental results.

Chapter 9 summarizes the publications emerging from the development of this thesis.

Chapter 2

Data collection

2.1. Database description

Two groups of subjects were evaluated in this thesis: one with healthy individuals and another one with patients with confirmed diagnosis of functional oropharyngeal dysphagia produced by neurological or neuromuscular etiologies. All subjects were selected by convenience sampling. The following inclusion criteria were set for each group:

- *Healthy individuals*: age above 18 years old, male or female, healthy; absence of diagnosed dysphagia; and absence of central or peripheral neuropathies, or neuromuscular pathologies.

- *Patients with dysphagia*: age above 18 years old, male or female; clinical manifestations of oral or oropharyngeal dysphagia; confirmed diagnosis of neurological or neuromuscular etiology, responsible of oral or oropharyngeal dysphagia; Spanish EAT-10$\geq$3.

Additionally, the following exclusion criteria were defined:

- *Healthy individuals*: the presence of dental pathology; the presence of congenital malformations in the mouth; to have active inflammatory processes in the mouth, head or neck; to have strange elements in the mouth such as piercing, retainers, braces, or dental prosthesis; to have diagnosed cognitive impairment; and to have cardiorespiratory impairments.

- *Patients with dysphagia*: the presence of esophageal or mechanical dysphagia; patient with active treatment for cancer in facial or cervical regions; the presence of orofacial or cervical edema or hematoma; recent surgical dissection in face or neck (the last three months); severe hypoxemia (oxygen saturation below 80%); and to have a deep brain stimulation device.

The patients and controls who fulfilled these criteria are described in Tables 2.1 and 2.2, respectively. Each subject signed a language-specific informed consent approved by the Ethics Committee of the Universidad Pontificia Bolivariana (Medellín, Colombia), by authorization issued on 07.01.2017. The Ethics Committee is conformed by the internal Resolution 80 of 12.17.2008.

Table 2.1: Clinical information of the patients with dysphagia. The use of data in the different experiments is indicated with bullets.

Code	Age	Sex	Etiology	EAT-10	#1	#2	#3	#4	Code	Age	Sex	Etiology	EAT-10	#1	#2	#3	#4
P001	49	F	Myasthenia gravis	3	•	•	•	•	P052	52	F	Motor neuron disease	13	•	•	•	•
P002	76	M	Motor neuron disease	3			•	•	P053	27	M	Myasthenia gravis	28	•	•		•
P004	39	M	Dementia	14	•	•		•	P054	59	M	Cerebral palsy	13			•	•
P005	62	M	Motor neuron disease	18			•	•	P055	31	M	Cerebral palsy	9	•	•		•
P006	19	M	Duchenne muscular dystrophy	33		•		•	P056	67	F	Amyotrophic lateral sclerosis	27				•
P007	71	F	Dementia	36			•	•	P057	26	M	Rubinstein-Taybi syndrome	21		•		•
P008	76	F	Dementia	24			•	•	P058	26	M	Stroke	4	•	•		
P010	20	M	Stroke	34		•		•	P059	28	F	Cerebral palsy	13	•	•		•
P011	50	M	Neuropathy	15	•		•	•	P060	31	M	Stroke	15	•	•		
P013	71	M	Dementia	35			•	•	P063	84	F	Parkinson's disease	15				•
P014	41	M	Ischemic stroke	35	•	•	•	•	P064	65	F	Neurocognitive disorder	13		•	•	•
P015	68	F	Ischemic stroke	20			•	•	P065	80	M	Dementia	13			•	•
P016	42	M	Ischemic stroke	ND	•	•	•	•	P067	54	F	Myasthenia gravis	10	•	•		•
P017	57	M	Dementia	31	•		•	•	P070	40	M	Dementia	8	•	•	•	•
P018	50	F	Myasthenia gravis	13	•	•	•	•	P081	66	M	Dementia	ND			•	
P019	44	F	Ischemic stroke	29	•	•	•	•	P082	20	M	Muscular dystrophy	15	•	•	•	•
P020	68	M	Dementia	9			•	•	P083	67	F	Ischemic stroke	10				•
P021	65	F	Muscular dystrophy	ND				•	P086	79	F	Ischemic stroke	35			•	•
P022	61	M	Dementia	13			•	•	P087	60	F	Inflammatory myopathy	25	•	•	•	•
P023	73	M	Myasthenia gravis	40			•	•	P088	66	F	Dementia	4			•	•
P025	67	M	Myasthenia gravis	8			•	•	P089	65	F	Dementia	10			•	•
P026	53	F	Myasthenia gravis	15	•	•	•	•	P090	42	M	Muscular dystrophy	12		•	•	•
P027	77	M	Ischemic stroke	12				•	P091	38	M	Muscular dystrophy	17		•		•
P028	64	F	Motor neuron disease	20	•	•			P092	60	M	Ischemic stroke	9			•	•
P029	65	F	Amyotrophic lateral sclerosis	10		•	•	•	P095	50	M	Amyotrophic lateral sclerosis	17				•
P030	64	M	Dementia	3			•		P096	67	M	Ischemic stroke	17				•
P031	67	F	Dermatomyositis	35				•	P097	65	F	Neurocognitive disorder	16				•
P032	58	M	Ischemic stroke	14	•		•	•	P098	69	M	Progressive Supranuclear Palsy	22				•
P033	71	F	Dementia	21			•	•	P099	28	M	Cerebellar ataxia	17				•
P036	48	F	Myasthenia gravis	6	•	•	•	•	P100	64	F	Frontotemporal dementia	18				•
P037	70	M	Ischemic stroke	16				•	P101	73	M	Motor neuron disease	17				•
P038	19	M	Cerebral palsy	32		•			P102	68	M	Motor neuron disease	22				•
P039	75	F	Neuropathy	8			•	•	P103	65	M	Parkinson's disease	17				•
P040	33	M	Stroke	13	•	•		•	P104	61	M	Parkinson's disease	15				•
P041	77	M	Thalamotomy (essential tremor)	9				•	P105	37	F	Muscular dystrophy type 1	13				•
P042	68	F	Myasthenia gravis	16			•	•	P106	46	M	Stroke grave	5				•
P043	55	M	Dementia	18	•		•	•	P107	60	M	Parkinson's disease	9				•
P044	66	F	Cerebellar ataxia	18			•	•	P108	73	M	Parkinson's disease	10				•
P045	55	M	Myasthenia gravis	17	•		•	•	P109	71	F	Amyotrophic lateral sclerosis	25				•
P046	53	F	Myasthenia gravis	19	•	•	•	•	P111	57	M	Parkinson's disease	8				•
P047	58	F	Motor neuron disease	27	•	•	•	•	P112	66	F	Juvenile Parkinson's disease	23				•
P048	78	M	Amyotrophic lateral sclerosis	18			•	•	P113	69	M	Parkinson's disease	16				•
P049	57	M	Ischemic stroke	26	•				P114	50	F	Ischemic stroke	12				•
P051	50	M	Motor neuron disease	9	•		•	•									
TOTAL					19	16	33	40	**TOTAL**					10	14	13	40

Surface electromyography (sEMG), accelerometry-based cervical auscultation (Acc), and speech recordings were collected from two groups of volunteers divided into healthy individuals (controls) and patients with functional oropharyngeal dysphagia. Patients were also classified according to neurological or neuromuscular conditions. Four experiments were performed:

- Experiment #1: Only sEMG (unimodal)

- Experiment #2: sEMG and Acc (bimodal)

- Experiment #3: Only speech (unimodal, pre-swallowing recordings)

- Experiment #4: sEMG, Acc and -post swallowing- speech (multimodal)

The experiments were carried out as the total database was built, which meant that each one of them was performed using a subset of the total volunteers. Although one single database was created, this was partitioned according to the experiment requirements, such as the balance of sex and age. Next, the demographic data of the volunteers and the acquisition protocol designed for each biosignal are presented in Tables 2.1 and 2.2.

Table 2.2: Demographic information of the healthy individuals. The use of data in the different experiments is indicated with bullets.

Code	Age	Sex	Experiment #1	#2	#3	#4
C482	54	F				•
C537	28	M	•			
C541	59	F	•			
C542	28	M	•			
C543	58	M	•			
C550	36	M	•		•	
C553	84	F	•		•	
C554	62	F	•		•	
C555	32	M	•			
C556	29	F	•			
C563	26	F		•		
C564	34	M	•		•	
C565	30	F	•	•		
C568	32	M	•	•		
C571	32	F		•		
C572	26	M		•		
C575	30	M		•		
C578	28	F	•	•		
C585	29	M	•	•		
C587	26	F		•		
C588	25	F		•		
C590	34	M	•	•		
C591	28	M	•	•		
C592	26	M		•		
C599	32	M	•	•		
C608	27	M		•		
C613	25	F		•		
C616	51	F	•	•	•	
C618	27	F		•		
C628	54	F	•	•		
C629	33	F	•			
C630	66	F	•		•	
C631	63	F	•	•	•	
C632	76	M	•	•		
C633	55	M	•	•	•	•
C634	68	M	•	•	•	•
C635	38	M	•	•	•	
C636	35	F	•		•	
C637	26	F		•		
C638	57	F	•	•	•	•
C639	53	M	•	•	•	•
C640	43	M	•	•	•	•
C642	57	F		•	•	
C643	49	F	•	•	•	
C646	52	M				•
C647	38	F				•
C648	58	M				•
C649	55	M				•
C650	48	F				•
C651	40	M				•
TOTAL			31	30	16	12

Code	Age	Sex	Experiment #1	#2	#3	#4
C653	39	M				•
C655	54	M				•
C657	51	F				•
C658	53	F				•
C660	48	F				•
C661	51	F				•
C665	78	F				•
C666	75	F				•
C667	22	F				•
C668	29	M				•
C669	23	M				•
C670	35	M				•
C671	61	F				•
C672	59	F				•
C673	62	M				•
C674	64	M				•
C675	22	M				•
C680	58	F				•
C681	61	F				•
C682	22	M				•
C683	64	F				•
C684	56	F				•
C686	73	M				•
C687	68	M				•
C688	53	M				•
C689	42	F				•
C690	68	F				•
C691	64	F				•
C692	68	F				•
C693	56	F				•
C694	70	F				•
C695	71	F				•
C696	90	F				•
C697	58	M				•
C698	68	F				•
C699	65	F				•
C700	60	M				•
C701	62	M				•
C702	60	F				•
C703	76	M				•
C704	63	M				•
C705	59	M				•
C706	72	M				•
C707	62	M				•
C708	65	F				•
C709	65	F				•
C710	64	M				•
C711	61	F				•
C712	72	M				•
C713	64	M				•
TOTAL			0	0	0	50

Code	Age	Sex	Experiment #1	#2	#3	#4
C714	60	F				•
C715	62	F				•
C716	67	F				•
C717	66	M				•
C718	75	F				•
C719	58	M				•
C720	66	F				•
C721	82	F				•
C723	69	M				•
C724	60	M				•
C725	66	F				•
C726	63	F				•
C727	61	M				•
C728	69	F				•
C729	64	F				•
C730	63	M				•
C731	68	F				•
C732	66	F				•
001E*	49	M			•	
002E*	69	F			•	
003E*	72	M			•	
004E*	53	F			•	
005E*	50	F			•	
006E*	64	M			•	
007E*	63	F			•	
008E*	70	F			•	
009E*	66	M			•	
010E*	53	F			•	
011E*	71	F			•	
012E*	63	F			•	
016E*	50	F			•	
018E*	65	M			•	
020E*	75	M			•	
021E*	73	M			•	
022E*	78	M			•	
023E*	63	M			•	
024E*	70	M			•	
025E*	68	M			•	
026E*	70	M			•	
029E*	79	M			•	
030E*	50	F			•	
031E*	56	M			•	
033E*	68	M			•	
034E*	64	F			•	
035E*	62	F			•	
036E*	74	F			•	
037E*	53	F			•	
038E*	56	F			•	
TOTAL			0	0	30	18

* Healthy controls of the PC-GITA database, composed of Colombian Spanish speakers [Oroz 14]. Only pre-swallowing recordings were analyzed (see Section 4.4).

2.1.1. Database description per experiment

Clinical and demographic data for each experiment are detailed below, including the matching-related distribution and statistical tests.

Experiment #1   The database consisted of two groups: one with 31 healthy subjects (14 female and 17 male, 45.29 ± 16.22 years old), and the other one with 29 patients with dysphagia (13 female and 16 male, 45.69 ± 11.92 years old). Both groups were matched by age [Mann-Whitney U test, $Z_{0.95} = 0.17, p = 0.865$] and sex [$\chi^2(95) = 6.74 \times 10^{-4}, p = 0.979$]. Further details are provided in Table 2.3.

Table 2.3: Clinical and demographic data of the experiment #1. The age is given in mean $\pm$ standard deviation.

	Patients	**Controls**
Sex [M/F]	16/13	17/14
Age [M/F]	43.25 ± 11.45 / 48.69 ± 12.25	41.41 ± 15.02 / 50.00 ± 16.90
Etiology [M/F]	Multiple sclerosis: 2/6 Ischemic stroke: 4/1 Dementia: 4/0 Motor neuron disease: 1/3 Traumatic brain injury: 3/0 Muscular dystrophy: 0/1 Cerebral palsy: 1/1 Neuropathy: 1/0 Inflammatory myopathy: 0/1	N/A

Experiment #2   The database comprised two groups: one with 30 control individuals with normal deglutition (15 female and 15 male, 39.10 ± 15.05 years old), and the other one with 30 patients with dysphagia (15 female and 15 male, 41.23 ± 14.45 years old). Both groups are matched by age [Mann-Whitney U test, $Z_{0.95} = 0.666, p = 0.505$] and sex [$\chi^2(95) = 0, p = 1.00$]. Further details are provided in Table 2.4.

Experiment #3   Two groups of Colombian Spanish native speakers were assessed: one with 46 healthy subjects (23 female and 23 male, 60.17 ± 11.93 years old), and the other one with 46 patients with dysphagia (23 female and 23 male, 60.04 ± 12.37 years old). Both groups were matched for age [$t(99) = -0.0515, p = 0.96$] and sex [$\chi^2(95) = 0, p = 1.00$]. Further details are provided in Table 2.5.

Table 2.4: Clinical and demographic data of the experiment #2. The age is given as mean ± standard deviation.

	Patients	Controls
Sex [M/F]	15/15	15/15
Age [M/F]	31.60±8.49 / 50.87±12.71	39.80±15.98 / 38.40±14.59
Etiology [M/F]	Multiple sclerosis: 1/6 Cerebrovascular disease: 2/1 Neurocognitive disorder: 1/1 Motor neuron disease: 0/4 Traumatic brain injury: 4/0 Muscular dystrophy: 3/0 Cerebral palsy: 2/1 Myopathy: 0/2 Parkinson disease: 1/0 Cerebellar ataxia: 1/0	N/A

Table 2.5: Clinical and demographic data of the experiment #3. The age is given as mean ± standard deviation.

	Patients	Controls
Sex [M/F]	23/23	23/23
Age [M/F]	59.8±11.9/60.3±13.0	61.3±13.4/61.2±10.2
Etiology [M/F]	Ischemic stroke: 4/3 Dementia: 9/6 Muscular dystrophy: 1/2 Spinocerebellar ataxia: 0/1 Motor neuron disease: 4/3 Multiple sclerosis: 2/6 Myastenia gravis: 1/0 Neuropathy: 1/1 Cerebral palsy: 1/0 Inflammatory myopathy: 0/1	N/A

Experiment #4 Two groups of Colombian Spanish native speakers were included: one with 80 healthy subjects (43 female and 37 male, 58.90 ± 13.32 years old), and the other one with 80 patients with dysphagia (35 female and 45 male, 57.41 ± 15.73 years old). Both groups were matched for age [Mann-Whitney U test, $Z_{0.95} = 0.101, p = 0.92$], and sex [$\chi^2(95) = 1.60, p = 0.21$]. Further details are provided in Table 2.6.

2.2. Protocol of swallowing and speech tasks

Prior to data collection and signal acquisition, all subjects were evaluated by a neurologist. In this stage, inclusion and exclusion criteria were verified. Additionally, a neurolo-

Table 2.6: Clinical and demographic data of the experiment #4. The age is given as mean $\pm$ standard deviation.

	Patients	**Controls**
Sex [M/F]	37/43	45/35
Age [M/F]	54.51$\pm$18.26/60.17$\pm$13.69	55.26$\pm$16.99/61.74$\pm$11.66
Etiology [M/F]	Ischemic stroke: 13/5	
	Dementia: 8/6	
	Myastenia gravis: 4/7	
	Parkinson's disease: 8/2	
	Motor neuron disease: 5/3	
	Amyotrophic lateral sclerosis: 2/3	
	Muscular dystrophy: 3/2	
	Cerebral palsy: 4/1	N/A
	Neurocognitive disorder: 0/2	
	Neuropathy: 1/1	
	Spinocerebellar ataxia: 1/1	
	Inflammatory myopathy: 0/1	
	Dermatomyositis: 0/1	
	Rubinstein-Taybi syndrome: 1/0	
	Essential tremor: 1/0	

gical evaluation was made by a neurologist or by a neuro-rehabilitation specialist with the aim to validate inclusion and exclusion criteria. The analysis of clinical variables collected from this evaluations is out of the scope of this thesis.

For the acquisition of sEMG and Acc signals, the protocol introduced by [Samp 14] was adapted, and subjects were asked to swallow the following consistencies: 5, 10, and 20 mL of yogurt (namely yogurt$_5$, yogurt$_{10}$, and yogurt$_{20}$, respectively), saliva (dry swallow), 5, 10 and 20 mL of water (namely water$_5$, water$_{10}$, and water$_{20}$, respectively), and 3 g of cracker approximately. The latter was not used to acquire Acc signals due to the high propagation of the masticatory vibrations. These consistencies have been used to assess penetration/aspiration in dysphagia [Samp 14]. Water and yogurt were delivered to the oral cavity with a 1.5 oz cup. In patients, only safe consistencies and volumes were delivered by a trained speech-language pathologist. One swallow per consistency was executed, in order to prevent fatigue-like effects associated with multiple swallows while guaranteeing the majority of swallowing tasks per subject.

Speech recordings were collected before and after the aforementioned swallowing tasks. The following speech tasks were recorded, based on the protocol described in [Vasq 18] for assessment of patients with Parkinson's disease:

- Sustained vowels /a/, /e/, /i/, /o/, and /u/, pronounced during at least 3 seconds

- Diadochokinetic task: rapid repetitions of the syllables pa-ta-ka

- Reading: a phonetically balanced text which contains all of the Spanish sounds (spoken in Colombia) [Oroz 14], is read by the subject. The text is as follows: *Ayer fui al médico. ¿Qué le pasa? Me preguntó. Yo le dije: ay doctor, donde pongo el*
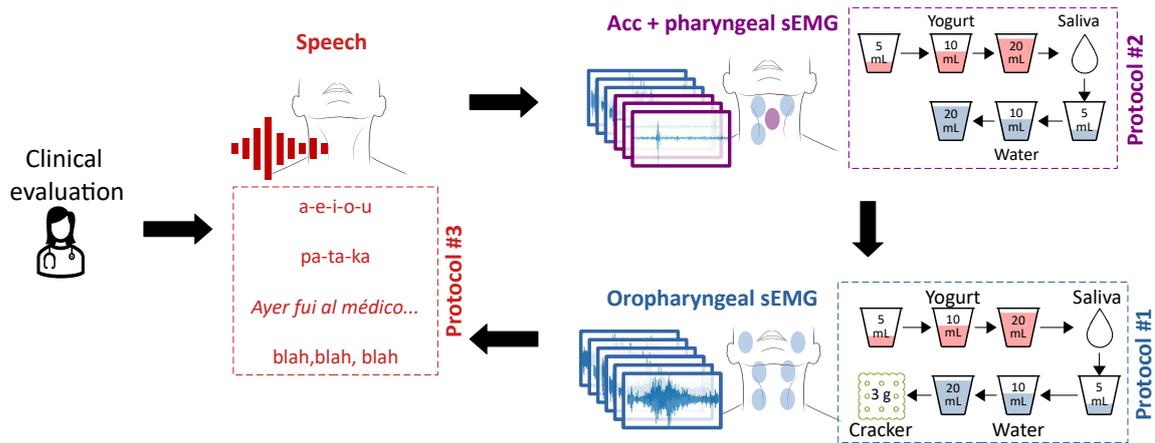
Figure 2.1: Acquisition protocol.

dedo me duele. - ¿Tiene la uña rota? - Sí. - Pues ya sabemos qué es. Deje su cheque a la salida [Translation: Yesterday I went to the doctor. What happened to you? He asked me. I told him: ah doctor! Where I put my finger it pains me. - Do you have a broken nail? - Yes. - Then we now know what is happening. Leave your check at the exit.]

- Continuous speech: a spontaneous monologue with approximately 90 s of duration, in which participants speak about what they did during the current day or week, their family, their job, or their interests

Figure 2.1 summarizes the complete protocol for data collection. Henceforth, this protocol will be divided into three minor protocols for readability:

- Protocol #1: oropharyngeal sEMG only

- Protocol #2: cervical accelerometry and supra/infrahyoid sEMG

- Protocol #3: speech recordings only

The instrumentation used for data collection is described in the following lines.

2.3. Instrumentation

2.3.1. Multi-channel sEMG acquisition

The electrical activity of the bilateral masseters, suprahyoid and infrahyoid muscles (see Figure 2.2), was acquired with the Noraxon $Ultium^{TM}$ EMG (Noraxon USA, CMMR > 100 dB, 16 bits A/D converter) and non-polarizable, bipolar, disposable and pre-gelled Ag/AgCl electrodes (Ref. 2228, 3M - 30 mm x 35 mm, 15 mm diameter in gel area, and interelectrode distance of 25 mm). The electrodes are placed in the most significant positions for evaluation of swallowing movements [Zare 17].

The swallowing tasks were video-recorded synchronously frame by frame with the sEMG signals to visualize the hyoid movement.
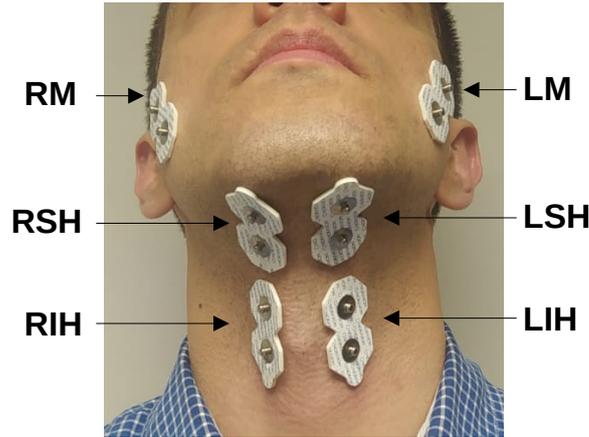
Figure 2.2: Placement of sEMG electrodes. RM and LM: right and left masseters, respectively; RSH and LSH: right and left suprahyoid muscles, respectively; RIH and LIH: right and left infrahyoid muscles, respectively.

sEMG signals were acquired with a sampling rate of $F_s = 2\,$kHz. Pre-processing, storage, and visualization of the raw signals were carried out with the MR3 software (Noraxon USA). Offline analysis was performed using custom scripts and open-source functions in Python 3. Signals were filtered with a 5th order band-pass Butterworth filter with cut-off frequencies of 10 and 500 Hz [Merl 99].

Signals were also filtered with a denoising method based on the Discrete Wavelet Transform. Optimal parameters for the sEMG related denoising were found in a study in the framework of this thesis and published in [Rold 20]: mother wavelet db5, five decomposition levels, soft thresholding, and minimax rule for threshold selection. See Appendix B for details.

2.3.2. Tri-axial Acc acquisition

The kinematic (mechanical) related activities were collected by Acc signals with the tri-axial accelerometer MMA7361 (NXP, Eindhoven, The Netherlands), the NI-DAQ 6215 (National Instruments), and the custom software MODAC (Acquisition Module of Accelerometry, Instituto Tecnológico Metropolitano, Medellín, Colombia). A sampling rate of 10 kHz and a bandpass filter with cut-off frequencies of 0.1 Hz and 3 kHz were applied [Dono 21b]. The accelerometer was placed on the cricoid cartilage (see Figure 2.3). This sensor measures laryngeal movements during swallowing in three axes: anterior-posterior (AP), superior-inferior (SI), and medial-lateral (ML). Notice that upward-forward movements are analyzed to describe the physio-mechanical characteristics of the pharyngeal phase [Mats 08b].

Likewise for sEMG, wavelet denoising was applied for Acc signals to maximize their signal-to-noise ratio (SNR). The SNR was computed from spectrograms in which the presence of background noise and swallow events were manually labeled. Since all parameters exhibited a similar behavior in the three axes, the following combination was selected: mother wavelet db2, nine decomposition levels, soft threshold and heuristic SURE for threshold selection, and no re-scaling.
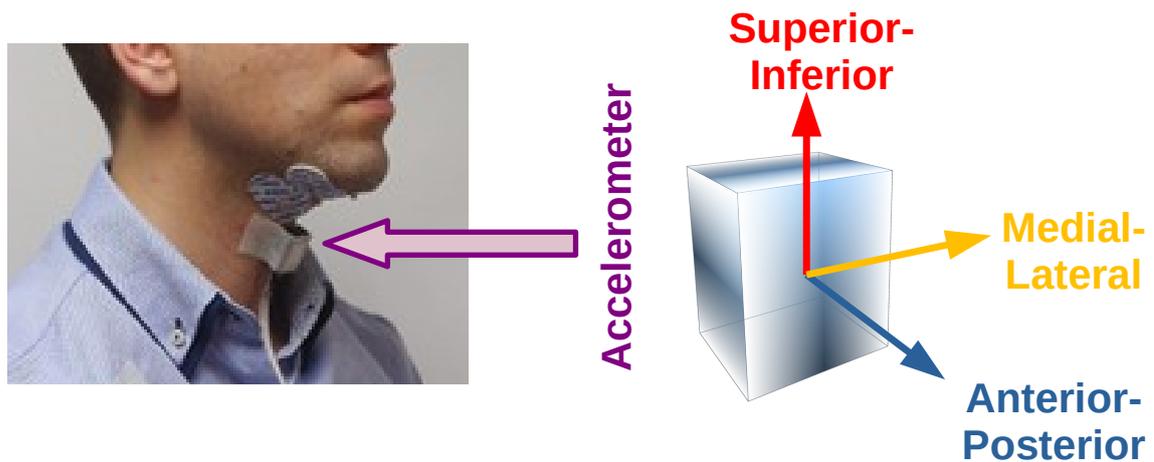
Figure 2.3: Placement of the accelerometer on the neck area. The three axes are also illustrated.

2.3.3. Speech signals acquisition

Speech signals were acquired with a Logitech H390 USB headset which is equipped with a noise-canceling microphone. The microphone was placed parallel to the lip's commissure (see Figure 2.4). The audio recordings were collected using the software Audacity® at a sampling rate of 44.1 kHz with 16 bit-resolution.

All signals were analyzed with custom codes in Python. Aiming to create homogeneous acoustic conditions along the complete set of speakers, a channel normalization process was applied, intended to simulate channel conditions of a regular mobile network. This normalization is based on the Global System for Mobile Communications standard (GSM) full-rate compression. The channel normalization process was performed with the Sound eXchange (SoX) software, as follows [ETSI 00]: each signal was converted to monophonic, its bit rate was reduced to 13 bps and down-sampled to 8 kHz, the compression factor was set at 8, and the band-pass filter was set between 0.2 and 3.4 kHz (bandwidth of a telephone channel). After the acoustic normalization procedure, the features described in Section 3.3.2 were extracted.

Figure 2.4: Placement of the headset for speech acquisition.

Chapter 3

Characterization of biosignals

This chapter introduces the concept of swallowing dimensions, which is created to address the dysphagia-related alterations comprehensively. Afterward, the context of the characterization of the signals evaluated in this thesis is presented. Finally, a detailed mathematical formulation of the features extracted in this thesis is provided.

3.1. Swallowing dimensions

The different phenomena implied in the swallowing process can be compartmentalized by well-defined deglutition aspects, in other words, by different ***swallowing dimensions***. Although this name has not been previously defined or used neither in the medical nor in the engineering-related field, this concept is proposed to address the swallowing phenomenon having in mind the interconnected anatomical structures and physiological events as well. The criteria to define a swallowing dimension are established as follows:

- To have differential aspects with respect to the other dimensions

- To have at least one assessment technique to evaluate it

- To be affected in well-described ways by the dysphagia-related etiologies

- To -potentially- allow the definition of normality patterns

With the aim to contribute to the establishment of a formal analysis for this phenomena, the proposed swallowing dimensions are explained next.

Electrophysiological: This dimension gives information about the state of the electrical communication between the nervous system and the muscles required to execute the normal deglutition. The electrophysiological study of swallowing investigates patterns in timing, activation amplitudes, and sequentiality in the oral and pharyngeal phases under different neurological conditions [Alfo 13]. The electrophysiology of swallowing can be assessed by needle or surface electromyography [Step 12, Suzu 20], and mechanomyography [Cons 17].

Mechanical: This aspect is related to kinematic events performed during the swallowing, such as hyoid bone and larynx movements, as well as opening and closure of the laryngeal vestibule and the upper esophageal sphincter [Dono 22a]. This dimension allows analyzing temporal patterns in all phases, although most of the related studies focused on the pharyngeal one. The -high resolution- accelerometry based cervical auscultation [Dono 22b], piezoelectric sensors [Kala 15], and VFSS [Zhan 21], provide information about the mechanical swallowing dimension.

Acoustic: This dimension is related to the sounds produced by different events during swallowing, particularly the glottic closure [Sanc 18]. The duration, intensity, and frequency of the acoustic swallowing signal are intended to analyze the pharyngeal phase mainly [Youm 05]. In general, sensors placed at the patient's neck are used for acoustic evaluation of swallowing [Youm 11]; the stethoscope and microphone-based cervical auscultation [Haml 94, Dudi 18b] are the most common methods for the acoustic evaluation of swallowing, but the accelerometry based auscultation has been also investigated from an acoustical point of view [Dudi 18b]. This dimension also covers speech production. The main assumption is that a person with swallowing problems might produce abnormal movement while swallowing and also while speaking. However, few dysphagia-related works have addressed acoustic studies based on speech recordings [Ipin 18].

Neurological: It describes only the central and peripheral neurological components rather than the effectors. The central components of the neurological dimension are cortical, nuclear, brainstem, cerebellum, and spinal cord [Suar 18]; the peripheral components are the cranial nerves, neuromuscular junctions, and the enteric nervous system [Suar 18]. Evaluations of the neurological dimension are performed by CSE and instrumental methods. The physical CSE identifies signs of oropharyngeal dysphagia by the neurological evaluation of the head and neck, as well as inspection, palpation, and auscultation of structures from the digestive and upper respiratory system, face, and neck [Suar 18]. Moreover, instrumental evaluations based on electroencephalography allow mapping of brain networks during swallowing in different conditions [Jest 16], whilst functional Magnetic Resonance Imaging and Positron Emission Tomography allow examinations of cortical representations of swallowing [Erte 03].

Cardiorespiratory: The normal swallowing shares anatomical space with breathing, so both processes must be well coordinated in order to protect the airway through the predominant pattern "exhale-swallow-exhale" [Mats 09]. When this coordination fails during dysphagia, the food or fluid could enter to the airway in aspiration episodes, producing bronchospasms or airway obstruction, with the subsequent mismatch of the ventilation-perfusion and reduction in oxygen saturation [Mari 17]. This justifies the proposal of a cardiorespiratory dimension, which can be assessed by pulse-oximetry (even though its use has been systematically rejected [Mari 17]), nasal airflow measurements [Inou 18], plethysmography [Mats 08a], and arterial blood gases [Han 19].

Symptomatic: The signs and symptoms of abnormal swallowing are well characterized [Fine 03, Clav 15]. Early screening in populations at risk is a crucial step of an

appropriate dysphagia treatment [Spey 13]. This dimension is mainly evaluated by CSE in two ways, by the functional health status (FHS) and the health-related quality of life (HRQoL) [Spey 13]. FHS aims to identify how the dysphagia affects particular functional aspects, whilst HRQoL is intended to observe how the patient perceives such condition not only in terms of the physical burden but also in the economic, psychological, and social aspects (see Figure 1.4). FHS and HRQoL are evaluated by anamnesis and self-administered questionnaires (or scales), respectively. However, FHS and HRQoL are frequently evaluated jointly in clinical practice, since it is hard to separate the disease-related functioning and the disease-related alterations of the quality of life perceived by the patient [Spey 13].

Structural: The normal swallowing process not only depends on the appropriate neurological functioning but also on the integrity of the anatomical structures involved in the deglutition. Structural alterations can produce dysphagia [Hira 18]. Structural dysphagia refers to strictly mechanical alterations despite an intact nervous system [Pfei 16]. In general, diverticula (e.g. Zenker's diverticulum), head and neck cancer (malignancy, esophageal tumors), muscular rings, inflammation-derived scarring, and fibrosis can cause structural oropharyngeal dysphagia [Chen 16a, Barr 14, Pfei 16]. The methods to evaluate structural lesions are the instrumental explorations based on X-Rays (i.e. VFSS) and the clinical exploration [Clav 15, Wang 17].

In this thesis, three swallowing dimensions were addressed: the electrophysiological by surface electromyography (sEMG), the mechanical by accelerometry-based cervical auscultation (Acc), and the acoustic by speech recordings. The following sections explain the characterization of such signals in the context of swallowing evaluation.

3.2. Characterization of electrophysiological and mechanical dimensions

3.2.1. Context of electrophysiological characterization in swallowing

The sEMG is intended to record muscular activity non-invasively. Even though there is a huge amount of studies oriented to the sEMG-related analysis in large muscles, specifically in upper and lower limbs, such studies are not necessarily comparable to those performed in swallowing-related muscles, because they differ in terms of size, signal-to-noise ratio (SNR), discharge frequency, amplitude, and temporal pattern [Rest 17, Klah 99]. Also, not all swallowing-related muscles can be assessed via sEMG because of their depth. Only some muscles of the face (e.g. *orbicularis oris*, buccinator), mastication (e.g. *temporalis* and masseter), suprahyoid (mylohyoid, stylohyoid, anterior belly of digastric), and infrahyoid regions (sternohyoid, superior and inferior bellies of omohyoid), allow the measure of their electrophysiological activity with surface electrodes [Step 12]. Other swallowing-related muscles are hindered by others as well as anatomical structures that impede their recording. These ideas motivate the research on the characterization of sEMG signals of swallowing-related muscles.

Feature extraction in sEMG could be affected by many intrinsic attributes such as the individual skin formation, blood flow velocity, skin temperature, tissue composition, and

the measuring site [Chow 13, Step 12]. For instance, small differences in fat content of the swallowing-related regions affect greatly the amplitude of the measured sEMG [Step 12]. Several features have been used for quantitative characterization, at the time, frequency, time-frequency, and non-linear dynamics domains [Phin 13, Phin 12a, Phin 12b, Engl 01]. Notwithstanding, not all these features could be extrapolated to swallowing tasks. For example, the maximal voluntary contraction is one of the most used indexes in sEMG analysis, including masseter and orbicularis oris signals [Sfor 11, Lope 17]; however, it is not practical for pharyngeal phase analysis.

In general, only amplitude-based estimations (root mean square, average rectified and filtered sEMG), duration, and frequency domain features (median frequency and spectral coherence) have been used in swallowing assessment [Vaim 04a, Step 12, Poor 17]. Other indexes usually analyze the abnormal timing (prolongation), the abnormal amplitude (attenuated), and the abnormal shape (lack of peaks) [Vaim 09]. Although some authors have found a significant correlation between timing and bolus volume and viscosity, other researchers did not [Palm 99, Poor 17, Watt 15]. More consistent results have been reported for the amplitude [Palm 99, Perl 99]. Although amplitude could be an inappropriate feature for inter and intra-subject analysis in the multichannel acquisition, it can be partially solved with normalization techniques [Step 12]. On the other hand, gender and age do not affect either the amplitude or the duration [Enge 12, Watt 15, Dell 18]. Otherwise, variation of these features is specific for the clinical conditions and reveals different patterns per disease [Cons 18, Grac 12, Vaim 08, Vaim 06, Crar 97, Enge 13, Arch 13, Erco 13, Hsu 13].

3.2.2. Context of mechanical characterization in swallowing

There is no consensus about what are the key features that represent the cervical auscultation signals [Dudi 15a]. Notwithstanding, statistical and frequency domain features have shown significant differences between A-P and S-I in healthy subjects [Lee 08]. These features depend on the age group and bolus characteristics, although viscosity has a larger influence than volume [Hann 10, Youm 11, Sejd 09]. The thickness of liquid boluses seems to affect specific features, and viscosity strongly influences the frequency and time-frequency domain features [Lee 10, Hann 10, Jest 14]. Entropy rate, Lempel-Ziv complexity (LZC), wavelet entropy, and frequency domain features have shown statistically significant differences between healthy and non-healthy thin swallows [Dudi 18b, Mova 17b]. In discrete swallows, entropy rate and LZC suffer variations in A-P and S-I axes, as well as the skewness (only in A-P). Otherwise, for sequential swallows, the entropy rate is sensitive in both axes but LZC is sensitive only in S-I axis, as well as the wavelet energy [Lee 10]. It is not clear yet if the aforementioned features have statistically significant differences by gender in healthy or non-healthy subjects since contradictory results have been found [Dudi 15c, Dudi 15b, Mova 17a, Dudi 18b, Jest 13].

Another meaningful feature is the swallowing duration [Youm 11]. It is affected by the bolus consistency [Hann 10, Sejd 09]: dry swallows last longer than wet swallows. The mentioned feature has shown dependency on liquid thickness in discrete and sequential swallows [Lee 10]. The duration measured by accelerometers has shown significant dependence on gender (possibly due to anatomical differences in the oropharyngeal mechanism), and age (explained by decoupling of oral and pharyngeal phases in elderly people) [Hann 10, Youm 11], but not on body mass index in healthy subjects [Sejd 09]. However,

it is difficult to make conclusions since some works have reported contrary results, i.e. the duration does not depend significantly on the sex and age [Dudi 15b].

The frequency of spontaneous swallows may be another feature for dysphagia evaluation. It has been used in patients with Parkinson's disease and acute stroke acquired with microphones, with variable sensitivity and specificity [Crar 13a, Gola 14]. This feature decreases in elderly individuals and patients with dysphagia [Crar 13b, Crar 13a].

Detection of penetration-aspiration (P/A) in patients with dysphagia has been approached with time and time-frequency domain characterization. Stationarity, normality, dispersion ratio, zero crossings, and energy features were used for aspiration detection in children with moderate results [Lee 06]. In contrast to simple statistical features, wavelet packet was applied on A-P and S-I axes successfully [Sejd 13, Dudi 16]; however, results could be overestimated since it is not clear if cervical auscultation is capable to detect the occurrence of aspiration by itself [Dudi 16].

3.2.3. Feature extraction in sEMG and Acc signals

Feature extraction has paramount importance to obtain better classification performance [Nazm 16]. Features in different domains described next were extracted from each sEMG channel and Acc axes by the sliding window method illustrated in Figure 3.1. This procedure retrieves one vector per feature and acquisition channel. Subsequently, six functionals were extracted: mean, standard deviation (SD), skewness, kurtosis, maximum and minimum. They were computed per feature vector to create a static and compressed representation of each feature. Window size for Acc and sEMG was experimentally set at 100 ms and 250 ms, respectively. Step size of 50 % in both cases. For missing data, imputation over the mean of each group (i.e. patients, controls) was performed.
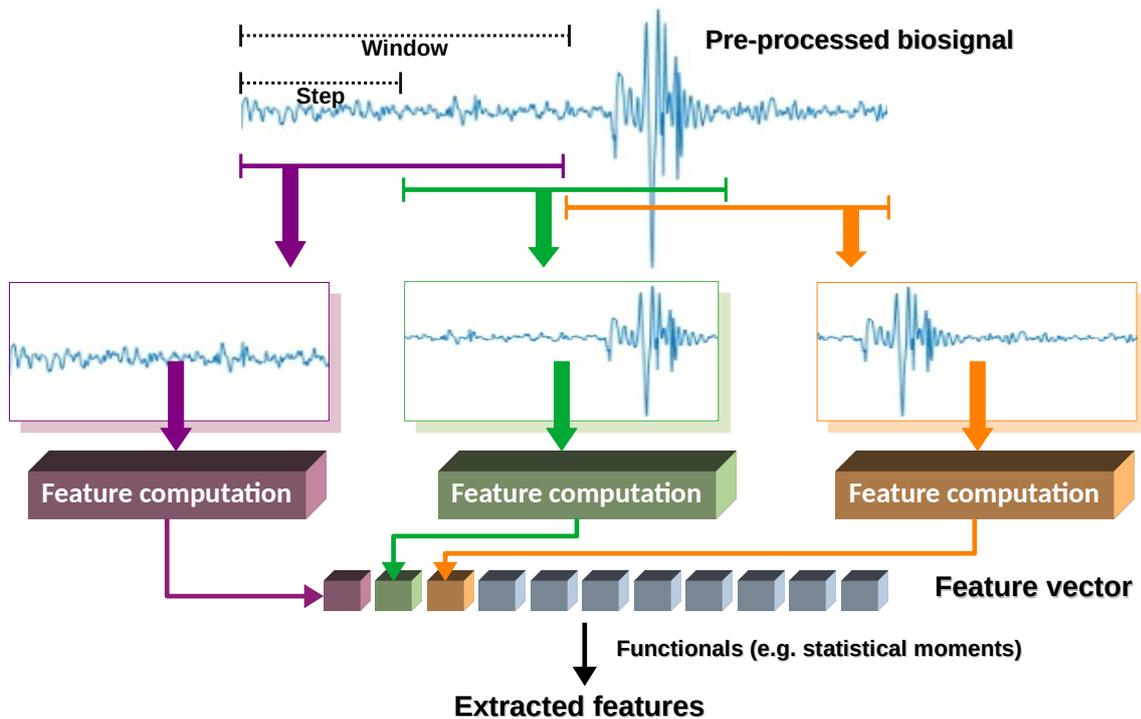
Figure 3.1: Sliding window method for feature extraction.

Time domain features

The implementation of time domain features is, in general terms, simple and does not demand high computational costs. An additional advantage is their direct clinical interpretability. However, they assume stationarity of the biosignal ignoring changes of statistical properties in time [Phin 12a]. The following time domain features were evaluated in this study: variance (VAR), root mean square (RMS), integrated EMG (iEMG), log-detector (LOG), waveform length (WL), the difference of absolute standard deviation (DASDV), Teager-Kaiser Energy Operator (TKEO), zero-crossing (ZC), Willison amplitude (WAMP), and myopulse (MYOP). The WAMP, ZC and MYOP are driven by the following thresholding function:

$$\phi(x) = \begin{cases} 1 & \text{if } x \geq \epsilon \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

where the threshold is given by $\epsilon = \text{mean} + h \times \text{std}$, with $h = 3$ [Soln 10], and mean and std are computed from the first 50 ms of each recording (without activation). The mathematical formulation of each feature is shown in Table 3.1.

Table 3.1: Mathematical formulations of time domain features (based on [Phin 12a])

Feature	Equation	Feature	Equation
VAR	$\dfrac{1}{N-1} \sum_{i=1}^{N} x_i^2$	RMS	$\sqrt{\dfrac{1}{N} \sum_{i=1}^{N} x_i^2}$
iEMG	$\sum_{i=1}^{N} |x_i|$	LOG	$\exp\left(\dfrac{1}{N} \sum_{i=1}^{N} \log(|x_i|)\right)$
WL	$\sum_{i=1}^{N-1} |x_{i+1} - x_i|$	DASDV	$\sqrt{\dfrac{1}{N-1} \sum_{i=1}^{N-1} (x_{i+1} - x_i)^2}$
WAMP	$\sum_{i=1}^{N-1} \phi(|x_i - x_{i+1}|)$	MYOP	$\dfrac{1}{N} \sum_{i=1}^{N} \phi(x_i)$
ZC	$\sum_{i=1}^{N-1} (x_i \times x_{i+1} < 0) \cap \phi(|x_i - x_{i+1}|)$	TKEO	$x_i^2 - x_{i+1} x_{i-1}$

Abbreviations: x_i: i-th sample of the sEMG signal; N: length of the sEMG signal; $\phi(\bullet)$: thresholding function.

Frequency domain features

These features typically describe properties of the power spectral density. The frequency domain features extracted in this study were: frequency ratio (FR), mean power

(MNP), mean frequency (MNF), median frequency (MDF), and peak frequency (PKF). The mathematical formulation of each feature is shown in Table 3.2. Low and high bands were considered to estimate FR. Since sEMG energy is mostly contained between 10 and 500 Hz [Merl 99], the low band was set between 10 and 250 Hz, whilst the high band was set between 250 and 500 Hz. These bands were also considered for Acc signals, but it was expected more spectral content below 100 Hz [Lee 09a].

Table 3.2: Mathematical formulations of frequency domain features (taken from [Phin 12a])

Feature	Equation	Feature	Equation
FR	$\displaystyle\sum_{j=LLC}^{ULC} P_j \Big/ \sum_{j=LHC}^{UHC} P_j$	MNP	$\displaystyle\frac{1}{M}\sum_{j=1}^{M} P_j$
MNF	$\displaystyle\sum_{j=1}^{M} f_j P_j \Big/ \sum_{j=1}^{M} P_j$	MDF	$\displaystyle\sum_{j=1}^{MDF} P_j = \frac{1}{2}\sum_{j=1}^{M} P_j$
PKF	$\displaystyle\operatorname*{argmax}_{f}\{P\}$		

Abbreviations: M: length of the power spectral density; P_j: power spectral density evaluated at the *j*-th frequency f_j; ULC and LLC: upper and lower cutoff frequency of the low frequency band, respectively; UHC and LHC: upper and lower cutoff frequency of the high-frequency band, respectively.

Time-frequency domain features

The time-frequency domain has been widely explored in sEMG signals. In fact swallowing-related electrophysiological events have been analyzed with continuous and discrete wavelets [Rest 17], in order to filter as well as to improve burst detection in the muscles assessed in the current work. Furthermore, wavelet-related energies and entropy have been successfully applied in kinematic signals for the automatic detection of swallowing-related events [Lee 10, Rebr 18].

The Discrete Wavelet Transform (DWT) is suitable to analyze non-stationary signals [Duan 16]. The DWT scales and shifts the signal $x[n]$ in discrete steps by the following dyadic expression in time and frequency axes [Zhan 10]:

$$\text{DWT}(j,k) = \frac{1}{\sqrt{2^j}} \sum_n \psi\left(2^{-j}n - k\right) x[n] \tag{3.2}$$

where $\psi(\bullet)$ is a function called *mother wavelet*; j, k and n are integers; j is the number of decomposition levels and n is the time counter; and the term $1/\sqrt{2^j}$ is an energy normalization factor [Zhan 10]. The DWT does not retrieve always an analytical solution, so numerical algorithms are required [Duan 16]. Mallat proposed a decomposition-based algorithm to define a complete and orthogonal multiresolution representation called "the

wavelet representation" [Mall 89]. Figure 3.2 illustrates such algorithm. The original discrete signal is decomposed in two downsampled signals, one obtained by the application of a low-pass decomposition filter $\tilde{H}$, namely the first approximation coefficient cA_1, and another obtained by the application of a high-pass decomposition filter $\tilde{G}$, namely the first detail coefficient cD_1. Afterward, cA_1 is decomposed again in the other two downsampled coefficients cA_2 and cD_2. Subsequently, each approximation coefficient is decomposed until the maximum decomposition level J is achieved.

(a) Decomposition at level j (b) Decomposition tree for J levels
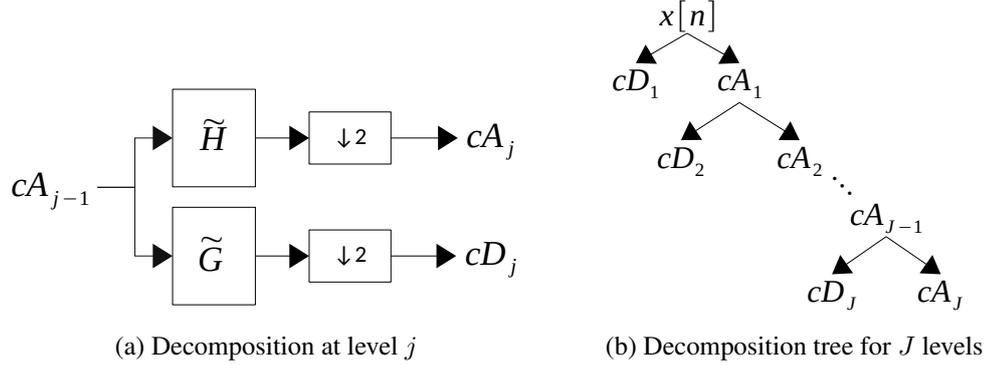
Figure 3.2: Wavelet decomposition. $\tilde{H}$ and $\tilde{G}$ denote the decomposition low-pass and high-pass filters, respectively; $\downarrow 2$ indicates a downsampling by 2; cA_j and cD_j are the j-th approximation and detail coefficients, respectively.

Relative energies of the decomposition levels from the discrete wavelet transform were computed. Each detail coefficient is a decimated, dilated, and translated version of the original signal. Five and nine decomposition levels were applied for sEMG and Acc signals, respectively. Thus, the energies of the decomposition coefficients were distributed in the following frequency ranges:

- sEMG: cD_1 in 500∼1000 Hz, cD_2 in 250∼500 Hz, cD_3 in 125∼250 Hz, cD_4 in 62.5∼125 Hz, cD_5 in 31.25∼62.5 Hz, and cA_5 in 0∼31.25 Hz.

- Acc: cD_1 in 1500∼3000 Hz, cD_2 in 750∼1500 Hz, cD_3 in 375∼750 Hz, cD_4 in 187.5∼375 Hz, cD_5 in 93.7∼187.5 Hz, cD_6 in 46.9∼93.7 Hz, cD_7 in 23.4∼46.9 Hz, cD_8 in 11.7∼23.4 Hz, cD_9 in 5.9∼11.7 Hz, and cA_9 in 0.1∼5.9 Hz.

The relative energy of each detail coefficient E_{Dj} was computed as follows [Lee 10]:

$$E_{Dj} = \frac{1}{E_T} \sum_{i=1}^{N_{cDj}} \{cD_j\}^2 \times 100\% \tag{3.3}$$

where N_{cDj} is the length of the j-th coefficient. This equation also applies to the energy of the J-th approximation coefficient E_{A_J}. The total wavelet energy E_T was estimated as follows:

$$E_T = \frac{1}{100} \left(E_{A_J} + \sum_{j=1}^{J} E_{Dj} \right) \tag{3.4}$$

Aiming to evaluate the information distribution in the time-frequency domain, the wavelet entropy W_{ent} was computed as follows:

$$W_{\mathrm{ent}} = -E_T \log_2 E_T \tag{3.5}$$

Nonlinear dynamics -NLD- related features

Nonlinear measures have been used in studies of accelerometry-based cervical auscultation signals in patients with dysphagia [Dudi 15a], as well as for sEMG-related application in large muscles [Phin 12b]. Extraction of NLD features requires representing the signal in an embedded space (also known as phase-space) with M reconstructed points, as defined by [Rose 93]:

$$\mathbf{X} = [\mathbf{X}_m(1) \ \mathbf{X}_m(2) \ \cdots \ \mathbf{X}_m(M)]^T \tag{3.6}$$

Such trajectories form figures known as diffeomorphic *attractors*, because they hold the topological properties of the dynamical system [Trav 17]. Each row of $\mathbf{X}$ is a phase-space vector; $\mathbf{X}_m(i)$ is the state of the system at the discrete time i and it is reconstructed as follows:

$$\mathbf{X}_m(i) = [x_i \ x_{i+\tau} \ \cdots \ x_{i+(m-1)\tau}] \tag{3.7}$$

Where τ and m are the embedding parameters, namely lag and embedding dimension, respectively. Thus, the trajectory matrix $\mathbf{X} \in \mathbb{R}^{M \times m}$. According to Taken's theorem, $\tau > 0$ and $m > 2d$, where d is the number of axes of the original system. The parameter m was found with the method proposed by [Cao 97], and τ was chosen as the distance where the auto-correlation function drops below $1 - 1/e$ times its maximal value [Rose 93]. For a time series with N points, the number of the reconstructed points M is related to the embedding parameters by the formula $M = N - (m-1)\tau$. Based on the aforementioned basis, the extracted NLD-related features are presented below.

Sample entropy - SampEn
This measure quantifies the rate of information production. It is widely used in the study of biological signals. It was introduced by [Rich 00], and it is intended to compute the complexity and regularity of time series. This is an improvement of the approximate entropy since it does not depend on the signal's length. It was computed as follows:

$$\mathrm{SampEn}(m, r) = \lim_{N \to \infty} \left[-ln \frac{A^m(r)}{B^m(r)} \right] \tag{3.8}$$

Where r is the tolerance for accepting matches, and $A^m(r)$ and $B^m(r)$ are the probabilities that two phase-space vectors will match for $m + 1$ and m points, respectively [Rich 00]. One match is met when the distance between two phase-space vectors is smaller than r.

Largest Lyapunov exponent - LLE
LLE measures how sensible is a dynamic system to changes of its initial conditions, that is, it allows to assess the divergence differences in the trajectories of the system that could

derive in the presence of chaos. The average divergence between two trajectories at time t is given by [Rose 93]:

$$d(t) = Ce^{\lambda_1 t} \tag{3.9}$$

where λ_1 is the LLE and C is a constant. For small data sets, the LLE is calculated from the slope of the following equation [Rose 93]:

$$\ln d_j(i) \approx \ln C_j + \lambda_1(i\Delta t) \tag{3.10}$$

where $d_j(i)$ is the distance from the j-th point of the trajectory to its nearest neighbor after i discrete-time steps, and Δt is the sampling period. Note that the equation 3.10 describes a family of lines, thus the LLE is computed using the least-squares method from the line [Rose 93]:

$$y(i) = \frac{1}{\Delta t} \langle \ln d_j(i) \rangle \tag{3.11}$$

where $\langle \rangle$ denotes the average slope over the family of lines.

Hurst exponent - HE

This is a measure of the "long-term" memory, i.e. long-term dependencies in a time series [Trav 17]. For a given time series with a range R and standard deviation σ, the following relationship is observed [Hurs 57]:

$$R/\sigma = (N/2)^K \tag{3.12}$$

where N is the number of samples of the time series, and K is the called Hurst exponent. For white noise (uncorrelated series), $K = 0.5$; for series with negative auto-correlation, $0 < K < 0.5$, while for series with positive auto-correlation $0.5 < K < 1$ [Trav 17].

Correlation dimension - CDim

The CDim is intended to describe the geometry of chaotic attractors, i.e. the CDim is a measure of the space dimensionality of such attractors [Trav 17]. For an M points trajectory and any positive number r, the correlation sum is defined as the fraction of pairs of points with distance smaller than r [Gras 04]:

$$C(r) = \frac{2}{M(M-1)} \sum_{i<j} \theta(r - ||\mathbf{X}_m(i) - \mathbf{X}_m(j)||) \tag{3.13}$$

where θ is the Heaviside step function. The dimension of the system is defined as follows:

$$D = \lim_{r \to 0} \frac{\log C(r)}{\log r} \tag{3.14}$$

If $C(r)$ decreases proportionally to r^D, then D is called the correlation dimension. The Grassberger-Procaccia Algorithm is intended to compute the CDim, by fitting a straight line into the log-log plot of $C(r)$ vs. r.

Detrended fluctuation analysis - DFA

The DFA, like the HE, measures long-term dependencies in the time series. Additionally, the DFA combines the benefits of the time and time-frequency domains, without the requirement of proper selection of a wavelet basis function [Phin 12b]. In this way, it has

been used for the characterization of sEMG signals in upper limbs, improving the classification accuracy of movements in combination with traditional feature spaces [Phin 12b]. The extraction process of DFA was made by a modified version of the random walk process [Phin 12b]: 1) the iEMG of the detrended signal was computed: 2) the resulting profile was windowed without overlapping; 3) each resulting frame was least-square fitted, thus a local trend appeared per frame; 4) the profile of each frame was detrended with the trend found in the previous step; 5) the RMS was computed of each detrended profile of each frame; 6) the process was repeated with different windows' lengths; 7) a log-log plot of the previously found RMS vs. window's length was computed, and the slope of such curve represented the scaling exponent [Aria 10].

Shannon Entropy

The Shannon entropy is commonly used as a measure of disorder in biological signals, e.g. in the characterization of kinematic events related to swallowing [Dudi 15b]. The Shannon entropy of a sequence X is computed as follows:

$$\text{Shannon}(X) = -\sum_{i=1}^{M} p(x_i) \log_2 p(x_i) \tag{3.15}$$

where $p(x_i)$ is the probability that $X = x_i$.

Lempel-Ziv Complexity (LZC)

The LZC indicates the presence of patterns in the data, by conversion of the raw data into a binary dictionary. After the computation of the complex envelope of the signal, the binarization process was done by thresholding. The count of different patterns in the binary signal retrieved the LZC [Scha 15]. Finally, the binary sequence is scanned from the left to the right; when a segment (symbol) that has not been seen before is found, it is stored in a dictionary. The LZC is the number of binary symbols in the dictionary [Trav 17].

3.3.   Characterization of the acoustic dimension

3.3.1.   Context of acoustic characterization in swallowing

Although there is a huge amount of papers in the field of computational paralinguistics, i.e. the study about *how* something is said[1] [Pir 15], few works have assessed voice-quality changes in patients with dysphagia from a quantitative point of view. These limited works have characterized voice changes with the following features: fundamental frequency (F0), the relative average perturbation (RAP, variability of pitch-to-pitch interval), shimmer (cycle-to-cycle amplitude variability), jitter (cycle-to-cycle frequency variability), the harmonic-to-noise ratio (HNR) and the voice turbulence index (VTI). Even though such features have been used also for pathological speech analysis, they are highly dependent on gender and acoustic environment [Meky 15]. Furthermore, contradictory results in dysphagic or aspirated patients have been found. Additionally, the lack of a consensus about whether speech evaluation is an appropriate strategy to evaluate swallowing impairments [Dos 21], is remarkable.

---

[1]Contrary to the Automatic Speech Recognition, which is concerned about *what* is said [Pir 15]

Whereas Ryu et al. found that RAP, shimmer, HNR and VTI increase significantly in patients with dysphagia and reported that these indexes have high sensitivity in pair-wise combinations [Ryu 04], Chang et al. found that these features do not identify the presence of P/A [Chan 12]. Kang et al. found that RAP, jitter, and HNR, as well as F0 (the latter for the manly population only), could be useful for the detection of aspiration risk before swallowing tasks [Kang 17], but Ko et al. did not find a correlation between the maximum phonation time extracted from the sustained vowel /a/, and the presence of P/A in patients with Parkinsonism [Ko 18]. Otherwise, Waito et al. showed that jitter, shimmer, and SNR have poor sensitivity [Wait 11], whilst Groves-Wright found that RAP, jitter, shimmer and absolute shimmer increase significantly when the material is observed on the vocal folds during phonation [Grov 07]. Furthermore, de Bruijin et al. found (in patients with oral and oropharyngeal cancer), that intensity of the vowels /a/, /i/, and /u/ is highly correlated to the amount of P/A and swallowing impairment, whilst F0 is significantly associated to swallowing inefficiency [Brui 13]. Likewise, Vogel et al. also found that F0 has statistically significant differences between patients with dysphagia and progressive ataxia, and controls (pitch control on vowels), as well as reduced speech rate in readings and automatic series (weekdays task), and increased pauses in monologues [Voge 17]. Otherwise, patients with aspiration risk exhibit reductions in the perturbation and noise parameters, apparently due to the increase in their vocal effort after swallowing, improving voice quality [Kang 17]. Additionally, Ramig et al. hypothesized that wet voice could be related to shimmer increase, but from a longitudinal case study in a single patient with ALS [Rami 90]. Otherwise, Murugappan et al. carried out an *in-vitro* experiment, which detected that wetness could increase the phonation threshold pressure, jitter, and shimmer, and reduce the amplitude of subglottal pressure [Muru 10].

Despite the aforementioned ideas, to the best of my knowledge, perceptual features have been characterized in only two papers by acoustic parameters in patients with dysphagia: [Zhao 22] extracted some of the features described previously to classify between healthy and dysphagic individuals (jitter, RAP, PPQ, F0, shimmer, among others), but their discrimination capability is not clear because statistical analyses were not provided, and authors applied feature selection without semantic. On the other hand, [Park 22] reported that APQ11-shimmer and RAP from the sustained vowel /e/ are potential digital biomarkers to detect the severity of dysphagia according to the oral feeding and risk of respiratory complications in post-stroke patients.

3.3.2. Feature extraction in speech signals

Different features have been studied in the field of automatic speech processing and understanding with several aims, for instance, to model abnormal patterns observed in the speech produced by patients suffering from Parkinson's disease [Oroz 18]. Since there is a relationship between several health conditions like Parkinson's disease and dysphagia, and considering that there is a lack of quantitative analyses on this topic, Parkinson-related features were extracted and evaluated to model abnormal patterns in the speech of patients with dysphagia. To address this task, the open-source software for Parkinson's speech analysis called Neurospeech [Oroz 18], was adapted. The modules related to phonation, articulation, and prosody were implemented in combination with the software Praat [Boer 01] which is commonly used to extract different phonation and articulation features. The cus-

tomization of this software for dysphagia-specific analysis was performed in Python 3. Details of the extracted features are presented below.

Phonation analysis in sustained vowels

Phonation is perhaps the most typical speech dimension in studies with dysphagic patients [Chan 12, Wait 11, Brui 13, Farn 17]. The main reason is that phonation changes are usually related to food or liquid residuals at the laryngeal vestibule [Muru 10], which is directly linked to abnormal vibration of vocal folds and abnormal control of the air that is generated in the lungs to produce speech. In spite of this, descriptions of how phonation is influenced by swallowing impairments are scarce [Yama 18]. In this study, phonation features were extracted from the sustained Spanish vowels: /a/, /e/, /i/, /o/, and /u/.

Fundamental frequency

Denoted as F_0, it is the frequency at the maximum correlation of the sound signal [Boer 93], i.e. the fundamental frequency related to the vibration of the vocal folds. The perceived fundamental frequency is associated with the concept of *pitch*. Without loss of generality, in this study, F_0 and pitch are equivalent variables. F_0 was extracted from each frame using Praat. Since Praat extracts 4 pitch values per frame of 40 ms, the mean of those values is computed to create the F_0 value for the given frame. Afterward, the mean, standard deviation, skewness, and kurtosis of the resulting feature vector from each recording were computed (see Figure 3.1). These four statistical functionals are also estimated for all of the features considered in this study.

Jitter

It measures short-term temporal perturbations of the voice signal. Details can be found in [Hadj 02]. Its computation is according to:

$$\text{Jitter}(\%) = \frac{100}{N \cdot \max(F_0)} \sum_{i=1}^{N} |F_0(i)] - \max(F_0)| \,, \tag{3.16}$$

where N is the number of frames in the utterance, $\max(F_0)$ is the maximum value of F_0 in all frames, i.e. $\max\{F_0(1), ...F_0(N)\}$, and $F_0(i)$ is the value of F_0 in the $i-$th frame.

Shimmer

It measures short-term amplitude perturbations in the voice signal and it is defined as follows [Hadj 02]:

$$\text{Shimmer}(\%) = \frac{100}{N \cdot \max(A)} \sum_{i=1}^{N} |A(i) - \max(A)| \,, \tag{3.17}$$

where N is the number of frames in the utterance, $\max(A)$ is the maximum amplitude value of the signal, and $A(i)$ is the amplitude of the $i-$th frame, i.e. the maximum value of such frame.

Amplitude and Pitch Perturbation Quotients – APQ and PPQ
These measures are used to estimate the long-term variability of the peak-to-peak amplitude and F_0 of the signal, respectively [Hadj 02].

$$PQ = \frac{1}{L} \sum_{i=1}^{L} \frac{\left| \frac{1}{k} \sum_{j=1}^{k} D(i+j-1) - D(i+m) \right|}{\left| \frac{1}{M} \sum_{n=1}^{M} M \cdot D(i) \right|}, \quad (3.18)$$

where $L = M - (k-1)$, $D(i)$ is the pitch period sequence (PPS) when computing the PPQ or the pitch amplitude sequence (PAS) when computing the APQ, M is the length of PPS or PAS, k is the length of the moving mean (tipically 11 for APQ and 5 for PPQ), and $m = (k-1)/2$. In this case, a window length of 150 ms was used with 50% overlapping [Oroz 18].

Energy
It is related to speech loudness. The energy content in logarithmic scale within one frame is computed as follows:

$$logE = \log\left(\sum_{j=1}^{N} |x_i(j)|^2 \right), \quad (3.19)$$

where $x_i(j)$ is the j-th sample of the i-th frame and N is the frame's length.

Articulation analysis in sustained vowels

The articulation relates to passive and active modifications of the sound energy in the vocal tract (supra-glottic cavities), composed of two kinds of signals, non-periodic (noise) and periodic (glottic sound) [Farn 17]. While the consonants have an oral source of noise and a laryngeal source of the sound, the vowels have a laryngeal source *articulated* by the tongue, palate, pharynx wall and lips [Farn 17]. The following features are extracted from recordings of sustained vowels with the aim to model articulation characteristics of the speaker:

Formants
The acoustic analysis of vowels is mainly based on the extraction of the first and second formants, F_1 and F_2 respectively, since these are the most relevant metrics for the production and perception evaluation of vowels [Nade 19]. They are generated by an energy reinforcement in the vocal tract [Farn 17]. In this case, formants are computed using the Praat software [Boer 01]. Furthermore, the first and second derivatives of each formant are also computed to include information about the dynamics of vowel production.

Teager-Kaiser Energy operator (TKEO)
It is a nonlinear operator widely used in the signal-processing field. For example, it has been used for burst detection in electromyographic recordings in swallowing analysis [Rest 17], and also to model different speech disorders [Oroz 15]. The TKEO is defined by:

$$TKEO\{x[n]\} = x[n]^2 - x[n-1] \cdot x[n+1], \quad (3.20)$$

where $x[n]$ is the n-th sample of the speech signal. The output of this operation describes the contour of the speech signal.

Mel-frequency cepstral coefficients (MFCC)

The vocal tract shape determines how the sound is produced. This shape is associated with the envelope of the short-time power spectral density of the utterance, which is associated with the MFCC introduced by [Davi 80]. MFCCs are also related to perceptual information in the human hearing range. A bank of triangular filters is built with central frequencies given by the Mel scale frequencies:

$$\text{Mel}(f) = 1127 \cdot \ln\left(1 + \frac{f}{700}\right), \tag{3.21}$$

Where f is the frequency in Hz. The filter bank is multiplied to the Fourier transform of each segment. The output of the filters is given by:

$$X_f(m) = \ln\left(\sum_{k=0}^{N-1} |X(k)H_m(k)|\right), \tag{3.22}$$

where m is a counter of filters, N is the length of each segment, k is a counter of frequency, $X(k)$ is the Fourier transform at the k-th frequency, and $H_m(k)$ is the frequency response of the m-th triangular filter. Subsequently, MFCCs are computed as follows [Sanc 18]:

$$\text{MFCC}[l] = \sum_{m=0}^{M-1} X_f(m) \cos\left(\frac{l\pi}{M}\left(m - \frac{1}{2}\right)\right); l = 1, ..., M, \tag{3.23}$$

where M is the number of filters. A total of 12 coefficients is extracted along with their first and second derivatives.

Vowel triangle analysis

The vowel triangle gives an idea of the capability of a speaker to hold the tongue in a certain position during the production of vowels. Particularly, it is based on the production of the so-called corner-vowels /a/, /i/, and /u/. The triangle is created on the F_1 vs. F_2 plane so each edge of it is given by the frequency value of the two formants. The vowel triangle is part of the features that evaluate the articulatory dimension of voice because it allows to evaluate vertical movements of the tongue (related to F_1) and also the tongue advancement (related to F_2) [Oroz 18, Nade 19]. In this study, three indexes from the vowel triangle were extracted:

Vowel Space Area (VSA)

It is a measure of the articulatory capability of a speaker [Oroz 18]. It has been used to analyze dysarthric speech, where it is expected to be compressed in comparison to healthy speech [Sapi 10]. The VSA is computed as follows:

$$VSA = \frac{|F_{1i}(F_{2a} - F_{2u}) + F_{1a}(F_{2u} - F_{2i}) + F_{1u}(F_{2i} - F_{2a})|}{2}, \tag{3.24}$$

where F_{ij} is the $i-$th formant associated with the vowel j.

Vowel Articulation Index (VAI)

It reflects the vowel centralization and it has been reported to be more sensitive to deficits in vowel articulation than the VSA [Rusz 13]. It is computed as shown in the following Equation:

$$VAI = \frac{F_{2i} + F_{1a}}{F_{1i} + F_{1u} + F_{2u} + F_{2a}}, \tag{3.25}$$

Formant Centralization Ratio (FCR)

It has been introduced by [Sapi 10] as an alternative to computing the VAI but minimizing the inter-speaker variability due to sex and age [Nade 19]. It is computed as follows:

$$FCR = \frac{F_{2u} + F_{2a} + F_{1i} + F_{1a}}{F_{2i} + F_{1a}}, \tag{3.26}$$

Articulation analysis in continuous speech

The level of articulation impairment has been assessed in continuous speech, particularly in patients with Parkinson's disease [Oroz 16a]. In this case, the Bark frequency scale has been used to model the transition between starting and stopping the vocal fold vibration in continuous speech recordings. Hence, two cases are possible in those transitions: onset, which describes the transition between unvoiced to voiced segments; and offset, which describes the transition from voiced to unvoiced segments. Voiced and unvoiced segments were found with the software Praat according to the presence or absence of F_0 values, respectively. Typically, 25 scales are computed from the speech signals as follows [Zwic 80]:

$$Bark(f) = 13 \arctan\left(0.76\frac{f}{\text{kHz}}\right) + 3.5 \arctan\left(\frac{f}{7.5\text{kHz}}\right)^2, \tag{3.27}$$

where f is the frequency in kHz. Considering that the channel normalization process down-samples the signal to 8 kHz, only 18 Bark bands can be computed per transition (onwards BBE_on and BBE_off, respectively), and the mean and standard deviation are obtained. The articulation analysis in continuous speech retrieves a feature vector per recording with 18 Bark-band energies, and 2 statistical moments in the voiced and unvoiced segments ($18 \times 2 \times 2 = 72$ features).

Diadochokinetic analysis

One of the most common methods to evaluate articulatory skills is based on the diadochokinetic (DDK) analysis, which consists of a rapid repetition of the pairs of plosive consonant-vowel. This task allows evaluating the movements of articulators such as lips, tongue, and velum [Rusz 11, Oroz 16a]. [Fest 16] reported that DDK has the ability to predict moderate or severe dysphagia. The most common DDK tasks involve the syllables /pa/-/ta/-/ka/, and combinations of them like /pa-ta-ka/, /pa-ka-ta/, or /pe-ta-ka/, [Oroz 18].

In this work, the participants were asked to produce rapid repetitions of /pa-ta-ka/. The DDK-related features extracted from recordings were [Oroz 18]:

Pitch (F_0): Details are presented in section 3.3.2.

Energy: Details are presented in Section 3.3.2.

DDK (syllable vocalizations): Syllable rate (syllable vocalizations per second), mean and variance of the syllable duration were computed.

Pauses related segments: Similarly to the previous feature, pauses rate, together with the mean and variance of the pause duration were computed.

The syllable vocalizations and pauses were extracted automatically with Praat.

Prosody analysis

This speech dimension is related to timing, intonation, speech rate, and pauses produced while naturally speaking [Oroz 18, Vasq 18]. Despite the high number of studies related to phonation and articulation analysis in pathological speech, the acoustic features of prosody have been scarcely analyzed [Ramo 20]. In this work, the following prosody features were considered: pitch (F_0), logarithmic energy, and duration of voiced and unvoiced segments. Maximum values along with the four functionals are estimated per feature. The voiced and unvoiced segments were extracted automatically with Praat. Table 3.3 summarizes the features as well as the functionals extracted from speech recordings.

3.4. Feature selection strategies

The sliding window method was used to return one vector per feature. Thus, a feature space $\mathbf{X}' \in \mathbb{R}^{m \times n}$ with m samples (individuals) and n features was obtained. Z-score normalization was applied to standardize the feature space, i.e.:

$$\mathbf{X} = \frac{\mathbf{X}' - \mu'_{\mathbf{X}}}{\sigma_{\mathbf{X}}}, \tag{3.28}$$

where $\mu'_{\mathbf{X}}$ and $\sigma_{\mathbf{X}}$ are the mean and standard deviation vectors of the analyzed population, respectively, computed on the columns of the matrix $\mathbf{X}'$. Thus, $\mathbf{X} = \left(x_j^{(i)} \right) \in \mathbb{R}^{m \times n}, i = 1, \ldots, n,$ and $j = 1 : m$, where $x_i^{(j)}$ denotes the normalized *j*-th feature of the *i*-th individual.

Different feature selection schemes were applied to prevent overfitting and to provide model interpretation, optimization, sparsity, and data understanding [Guyo 03]. There are three main approaches to select features [Bolo 13]:

- Filter methods: selection of features based on their intrinsic properties, individual or jointly, such as statistical independence, mutual information, correlation, or inherent power of discrimination. These methods are independent of the classification algorithm.

- Embedded methods: the feature selection depends on the classification algorithm in the training process.

Table 3.3: Summary of features and functionals evaluated by speech dimension.

Dimension	Features	Functionals
Phonation	APQ PPQ Jitter Shimmer F_0 (Δ^0, Δ^1, Δ^2) logE	(One feature per vowel) MV, SD, SK, Kurt
Articulation	*Sustained vowels* F_1 (Δ^0, Δ^1, Δ^2) F_2 (Δ^0, Δ^1, Δ^2) MFCC[$1 \sim 13$] (Δ^0, Δ^1, Δ^2) TKEO *Vowel triangle* VSA, VAI, FCR *Continuous speech* BBE_on[$1 \sim 18$] BBE_off[$1 \sim 18$]	(One feature per vowel) MV, SD, SK, Kurt MV, SD
Diadochokinesia	F_0 logE DDK & Pauses	SD Max, MV, SD Rate, MV, SD
Prosody	F_0 & logE Voiced & Unvoiced segments	Max, MV, SD, SK, Kurt Rate, MV, SD, SK, Kurt

Abbreviations: MV: mean value; SD: standard deviation; SK: skewness; and Kurt: kurtosis.

- Wrapper methods: it also depends on the classification algorithm, but the selection is performed by an optimization process (greedy search).

Only filter methods were considered in this thesis because they have lower computational cost and good generalization capability since they do not depend on the classifier [Bolo 13]. The latter is the most desired property for the application under study in this work since it seeks possible biomarkers of dysphagia, which should not depend on specific algorithms, in order to prevent the risk of overfitting, as well as lack of generalization or physiological interpretability. Next, the filter methods applied in this work are explained.

Minimal Redundancy - Maximal Relevance (mRMR)   The mRMR is one of the most robust methods for multivariate filter of features, and it has been also used successfully for channels selection in sEMG [Mesa 14]. This algorithm is based on the mutual information criterion. For two variables x and y, with marginal probabilities $p(x)$ and $p(y)$, respectively, and joint probabilistic function $p(x, y)$, the mutual information I is computed as follows [Ding 05]:

$$I(x, y) = \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \tag{3.29}$$

The mRMR method seeks to find a set of features S that are maximally dissimilar between them, as well as to maximize the relevance of these features regarding the classes $h = \{\text{healthy}, \text{dysphagic}\}$. This is summarized in the next optimization conditions [Ding 05]:

$$\min \frac{1}{|S|^2} \sum_{i,j \in S} I(i,j) \,, \; \max \frac{1}{|S|} \sum_{i \in S} I(h,i) \tag{3.30}$$

where $|S|$ denotes the number of features in S.

Principal components analysis (PCA) based selection   PCA is, in principle, a powerful technique to reduce dimensionality. It is based on an orthogonal transformation of the original feature space into another with the highest contribution in variance. Given the normalized feature matrix $\mathbf{X}$, the covariance matrix is found as $\mathbf{\Sigma} = \frac{1}{m}\mathbf{X^T X}$, $\mathbf{\Sigma} \in \mathbb{R}^{n \times n}$ [Daza 09]. The Singular Value Decomposition method (SVD) retrieves the eigenvectors $\mathbf{U}$ and eigenvalues $\mathbf{\Lambda}$ from $\mathbf{\Sigma}$. The trace of $\mathbf{\Lambda}$ returns the percentage of explained variance. The eigenvalues which sum an explained variance given by a threshold (e.g. 90%) are retained and their corresponding eigenvectors form a reduced matrix $\mathbf{V'}$. Then the uncorrelated principal components (PC) are found by $\mathbf{PC} = \mathbf{XV'}$. However, PCs are usually hard to interpret, that is, they have no semantics. In this way, we used the projection of $\mathbf{X}$ onto the PCs, i.e. $\text{proj}_{\mathbf{PC}}\mathbf{X}$. In this projection, the weight of each original feature is computed per PC. In this way, only those features with the highest weights on PCs that retain a certain amount of variance, are selected, without losing semantic and interpretability. A threshold of 0.8 of the explained variance was set *ad-hoc* to select the features for the classification stage.

Hypothesis tests   Hypothesis tests are used to make inferences about populations from samples [Maru 10]. In this case, such tests were used to determine the inherent capability of discrimination of the features extracted from each biosignal. The null hypothesis H_0 is that *there is no statistically significant difference between the values of the feature* $\mathbf{X_i}$ *extracted from healthy controls and patients with dysphagia*. The goal is to reject hypothesis; in this case, the feature $\mathbf{X_i}$ will be taken as a possible biomarker. The p-value of the statistical test is the probability to obtain a random difference which is higher than the observed one. A typical criterion says that if $p < 0.05$ one could reject H_0 [Moli 17]. Since the collected data come from two unpaired samples, the Mann–Whitney U test is the most recommended test to find features with significantly different values (the median) between groups [Maru 10].

Due to the recent criticism regarding the use of the p-value for statistical inference in diagnostic tests [Amrh 19], this kind of comparison was performed only in Experiment #3 (see Chapter 2). To compensate the problem of multiple comparisons, the Bonferroni correction was applied, with a family-wise error rate of 0.01. It also helps to prevent type-I errors (reject H_0 given that it is true), leading to more conservative conclusions.

Additionally, since hypothesis tests are highly dependent on the sample size [Tomc 14], the eta-squared (η^2) was computed in order to analyze the effect size as follows:

$$\eta^2 = \frac{Z^2}{N}, \tag{3.31}$$

where Z is the standardized value of the Mann-Whitney U, and N is the number of samples. If $\eta^2 = 0.01$, there is a small effect size, $\eta^2 = 0.09$ means medium effect size, and $\eta^2 = 0.25$ means large effect size [Frit 12].

Area under the ROC curve (AUC)   The analysis of the Receiver Operating Characteristics curve (ROC) is paramount in the Machine Learning community [Powe 11]. To deep inside its definition, the following concepts must be clarified:

- True positives (TP): number of patients with dysphagia that the algorithm classifies as with dysphagia

- True negatives (TN): number of healthy individuals that the algorithm classifies as healthy

- False positives (FP): number of healthy individuals that the algorithm classifies as with dysphagia

- False negatives (FN): number of patients with dysphagia that the algorithm classifies as healthy

The area under the ROC curve (AUC_{ROC}) is a univariate feature selection method used to measure the equilibrium between the true positive rate (TPR) and the false positive rate (FPR). For ordinal-like data of the features extracted in this thesis, one probability distribution is obtained per group, i.e. per class (see Figure 3.3).
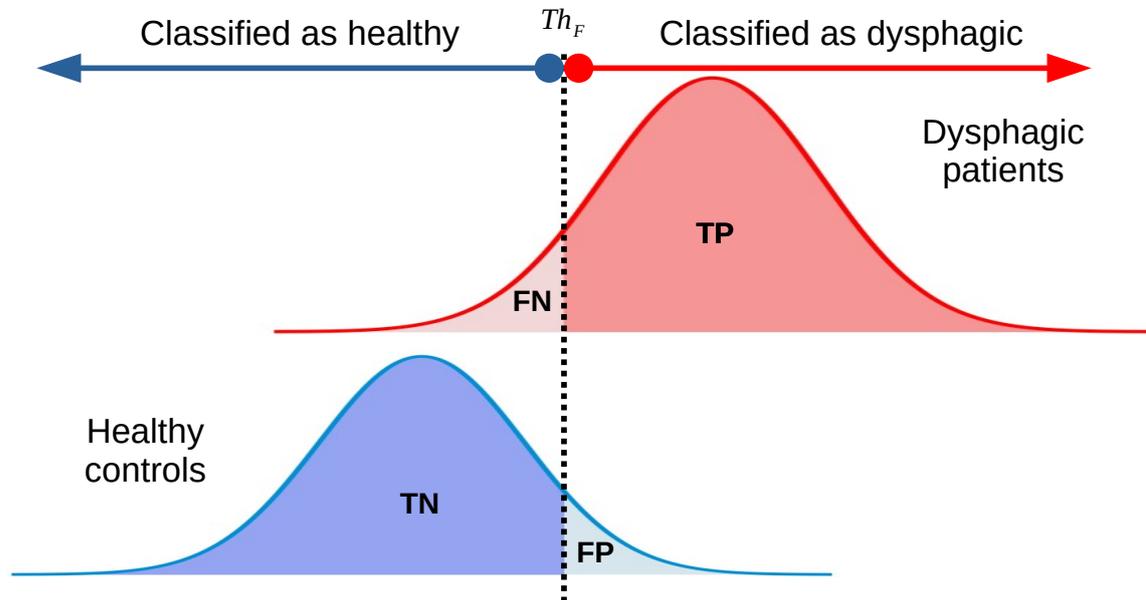
Figure 3.3: Illustration of the distribution values of a specific feature in healthy and dysphagic individuals. Th_F: feature threshold; TP: true positives; FN: false negatives; TN: true negatives; FP: false positives.

A moving threshold Th_F is set for each feature $\mathbf{X_i}$ in the range $[\min(\mathbf{X_i}), \max(\mathbf{X_i})]$. Thus, as Th_F increases, the number of TN, FN, FP, and TP changes, i.e. the number

of the correctly classified individuals changes. In Figure 3.3, low values of Th_F tend to classify all individuals as dysphagic, but high values of Th_F tend to classify all individuals as healthy. Each value of Th_F produces a point in the ROC (TPR vs. FPR):

$$(\text{FPR}, \text{TPR})_{\text{ROC}} \rightarrow \left(1 - \frac{\text{TN}}{\text{TN} + \text{FP}}, \frac{\text{TP}}{\text{TP} + \text{FN}} \right) \tag{3.32}$$

In the context of binary classification, an AUC score *between 0.7 and 0.8 is acceptable, between 0.8 and 0.9 is excellent, and more than 0.9 is outstanding* [Mand 10]. Thus, a selection threshold of 0.7 was defined, i.e., features with $\text{AUC}_{\text{ROC}} \geq 0.7$ were selected for classification. As a limitation, in contrast to the mRMR and PCA-based method, the AUC_{ROC} does not retrieve any information about correlations between features of the input space. This method was used in order to determine the individual discrimination capability of each feature.

Chapter 4

Feature analysis and proposal of swallowing biomarkers

In this chapter, the results of the signal characterization are presented per experiment. Several swallowing biomarkers are also proposed.

Since the experiments of this thesis were conducted in different stages of the research and with different databases[1], different combinations of feature selection methods were performed for each experiment. Table 4.1 summarizes the feature selection methods used for each experiment.

Table 4.1: Summary of feature selection methods used for each experiment.

Experiment	PCA-based	mRMR	H_0	AUC_{ROC}
Experiment #1	●	●		●
Experiment #2		●		
Experiment #3			●	
Experiment #4				●

4.1.   Experiment #1: Electrophysiological biomarkers

This section and the number 6.1 were excerpted with some modifications from the following journal paper: Roldan-Vasco, Sebastian, Andres Orozco-Duque, and Juan Rafael Orozco-Arroyave. "Swallowing disorders analysis using surface EMG biomarkers and classification models." Digital Signal Processing 133 (2023): 103815. This article can be found on the publisher's website at https://www.sciencedirect.com/science/article/abs/pii/S1051200422004328. The journal's homepage is located at https://www.sciencedirect.com/journal/digital-signal-processing, and the publisher's copyright information can be found at https://www.elsevier.com/

---

[1]Volunteers recruitment, including healthy and dysphagic individuals, lasted approximately five years, with the COVID-19 pandemic in between.

`about/policies/copyright/permissions`. In this experiment, the evaluation of sEMG-related features as biomarkers was performed

4.1.1. Contribution of individual features

The feature space used as input to the classifiers that will be detailed in Section 6.1, was formed with four functionals per feature (mean, standard deviation, maximum and minimum), which were extracted from every channel (two bilateral channels are considered per muscle group). Therefore, the representation space was formed as follows: 4 functionals $\times$ 6 channels = 24 features. Figure 4.1 shows a bubble-matrix chart to visualize the number of functionals retrieved by each feature (rows) extracted from each muscle group (columns). The bubble size indicates the number of swallowing tasks in which some functional of a specific feature was retrieved by a specific feature selection method. There are eight different sizes of the bubbles, with the smallest indicating that the feature was retrieved in just one swallowing exercise and the largest indicating that it was retrieved in all tasks. Otherwise, the quantity of functionals retrieved across all swallowing activities is shown by a color scale; the more functionals engaged, the darker the color. This graphic offers a broad overview of feature selection patterns across feature domains. Neither in terms of feature domain nor in terms of muscle groups, the PCA-based selection did not show any discernible pattern.

The majority of the features in all sEMG channels were retrieved even if the selection criterion was set at 80% of variance, regardless of the swallowing task. No clear pattern was found either with mRMR. In contrast to PCA, mRMR selected a sparse set in terms of features and muscle groups (see Figure 4.1). Adding up to the inconvenience of using these feature selection methods, mRMR does not allow to analyze the relevance of individual features in terms of discrimination and PCA is based on the maximization of the representation space variance, which limits specific analyses with individual features.

In contrast, a distinct pattern in terms of muscle groups and feature domains was recovered by the AUC_{ROC} based selection (see Figure 4.1). Features in the time domain extracted from supra- and infrahyoid muscles were more frequent than other domains and muscles (see Figure 4.1). In this way, VAR, RMS, iEMG, LOG, WL, and DASDV from bilateral suprahyoid muscles, WAMP and MYOP from the left suprahyoid, and the frequency domain feature MNP from bilateral suprahyoid, are clearly highlighted in terms of AUC for most of the swallowing tasks. Time domain features are also present in most of the swallowing tasks (water and yogurt) in the infrahyoid muscles. Figure 4.2 shows colormaps with information on the number of functionals selected by the AUC_{ROC} method, distributed along the muscle groups and swallowing tasks.

4.1.2. Proposal of electrophysiological biomarkers

In this experiment, different feature sets were evaluated, which had not been explored previously in the state-of-the-art of swallowing based on sEMG signals.

Despite PCA and mRMR selection methods retrieved good classification performances (shown in Section 6.1), they were not appropriate to find biomarkers that enable possible clinical interpretation. Even though the AUC_{ROC} based selection method does not allow to analyze how features and muscles interact or correlate among them, this selection criterion
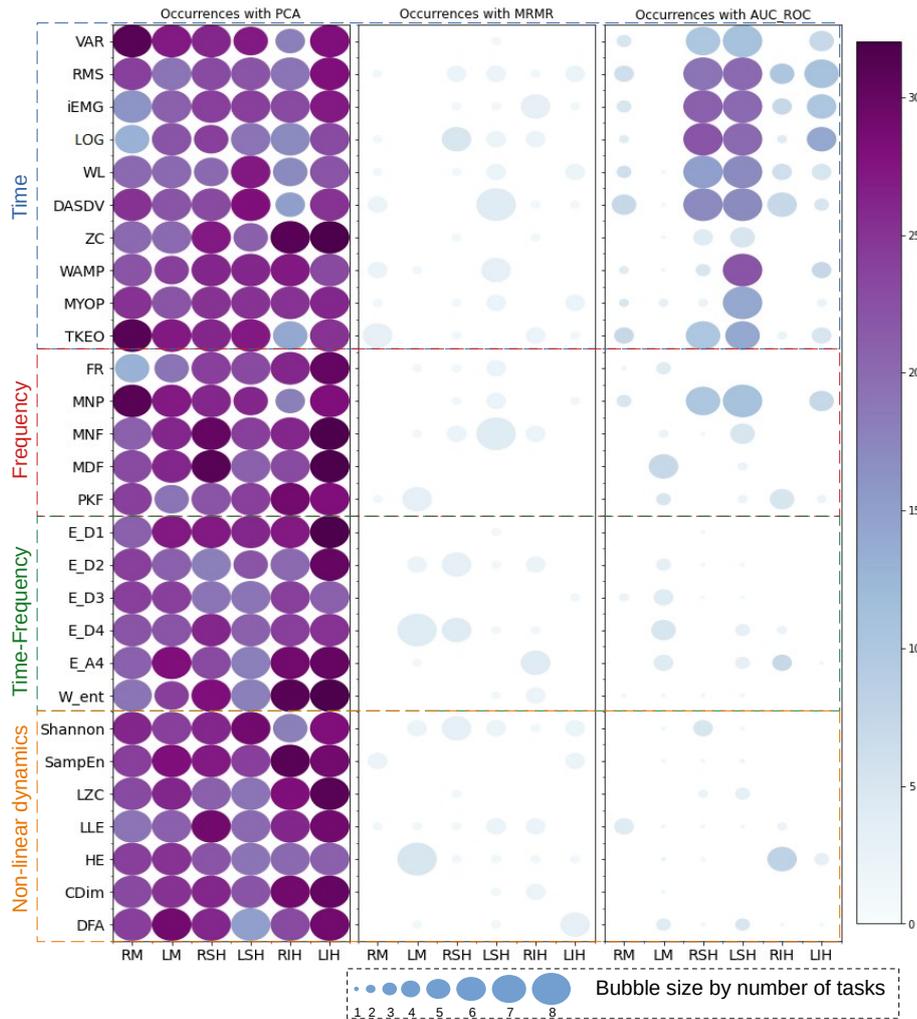
Figure 4.1: Bubble-matrix chart with the features obtained by each selection method. The bubble size shows the occurrence of features along the different swallowing tasks (rows) in different muscle groups (columns). The colormap indicates the number of functionals per feature.

allowed to find the features that yielded higher discrimination results, so this method was used for biomarkers selection.

It is noticeable that the time-domain features extracted from suprahyoid muscles were the most prevalent in all swallowing tasks. Figure 4.2 also illustrates that muscle activation is volume dependent: masseter-related features show good discrimination capability in water$_5$ and yogurt$_5$, while the intake of greater volumes does not show good classification performance. An effortful swallow was observed during the intake of such small volumes, especially for yogurt, which supports the fact that time domain features –most of them measures of contraction force–, retrieved high classification accuracies. Such effort is characterized by a suction-like praxis with slight masseteric contraction in many patients with dysphagia. This behavior could be related to the piecemeal deglutition produced by 20 mL in neurogenic patients, known as *dysphagia limit* [Aydo 15]. This double or even triple activation attenuates the effect of one struggling swallowing in terms of amplitude
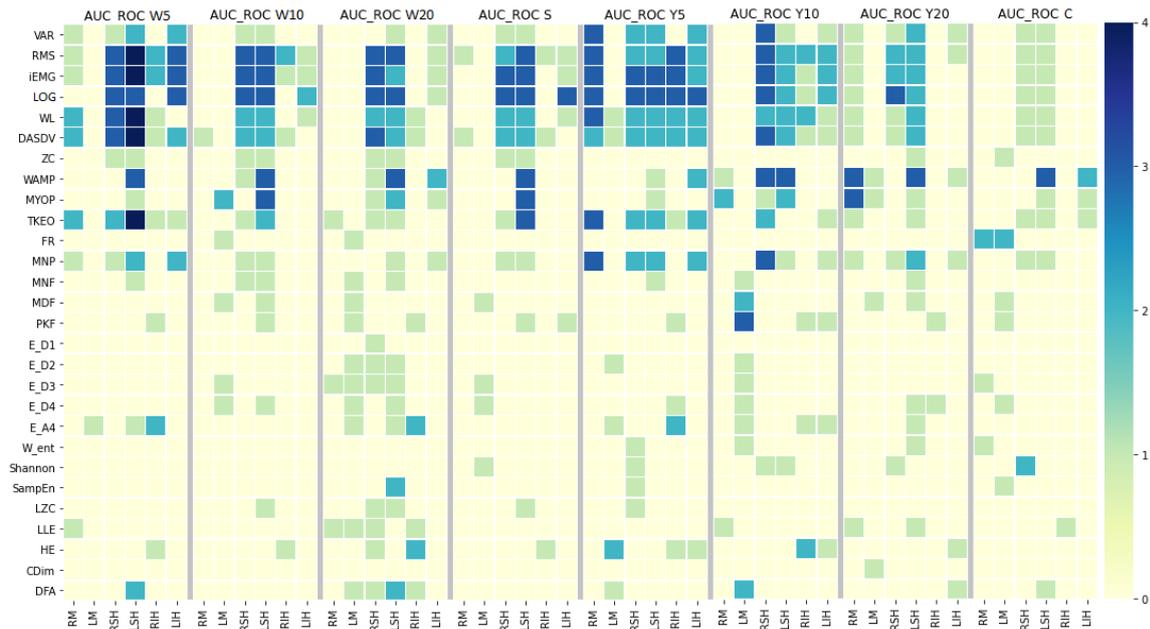
Figure 4.2: Colormaps of features selected by the AUC$_{ROC}$ method distributed along muscle groups and swallowing tasks. Wx: x mL of water; Yx: x mL of yogurt; S: saliva; C: 3 g of cracker.

or duration which are among the underlying aspects that are modeled with time-domain features.

Although other feature domains did not show consistent results for dysphagia discrimination, the MNP from suprahyoid muscles yielded AUC≥0.7 in most of the swallowing tasks (Figure 4.1); this result agrees with a paper developed in an early stage of this work [Rold 18], in which MNP was the most accurate frequency-domain feature for the classification of swallowing phases. NLD and time-frequency features did not exhibit individual capability of discrimination. Even though wavelets have been used for burst detection during swallowing tasks [Rest 17], the wavelet performance is highly dependent on the proper selection of the mother wavelet. The parameters used in this experiment were recommended for sEMG-based classification [Chow 13], but such recommendations were made on typical applications of sEMG, i.e. upper and lower limbs-related myoelectric control, whose muscle fibers have different activation patterns and discharge rate in comparison with the swallowing related muscles evaluated in this work. Regarding the NLD features, it is difficult to hypothesize a physiological reason behind the low performance, but it could be related to the absence of well-defined patterns of the phase spaces: no visual differences were detected in the graphics of the attractors of healthy individuals, and patients with dysphagia.

On the other hand, this experiment strengthens the findings of a previous one [Rold 18], and the following features are proposed as potential electrophysiological biomarkers for swallowing studies based on sEMG signals: VAR, RMS, iEMG, LOG, WL, DASDV, WAMP, MYOP, TKEO, and MNP. Despite MYOP and WAMP were not accurate for automatic detection of the swallowing phases in [Rold 18], in this experiment the threshold was modified (functionals of the first 50 ms instead of the whole recording), and the results improved.

4.2. Experiment #2: Electrophysiological and mechanical biomarkers

This section and the number 6.2 were excerpted with some modifications from the following journal paper: Roldan-Vasco, Sebastian, et al. "Analysis of electrophysiological and mechanical dimensions of swallowing by non-invasive biosignals." Biomedical Signal Processing and Control 82 (2023): 104533. This article can be found on the publisher's website at https://www.sciencedirect.com/science/article/pii/S1746809422009879?dgcid=author. The journal's homepage is located at https://www.sciencedirect.com/journal/biomedical-signal-processing-and-control, and the publisher's copyright information can be found at https://www.elsevier.com/about/policies/copyright/permissions. In this experiment, the evaluation of the combination of sEMG and Acc-related features was performed.

4.2.1. Electrophysiological and mechanical biomarkers

Table 4.2 shows the contribution of each feature to the classification results obtained with the mRMR selection method. Only the best results are shown. The last column illustrates the relative contribution of each feature. Notice that MYOP achieved the highest contribution, although for this feature the algorithm only selected measures extracted from the sEMG channels. This behavior was to some extent predictable because this feature is mainly intended to describe sEMG signals [Phin 12a]. Notice that MYOP was present for all sEMG channels, but for the specific case of the saliva intake, that feature was not required to achieve the best classification result.

When analyzing the results regarding domains, it can be observed that in time, there were other three features that achieved high occurrence: RMS, INT, and LOG. However, they were not as consistent as MYOP. For the frequency domain, these features were not systematically selected by the algorithm, along swallowing tasks. They show to be strongly related to the Acc axes. The MDF is shown neither in Table 4.2 nor in Figure 4.3, because is was not selected for any swallowing task. Regarding time-domain features, there were a total of 44 occurrences in which one or more measures were selected along the swallowing tasks and recording channels. Notice that 16 out of those 44 occurrences (36%) correspond to measures of the AP axis and 10 (23%) correspond to SI, both associated to Acc biosignals. Furthermore, W_{ent} also retrieved a high relative contribution, but it was not selected by the algorithm for the classification during the intake of water$_5$ and yogurt$_{10}$. Its participation to characterize information from different channels is not consistent. Finally, for the NLD features, it can be observed that SampEn and LLE, which were computed for the Acc axes only due to the results obtained in Experiment #1, the relative contribution was high. SampEn was retained for at least one Acc axis in all swallowing tasks, and LLE was retained in the AP axis for all water tasks and yogurt$_{10}$.

Figure 4.3 shows heat maps for the AUC values computed for the features selected according to Table 4.2. Since each feature provides six functionals per acquisition channel and task, only the maximum AUC is shown for displaying purposes. Notice that most of the AUC values achieved by the time, frequency and NLD domains vary between 70% and 80%, indicating that the selected features have good or acceptable capability of dis-

Table 4.2: Features selected per channel and swallowing task. Only the features that achieved the highest classification performance are shown. The bullet size represents the number of functionals retrieved; the last row shows the contribution of each channel per task; and the last column shows the relative contribution of each feature (occurrences).

Feature		Water - 5 mL	Water - 10 mL	Water - 20mL	Saliva	Yogurt - 5 mL	Yogurt - 10 mL	Yogurt - 20 mL	Relative contribution
Time	VAR								0.2
	RMS								0.7
	INT								0.9
	LOG								0.6
	WL								0.4
	DASDV								0.3
	WAMP								0.1
	MYOP								0.4
	ZC								1.0
	TKEO								0.4
Frequency	FR								0.7
	MNP								0.4
	MNF								0.5
	PKF								0.6
Time-frequency	E_a								0.5
	E_{D1}								0.2
	E_{D2}								0.6
	E_{D3}								0.4
	E_{D4}								0.6
	E_{D5}								0.4
	E_{D6} *								0.7
	E_{D7} *								0.1
	E_{D8} *								0.3
	W_{ent}								0.8
NLD	SampEn *								0.8
	LLE *								0.7
	LZC *								0.3
	Count								

Bullet color notation: SI (•), ML (•), AP (•), RSH (•), LSH (•), and RIH (•)

* Features computed for Acc only

Figure 4.3: Heat maps of the AUC for the features selected by the mRMR method, by acquisition channel and swallowing task.

crimination [Mand 10]. The AP axis yielded the highest AUC regardless of the swallowing task: 0.81 in water$_5$, water$_{10}$, and yogurt$_{20}$; 0.82 in water$_{20}$; 0.79 in saliva; and 0.77 in yogurt$_5$ and yogurt$_{10}$. The LSH channel also achieved AUC>70% for most of the features and intakes. However, its AUC decreases for yogurt (5 mL and 10 mL). In contrast, the other two sEMG-related channels did not achieve high AUC values in most of the cases. The LLE calculated for water$_{20}$ achieved the highest AUC along the experiments (82%). Actually, NLD features systematically achieved AUC>70% in all swallowing tasks. There were only exceptions, SampEn in yogurt$_{20}$ and LLE in water$_5$. Other features such as RMS, INT, MNF, LOG, FR, and W_{ent} also show values higher than 80% for different swallowing tasks.

4.2.2. Information per sEMG channel and Acc axis

The last row of Table 4.2 illustrates the number of features selected from each acquisition channel and axis. The Acc-related axes contribute consistently with most of the features in all tasks: in water$_5$, water$_{10}$, and water$_{20}$, with 70.5%, 76.2%, and 64.7%, respectively; in saliva, the contribution was 70%; and for yogurt$_5$, yogurt$_{10}$, and yogurt$_{20}$, the contribution of the Acc axes was 76%, 57.1%, and 63.6%, respectively. Additionally, the AP axis provided the highest contribution in terms of the number of features, in all swallowing tasks, with the exception of yogurt$_5$ in which the SI axis retrieved the highest number (36.0%). In contrast, the RIH channel retrieved the smallest contribution (< 12% in all swallowing tasks). Otherwise, the LSH muscle group retrieved the most of the sEMG-related features in almost every task: 60% in water$_5$, yogurt$_{10}$, yogurt$_{20}$, and saliva, and 50% in water$_{20}$.

Regarding the swallowing tasks, there is not a clear pattern related to the number of features necessary to discriminate between healthy and dysphagic individuals: whilst yogurt$_{10}$ requires 12 features, yogurt$_5$ requires 23. In this way, neither the consistency nor the volume seems to be correlated to the number of features required. Otherwise, water$_{10}$, water$_{20}$, saliva, and yogurt$_{10}$ were characterized by the elimination of one channel or axis: RIH, RSH, RIH, and ML, respectively.

4.2.3. Regularity of electrophysiological biomarkers and proposal of mechanical ones

The selection of a feature in one specific Acc axis does not mean that it will also be selected for the other axes because they have different movement range within healthy and dysphagic individuals. The mRMR method also discarded similarities in features selected in two or three acquisition channels. For instance, PKF has shown statistical dissimilarities between AP and SI axes [Lee 08], and VAR, LZC and W$_{ent}$ have exhibited tri-axial directional differences with variable consistencies [Mova 17a].

It also explains that the ML axis did not contribute to the best classification results in yogurt$_{10}$; although the ML movement is hard to detect visually because it is quite subtle, it may provide information about symmetry alterations of hyolaryngeal muscle contractions [Mova 17a]. Furthermore, one study reported statistically significant differences between ML-related features extracted from healthy and dysphagic populations [Dono 21b], which agrees with the results obtained for the other consistencies in which this axis provided information for the best classification results, even with more features than another Acc axis (see Table 4.2).

Some features extracted here have been intensively analyzed in Acc axes [Kuro 19]: standard deviation, PKF, MDF, W$_{ent}$, LZC and entropy rate. SampEn was computed instead of the Shannon entropy rate, because SampEn is based on a generalization algorithm of the Shannon entropy, originally developed for time series analysis in clinical research [Delg 19]. In general, works oriented to statistical comparisons between features extracted from healthy and dysphagic individuals, have reported some inconsistencies, maybe produced by database-related bias. Even though [Dudi 16] did not find statistically significant differences between safe and unsafe swallows regardless the liquid viscosity, the comparisons were made with dysphagic individuals only. In contrast, [Dudi 18b] found significant differences between healthy and non-healthy individuals, in all of the aforementioned features in AP and SI axes during thin swallows, as well as during viscous swallows except for PKF in AP axis, and skewness and kurtosis of the amplitude in SI. Moreover, [Dono 21a] found statistically significant differences between healthy and non-healthy patients in LZC (AP axis) and the bandwidth (SI axis) extracted during thin liquids delivered by cups. However, when the liquids were swallowed by a spoon, more differences were retrieved (LZC in AP and SI, MDF and bandwidth in the three axes, and W$_{ent}$ in AP). Another study reported significant statistical differences of standard deviation, LZC, entropy rate, PKF, MDF, and W$_{ent}$, in healthy and dysphagic individuals deferentially in the three axes, during swallows of thin liquids [Dono 21b]. Some of these features were also retrieved by the proposed approach using swallows of water, specifically the LZC (SI axis, 5 mL, AUC=0.75), the entropy (ML axis, 5 and 10 mL, AUC=0.73 and 0.79, respectively), the

PKF (M axis, 5 and 10 mL, AUC=0.76 and 0.69, respectively), and W_{ent} (AP axis, 10 mL, AUC=0.78).

Regarding sEMG-related features, in swallowing applications, only one work in the framework of this thesis was published [Rold 18], as mentioned in Experiment #1. The current approach retrieved consistent results in terms of the features selected: the LOG was returned in the current experiment in one sEMG channel during all volumes of yogurt, and it was one of the features with the highest occurrences for sEMG; DASDV was retained in $water_{20}$ and saliva; MYOP was the biomarker with the highest occurrence of the entire feature space and with the highest AUC in comparison with other extracted from sEMG channels in yogurt and $water_{20}$ (see Table 4.2). One divergent result was retrieved by MNF, which was not associated with any sEMG channel.

Analysis of the contribution of sEMG channels and Acc axes

The AP axis yielded the highest AUC values in all swallowing tasks (Table 4.2). This result is in line with the fact that the anterior and superior movements of the hyoid bone and the larynx characterize the pharyngeal phase as a response of a series of different neuromuscular activations [Mats 08b]. But, in general, AP, SI, and ML have shown some capability to detect aspiration/penetration in dysphagic patients [Sejd 13], because the anterior-posterior, superior-inferior, and in a lesser extent of medial-lateral movements, are critical during deglutition [Mao 19]. Thus, the analysis of the kinematic dimension of swallowing using triaxial Acc seems to be potentially useful for clinical dysphagia analyses.

Regarding sEMG, the mRMR method retrieved a few infrahyoid-related features, which is surprising because these muscles have paramount functions in the swallowing process. It could indicate that the features of this channel are redundant with respect to the suprahyoid ones because both have a close relationship with the hyoid and laryngeal movements [Suzu 20]. This unexpected behavior could be also associated with the low signal-to-noise ratio of the infrahyoid movements, maybe produced by cross-talk, shallowness, and size [Rest 17]. Additionally, the RSH did not contribute to the classification in the intake of $water_{20}$. In contrast, the LSH had a contribution in all swallowing tasks (Table 4.2).

The same muscles also achieved the highest AUC among the group of the sEMG-related features, for almost all swallowing tasks, contrarily to the AUC achieved by the few features selected for the RIH (see Figure 4.3). These observations agree with [Erte 03], who claimed that the best way to pick up EMG activity is superficially for suprahyoid, i.e., sEMG, while for infrahyoid the use of needle EMG is required.

Regarding the feature selection, although the use of the mRMR algorithm is reported in the literature as a good strategy for classification purposes, it limits the analysis of dysphagia biomarkers; mRMR does not consider information about the localization of the information sources (channels, axes), which is a valid strategy for classical classification schemes, but it avoids more complete analyses behind the deglutition phenomenon. The use of less *sophisticated* but easy to interpret selection methods, such as the AUC_{ROC}-based selection method, could be more suitable for this kind of analyses. Subsequently, this was considered for multi-modal analysis performed in Experiment #4 (see Section 6.4).

4.3.   Experiment #3: Acoustic biomarkers

This section and the number 6.3 were excerpted with some modifications from the following journal paper: Roldan-Vasco, Sebastian, et al. "Machine learning based analysis of speech dimensions in functional oropharyngeal dysphagia." Computer Methods and Programs in Biomedicine 208 (2021): 106248. This article can be found on the publisher's website at https://www.sciencedirect.com/science/article/abs/pii/S0169260721003229.   The journal's homepage is located at https://www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine, and the publisher's copyright information can be found at https://www.elsevier.com/about/policies/copyright/permissions. In this experiment, the evaluation of the features extracted from different speech dimensions was performed using statistical tests.

4.3.1.   Statistical tests for feature selection per speech dimension

To prevent optimistic or misleading results, statistical tests were made upon 80% of the original database described in Section 6.3, randomly chosen but ensuring age and gender balance between groups. Those features that retrieved significant statistical differences between groups in the 80% of the data, were used as input to train the classifiers.

Phonation features

Few features were significantly different between the two groups of subjects. None of the features were significantly different in all of the five vowels, but the standard deviation of logE significantly differentiates the two groups in four of the five vowels. Similarly, the standard deviation of jitter and F_0 showed significant differences between groups in three vowels (specifically, /i/, /o/, and /u/). Furthermore, the standard deviation of the jitter retrieved $p = 0.01$ for /o/, $p < 0.01$ for /i/, and $p < 0.001$ for /u/. Figure 4.4 illustrates the separability of the mentioned features in the vowel /i/.

These observations suggest that changes in the stability of F_0 during the vibration of the vocal folds is a potential biomarker of dysphagia. A complete list with the results of statistical tests is provided in Table A.1 of Appendix A.

Figure 4.5 shows the box-plots corresponding to the particular case of the standard deviation of logE measured over the five vowels. A clear (and significant) separability is observed for all vowels, except for /a/. Besides, it is important to highlight that such feature shows higher values in the dysphagic group than in the control one. Furthermore, there is more variability in the dysphagic group than in the healthy control. These two results are clear signs of abnormal vocal folds vibration in patients with dysphagia.

Articulation features extracted from sustained vowels

Figure 4.6 shows an illustrative example of the behavior of the vowel triangle in one healthy and another dysphagic individual. Note that the triangle of the control speaker is larger than the one of the patient, which is consistent with the fact that VSA showed significant differences between groups ($\eta^2 = 0.217$, $p < 0.001$). Formants were also less sparse for the control subjects, and this difference was reflected in the VAI ($\eta^2 = 0.183$, $p <$
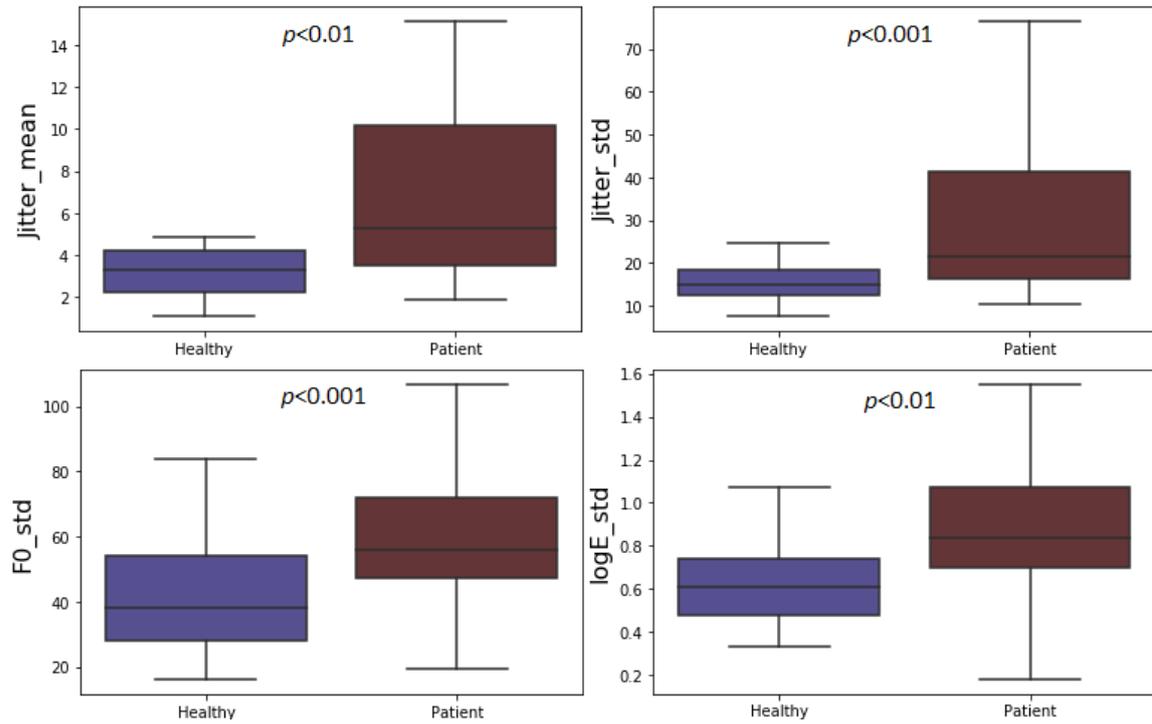
Figure 4.4: Box plots of phonation-related features with $p < 0.01$ extracted from the sustained vowel /i/.

0.001) and the FCR ($\eta^2 = 0.153$, $p < 0.001$). Figure 4.7 shows clear differences between the healthy controls and patients with dysphagia in the vowel triangle-related features.

The set of articulation features extracted from sustained vowels included a total of 172 additional features which correspond to four functionals computed over 43 measurements: 12 MFCCs, F_1, and F_2, with their corresponding first and second derivatives, and the TKEO. Due to space limitations, the results of the statistical tests are presented per functional in Tables A.2, A.3, A.4, and A.5 of Appendix A. The skewness and kurtosis values did not show significant differences between groups in almost none of the vowels and, in those features that had statistical differences, they appeared in only one or two vowels. Conversely, the standard deviation of MFCC1, ΔMFCC1 and Δ^2MFCC1 was significantly different in four or five vowels. This result is in line with previous studies in pathological speech signals where MFCCs coefficients showed good detection results [Frai 09]. However, their main limitation is the lack of direct interpretable outcomes for clinicians. When having a look at the results obtained with the formant F_1, which has a direct relationship with the tongue movement, the standard deviation of F_1, ΔF_1 and $\Delta^2 F_1$ was significantly different in four of the five vowels, while the mean of ΔF_1 showed significant differences in three of them. This is also a promising result that confirms abnormal patterns in the articulation of patients with dysphagia. For the rest of the features and their corresponding functionals, it is hard to find a well-defined pattern.

Articulation features extracted from continuous speech

Articulation features were also extracted from monologues. A total of 18 BBEs were measured in onset and offset transitions (Table A.6, Appendix A). The first Bark band
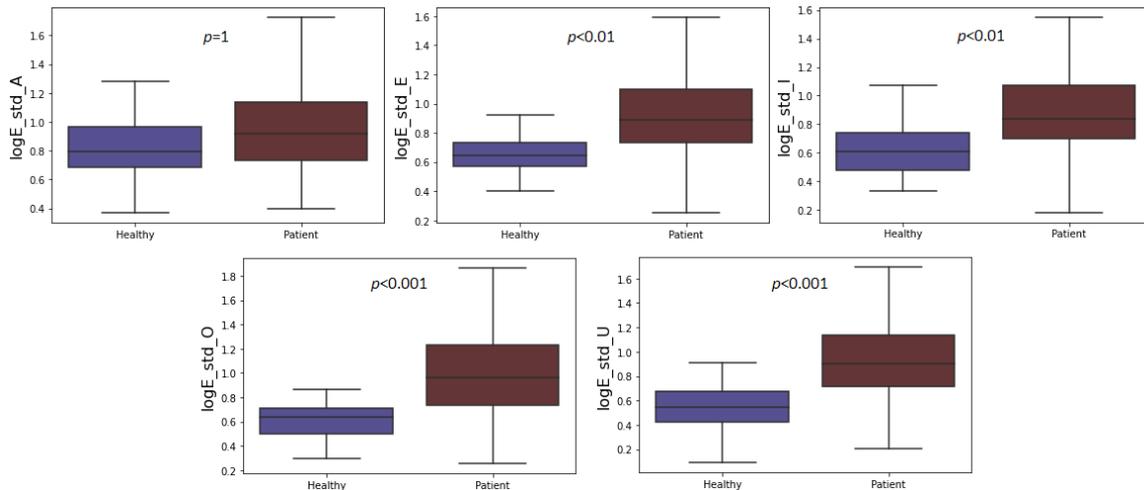
Figure 4.5: Box plots of the standard deviation of logE in the five vowels.
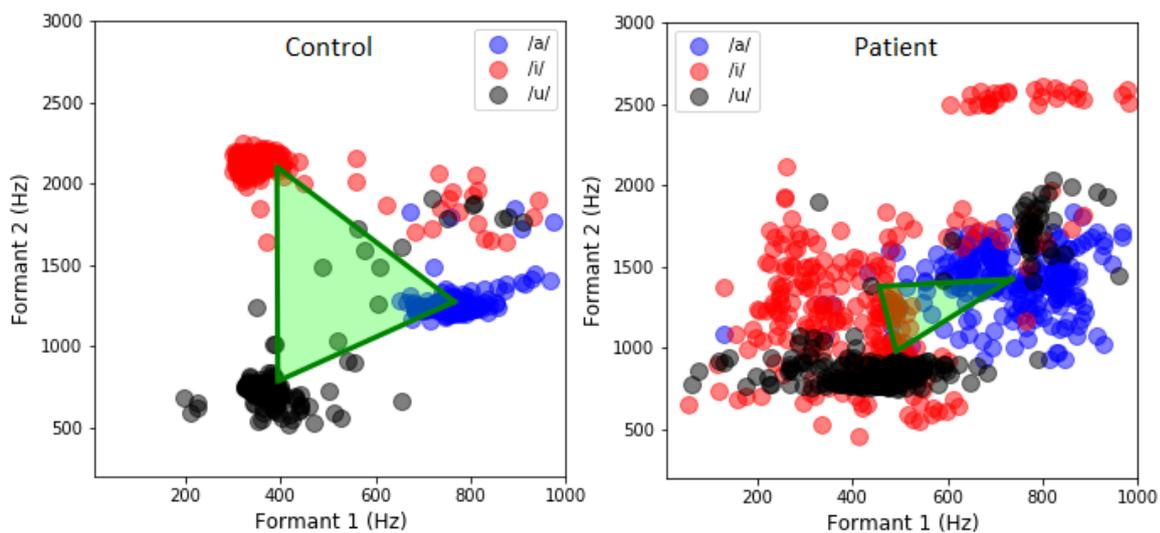
Figure 4.6: Vowel triangles computed in one healthy control (left), and one dysphagic patient (right).

showed statistical differences between groups systematically in both functionals computed over the onsets and offsets. This was also seen on the standard deviation of the second band. Figure 4.8 shows the box-plots resulting from evaluating all of the 18 Bark bands in onsets and offsets. Patients had higher values than the controls, especially in the mean of BBE_1 (onsets and offsets). This pattern was previously observed as present in dysarthric speech signals [Oroz 16b]: Parkinson's patients apparently made more effort to start the vocal fold vibration, which results in higher energy in onsets. In contrast, patients with dysphagia tend to have less standard deviation than healthy controls in offsets (See BBE_1, BBE_2, and BBE_[6 ∼ 13] in Figure 4.8).
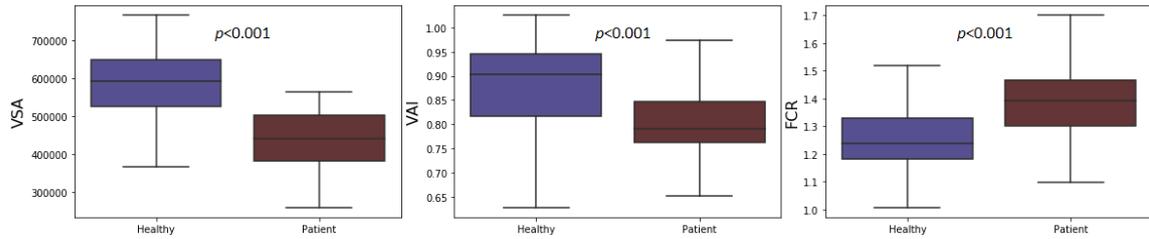
Figure 4.7: Box plots of vowel triangle-related features.
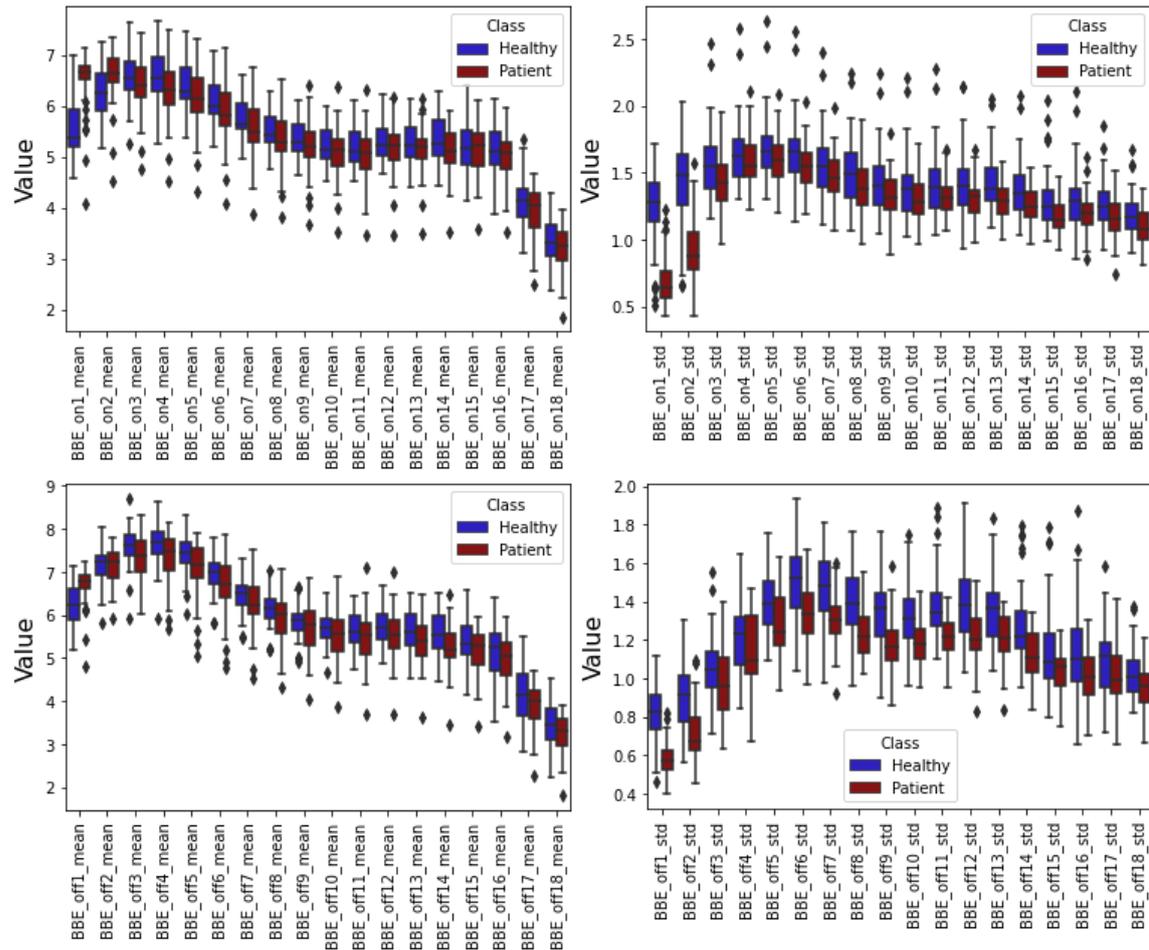
Figure 4.8: Box plots of the Bark-band energies for the onsets (first row) and offsets (second row) of the utterances. The first column shows the mean, and the second one shows the standard deviation.

Articulation features extracted from DDK tasks

Although DDK is related to the articulation dimension, speech tasks designed to evaluate them are different. Thus, their discrimination capability is not necessarily the same. This capability was evaluated from pa-ta-ka recordings. The results obtained are shown in Table A.7 (Appendix A). The mean and variance of the logarithmic energy, as well as the DDK-related features (mean, rate and regularity), were significantly different between groups. It is remarkable that despite the DDK-related features seeming to be suitable to

discriminate between groups (see Figure 4.9), the rate, duration, and regularity of pauses were not. Although all DDK-related features (mean, rate and regularity) retrieved sufficient evidence of significant statistical difference, DDK_mean and DDK_reg had the smallest corrected p-values (< 0.001). Not only the values of such features were higher in patients than in controls, but also the variability of data, which is an indicator of consistency in the DDK dimension from healthy recordings: $\text{DDK}_{\text{mean}}^{\text{Healthy}} = 58.79 \pm 19.39$ vs. $\text{DDK}_{\text{mean}}^{\text{Patients}} = 140.66 \pm 247.14$; $\text{DDK}_{\text{reg}}^{\text{Healthy}} = 4478.23 \pm 4529.37$ vs. $\text{DDK}_{\text{reg}}^{\text{Patients}} = 214581.62 \pm 1102980.95$. This is explained by the articulation impairments produced by the conditions leading to dysphagia (see section 4.3.2).
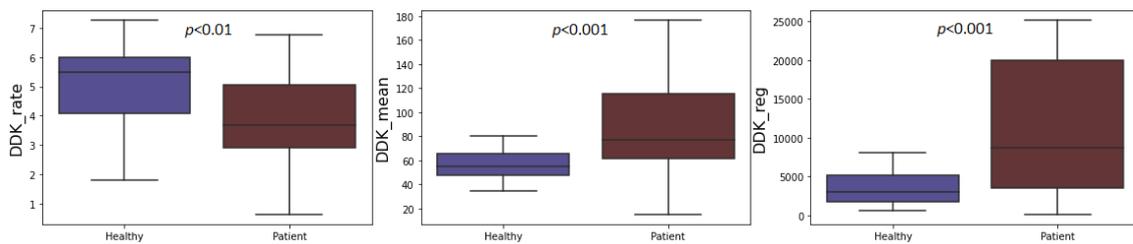
Figure 4.9: Box plots of the DDK features with the lowest p-values for controls vs. patients.

Prosody features

Results with the prosody features indicate that only the standard deviation was the feature with significant differences when comparing patients with dysphagia and healthy subjects (see Table A.8, Appendix A). Since the DDK-related features retrieved statistically significant differences, a careless interpretation could lead to expect similar performance not only for energy but also for voiced segment-related features. However, this is not true since diadochokinetic tasks are completely different from spontaneous speech, in which variations of intonation and rhythm are irregular since each volunteer made his/her own monologue. This kind of non-homogeneity in the recordings of all patients produced p-values of one for almost all prosody features, with the exception of logE and some statistical functionals of voiced and unvoiced segments. Consistently, the effect size of the energy indicates that this was the only feature with practical significance ($\eta^2 = 0.355$).

4.3.2. Speech features as an indicator of dysphagia

To the best of my knowledge, this experiment is the most comprehensive study of speech dimensions (i.e. phonation, articulation, diadochokinesia, and prosody) oriented to the quantitative analysis of swallowing impairment. Obtained results are discussed below.

Phonation: Some authors have worked with commonly used phonation-related features to determine changes in speech production from patients with dysphagia [Kang 17, Ryu 04]. Although [Ryu 04] found that RAP, shimmer, noise-to-harmonic ratio, and voice turbulence index have significant changes after swallowing in a heterogeneous group of patients, [Kang 17] only found significant variations for RAP. Despite the contradictory results, the use of such features is supported by the fact the presence of foreign material

at the larynx, e.g. in dysphagic patients with aspiration or penetration produces acoustic changes in phonation [Muru 10].

In this Experiment, different phonation-related descriptors were found that differentiate between "dysphagic" and healthy conditions. The jitter (mean and standard deviation), as well as the standard deviation of F_0 and logE were larger and more sparse in patients with dysphagia. This makes sense because the phonatory system may be affected by residuals of food or liquids in patients with dysphagia [Grov 07]. Figure 4.4 is an interesting example of the separability of such features in the vowel /i/. It is noteworthy that the mentioned features have significant statistical differences only in the vowels /i/, /o/, and /u/, but neither in /a/ nor in /e/ (except for logE). It could be related to the effort made by patients to pronounce and sustain the tongue and lips positions during the utterances: whilst Spanish vowels /a/ and /e/ have low and mid-front tongue positions, respectively, as well as unrounded lips position, the vowel /i/ has high-front tongue position, and vowels /o/ and /u/ have back positions of the tongue and rounded lips [Gari 19]. However, further and detailed phonatory analyses should be made to confirm such hypothesis.

Interestingly, most of the phonation-related features that retrieved significant changes, do so in the corner vowels (/i/ and /u/), similar to what was observed by [Brui 13] in patients with oropharyngeal cancer.

Articulation: Besides the confirmation of previously reported observations, articulation and prosody dimensions of speech were included in this work, which has been scarcely studied in dysphagia. Mathematical descriptors that could characterize patients with dysphagia based on such speech dimension were found. For the articulation dimension assessed via sustained vowels, the VSA, VAI and FCR from the vowel triangle (see Figure 4.7), the standard deviation of F1 and its first and second derivatives, the mean of its first derivative, the first derivative of F2, as well as the zero, first and second order derivatives of MFCC1, retrieved significant statistical differences between groups. On the other hand, the articulatory analysis in continuous speech showed that the mean and standard deviation of the first two BBEs had antagonistic results for onset and offset segments, i.e., the mean of BBEs was significantly higher in patients than in controls, but the standard deviation was converse.

DDK: On the other hand, the DDK analysis in pa-ta-ka recordings indicates that the most suitable descriptors to analyze voice changes in patients with dysphagia were logE (mean and standard deviation), and the DDK-related features, i.e. rate, mean duration, and regularity. The fact that DDK-related features differed in patients with dysphagia, not only agrees with manifestations of dysarthria [Dani 15] but also suggests that such dysarthria-related conditions could be most noticeable in such patients. This has been studied in patients with Parkinson's disease, and alterations in DDK are apparently related to irregular articulation due to imprecise articulation (e.g., lack of velar contact), which suggests impairment of velopharyngeal control [Mont 18]. In swallowing, the integrity of the velopharyngeal mechanism is paramount, because it provides the necessary seal to isolate the nasopharynx and the aerodigestive tract, preventing reflux [Smit 90]. The normal process requires properly opening or closing events at velopharyngeal junction [Clav 15]. Thus, the observed differences between DDK-related features in patients and controls could be explained by the delayed laryngeal vestibule closure present in the early oropharyngeal

phase in patients with dysphagia, which leads to unsafe swallowing [Clav 15]. It is consistent with the fact that the vowel triangle-related features also showed significant differences. Since these features were extracted in the most extreme articulatory positions of the tongue, it is expected to find large differences between healthy and dysphagic individuals. Further studies oriented to quantification of dysarthria in patients with dysphagia with well-defined scales, e.g. Frenchay dysarthria assessment, are needed to confirm such observation.

Prosody: In the case of the analysis of monologues, only the second statistical moment of the energy contour was statistically different between groups. Note that this feature was also present in the DDK analysis. However, unlike DDK, the duration and rate of voiced and unvoiced segments did not show significant differences in the monologue. Since that monologue is highly variable between individuals, other speech tasks such as the reading of pre-established texts could be considered to analyze the prosody dimension. Such types of texts have been applied to patients with Parkinson's disease [Vasq 18]. This was considered in Experiment #4.

Speech features selected for classification: The criterion of the selection of features for the feeding of the classifiers in Experiment #3 (see Section 6.3.1), was based on the significant differences ($p < 0.01$) between patients and controls. The combination of the selected feature subsets was also evaluated. For the features extracted from the vowels, the following were the criteria to include them in every speech dimension:

- Phonation: to show significant differences in at least three vowels.

- Articulation: to show significant differences in at least three vowels, with at least two of those vowels to be the corner ones (/a/, /i/, and /u/). For generalization purposes, once a certain feature was selected, it was extracted from the five vowels.

Table 4.3 summarizes the features that met the selection criteria and were used as input for the classifiers. For all speech-related dimensions, those features that retrieved corrected p-value smaller than 0.01, exhibited medium or large effect size, which confirms that such features actually represent substantial differences between healthy individuals and patients with functional oropharyngeal dysphagia.

It is important to clarify that, in the prosody analysis, considering that only the standard deviation of logE retrieved $p < 0.001$, also the mean and skewness were included in the feature set. Despite not showing statistical differences between groups, they were included because retrieved $p < 0.05$ and the η^2 value showed medium or large effect size. Figure 4.10 illustrates that the distributions of the mentioned features have clear patterns: the mean and skewness of logE were slower in healthy controls while the standard deviation was higher in this group.

4.4.   Experiment #4: Comprehensive biomarkers in three swallowing dimensions

Even though the previous experiments suggested some swallowing biomarkers for dysphagia analysis, the database size was considerably increased in Experiment #4. Thus, this

Table 4.3: Summary of features selected for classification purposes.

Dimension	Speech tasks	Features
Phonation	Sustained vowels	Jitter_mean, Jitter_std F_0_std logE_std
Articulation	Sustained vowels /a/, /e/, /i/, /o/, /u/	ΔF_1_mean $(\Delta^0, \Delta^1, \Delta^2)F_1$_std ΔF_2_std $(\Delta^0, \Delta^1, \Delta^2)$MFCC1_std
	Sustained vowels /a/, /i/, /u/	VSA, VAI, FCR
	Continuous speech	BBE_on1_mean, BBE_on1_std BBE_on2_std BBE_off1_mean, BBE_off1_std BBE_off2_std BBE_off[6~13]_std
DDK	Repetition of /pa-ta-ka/	logE_mean, logE_var DDK_rate, DDK_mean, DDK_reg
Prosody	Continuous speech	logE_mean, logE_std, logE_skew
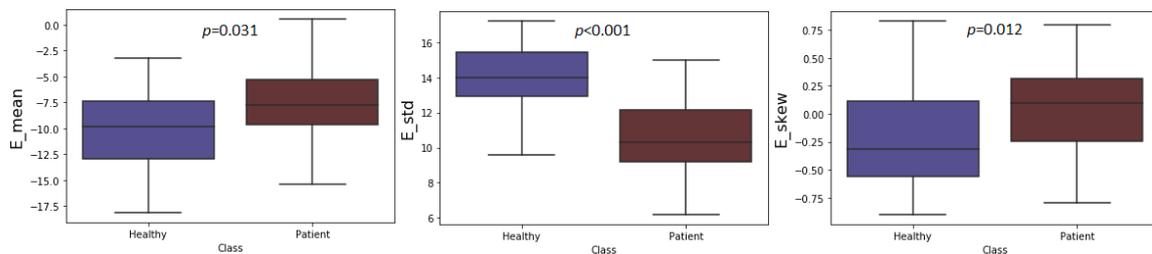
Figure 4.10: Box plot of the prosody-related features chosen for classification purposes.

experiment was intended to confirm or reject the aforementioned notions about biomarkers. The analysis of the largest database is provided per acquisition protocol.

Protocol #1: sEMG

Bearing in mind that AUC$_{ROC}$ showed a well-defined pattern in Experiment #1, it was used as the feature selection method with a threshold of 0.7. Protocol #1, i.e., only sEMG signals acquired in RM, LM, RSH, LSH, RIH, and LIH, retrieved 321 features. Figure 4.11 shows the number of features retrieved by each muscle group and swallowing task. For all swallowing tasks, most of the features were obtained from supra- and infrahyoid muscles. In contrast, masseter muscles retrieved so few features, even in cracker, which produces high activation of such muscles. This observation agrees with Experiment #1 (see Figure 4.2). Furthermore, features retrieved from the RIH achieved the highest AUC$_{ROC}$ (almost all with AUC$_{ROC} > 0.75$).

In contrast to the findings in Experiment #1, the most prevalent domain in the feature set was the frequency one. Actually, WAMP was not retrieved in any muscle or task.
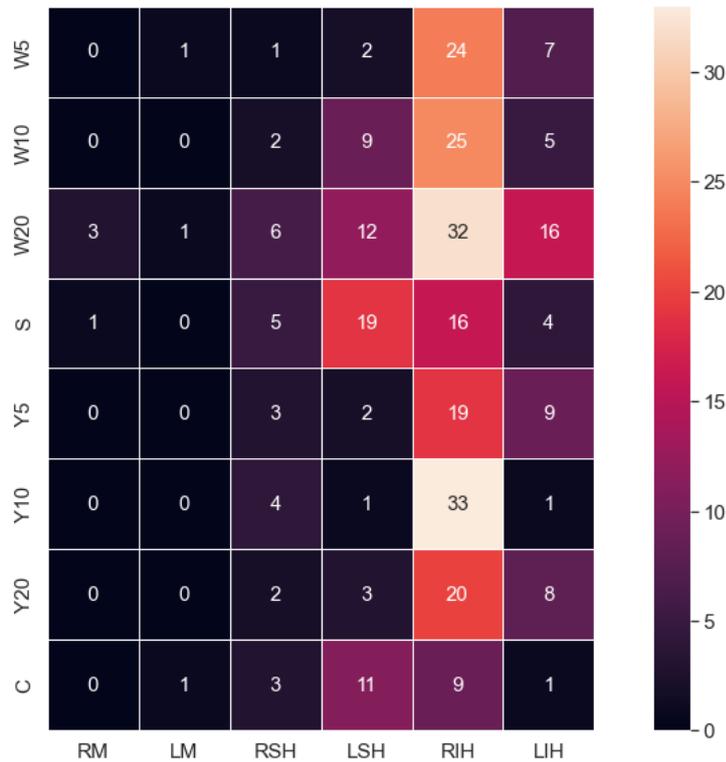
Figure 4.11: Colormap of sEMG related channels that retrieved features with $\text{AUC}_{\text{ROC}} \geq$ 0.7, in protocol #1. Wx: x mL of water; Yx: x mL of yogurt; S: saliva; C: cracker.

Figure 4.12 shows colormaps that illustrate the number of functionals retrieved by each feature and muscle group per swallowing task. Despite saliva retrieved a majority of time domain features, the other tasks were characterized mostly with MNF, MDF, and PKF from the frequency domain, as well as E_{D1}, E_{D2}, and W_{ent} from the time-frequency domain. Just one feature achieved $\text{AUC}_{\text{ROC}} = 0.80$: maximum of the DFA from the RIH in water$_{20}$. The remaining features achieved $0.70 \leq \text{AUC}_{\text{ROC}} < 0.80$.

Protocol #2: sEMG+Acc

Otherwise, Protocol #2 retrieved 308 features from both suprahyoid and right infrahyoid muscles, as well as the three Acc axes. Time domain features and MNP were the most retrieved by the selection method, like in Experiments #1 and #2 (see Figure 4.13). Additionally, E_{D5} and W_{ent} were retrieved by any sEMG channel in almost all tasks, with the exception of saliva.

Additionally, AP and ML axes retrieved the majority of the Acc-related features, in particular for saliva, and yogurt$_{20}$. However, sEMG retrieved more features than Acc, with a high contribution of the LSH (226 vs. 82, see Figure 4.14).

Protocol #3: speech

Regarding the speech-related features, a separate analysis per speech dimension was performed because each of them has specific features, extracted from pre-swallowing recordings in contrast to the Experiment # 3. Figure 4.15 shows that most of the phonation-
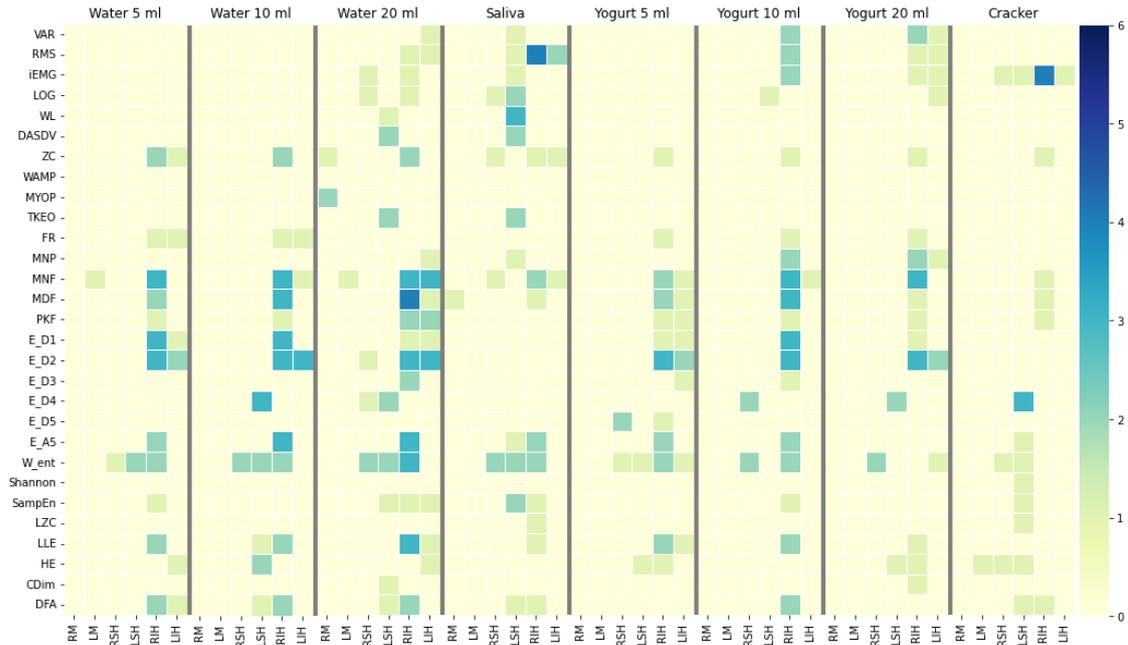
Figure 4.12: Heatmaps of sEMG related features selected by the AUC_{ROC} method distributed along the muscle groups and swallowing tasks, in protocol #1. Color intensity represents the number of functionals per feature.

related features were retrieved (or discarded) by the AUC_{ROC} criterion consistently among vowels, with the exception of APQ (not retrieved in /e/, and /i/). The most prevalent feature was the Jitter (four functionals in all vowels); in contrast, like in Experiment #3 with fewer data, PPQ was not retrieved for any vowel (see Table 4.3). Table C.9 shows the AUC_{ROC} of the features selected and illustrates that Jitter, the first two statistical moments of Shimmer, the standard deviation, and skewness of F_0, the standard deviation and kurtosis of $(\Delta, \Delta^2)F_0$, and the standard deviation of logE, were retrieved in all vowels. In this way, they are proposed as dysphagia biomarkers.

The vowels were also used for articulation analysis. Figure 4.16 shows the colormap extracted from sustained vowels in such speech dimension. Most of the features were also consistent among vowels and agreed with results obtained in Experiment #3 (see Table C.10): $(\Delta^0, \Delta^1, \Delta^2)F_1$ (standard deviation), ΔF_2 (mean), $(\Delta^0, \Delta^1, \Delta^2)\text{MFCC1}$ (standard deviation and additionally the mean for the latter), and MFCC5 (standard deviation). In contrast to Experiment #3, vowel triangle-related features were not retrieved.

Furthermore, each volunteer was asked to read the text described in Chapter 2, after swallowing tasks. Even though in Experiment #3 the individuals performed a free monologue before swallowing different consistencies, it is highlighted that BBE_on1_mean, BBE_on1_std, BBE_on2_std, BBE_off1_mean, BBE_off1_std, and BBE_off2_std were retrieved in both experiments. In Experiment #4 the AUC_{ROC} was high (between 0.79 and 0.85, see Table C.11). Notwithstanding, other features were also retrieved in this Experiment: the mean of BBE_on2 and BBE_off2, the standard deviation of BBE_on3, BBE_off3, BBE_on6, and BBE_off6, as well as the mean of BBE_on[7, 8, 10∼17], and the mean of BBE_off[13∼18].
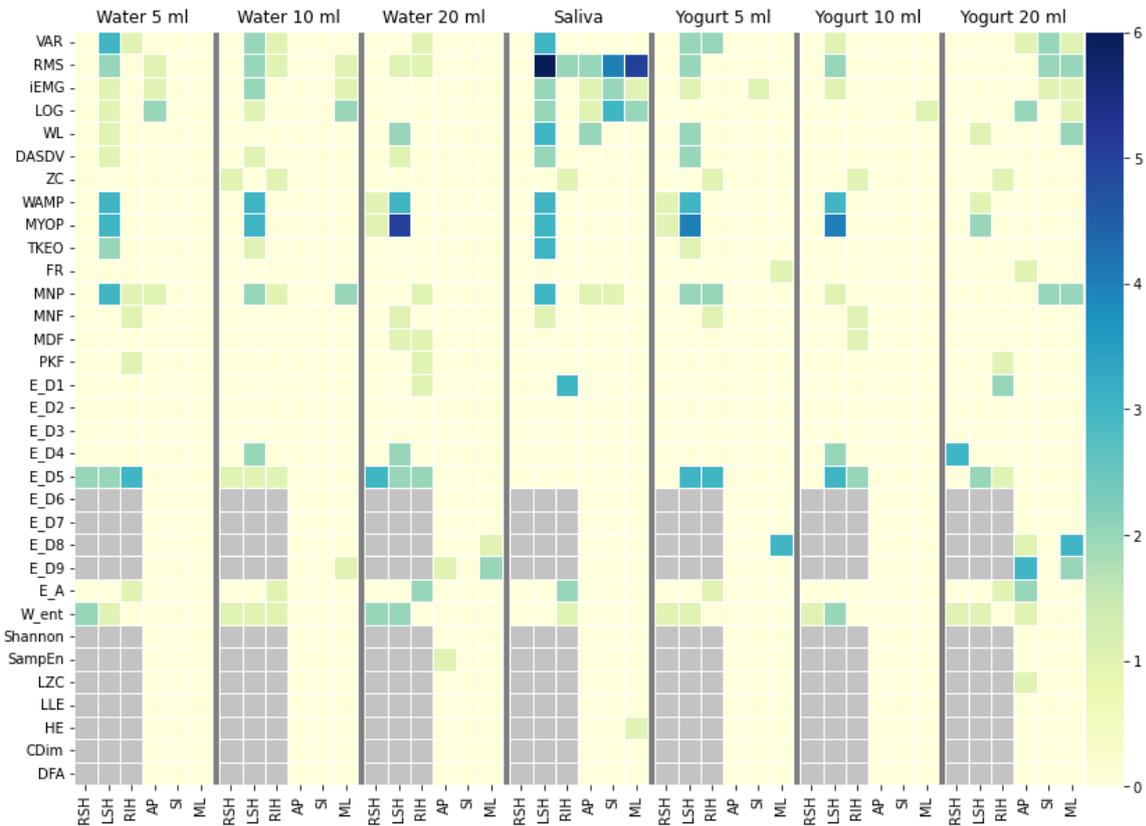
Figure 4.13: Colormaps of features extracted in protocol #2 and selected by the AUC_{ROC} method distributed along sEMG channels and Acc axes, by swallowing task. Color intensity represents the number of functionals per feature.

Consistent results between Experiments #3 and #4 were also consistent in DDK and prosody (see Table C.12) since almost the same features were retrieved in both experiments: the DDK mean and rate, as well as the variance of logE were selected from pa-ta-ka recordings, whilst the 2nd, 3rd, and 4th statistical moments of logE were retrieved from the reading of a preset text for prosody analysis. See Table 4.3 for comparison.

4.4.1.   A concluding analysis of multimodal biomarkers

This Experiment confirmed some findings obtained with the uni-modal experiments. Even though masseter muscles have paramount participation in the mastication of solids [Shaw 13], their related features showed a limited capability of discrimination between populations in the current experiment. Likewise in Experiment #1, did not retrieve a significant contribution in cracker. Additionally, sEMG was well-characterized by time domain features in saliva, and by frequency and time-frequency domains in the other tasks.

In contrast to Experiment #2, the majority of the selected features were extracted from sEMG instead of Acc, which may be related to the fact that such experiment implemented mRMR for selection, whilst in Experiment #4 the method was AUC_{ROC}, i.e., the selection was performed regardless the redundancy. Furthermore, the highest AUC_{ROC} values were achieved by suprahyoid-related features.
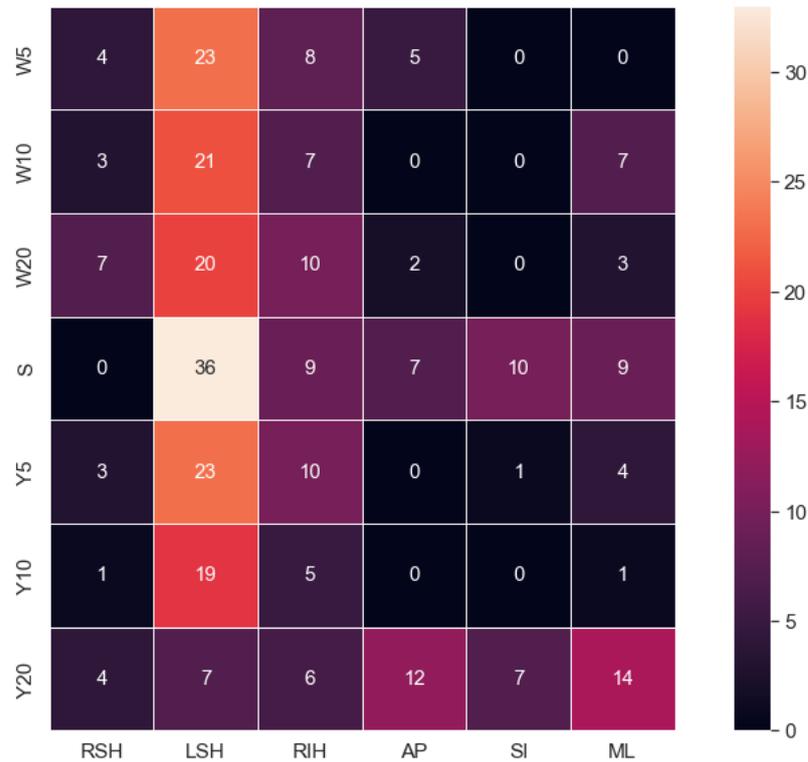
Figure 4.14: Colormap of acquisition channels that retrieved features with $AUC_{ROC} \geq 0.7$, in protocol #2. Wx: x mL of water; Yx: x mL of yogurt; S: saliva.
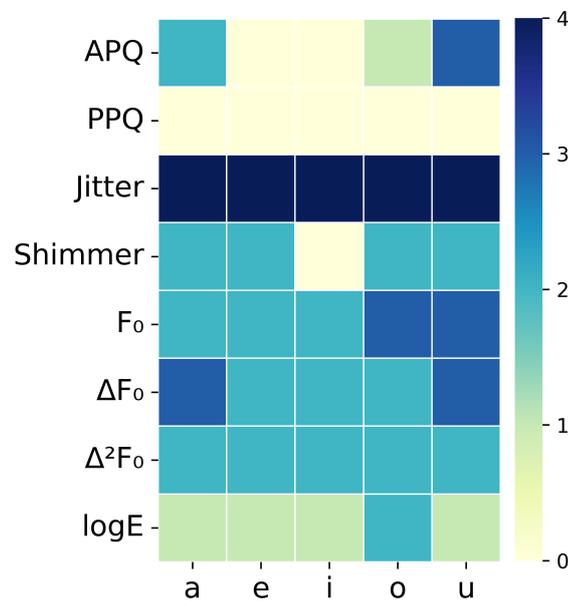
Figure 4.15: Heatmap of phonation-related features extracted in protocol #3 and selected by the AUC_{ROC} method distributed along vowels. Color intensity represents the number of functionals per feature.

Otherwise, AUC_{ROC} was higher in speech-related features than in extracted from sEMG (sEMG$_{P1}$ and sEMG$_{P2}$), and Acc; for instance, the mean of the Jitter achieved $AUC_{ROC} \geq$
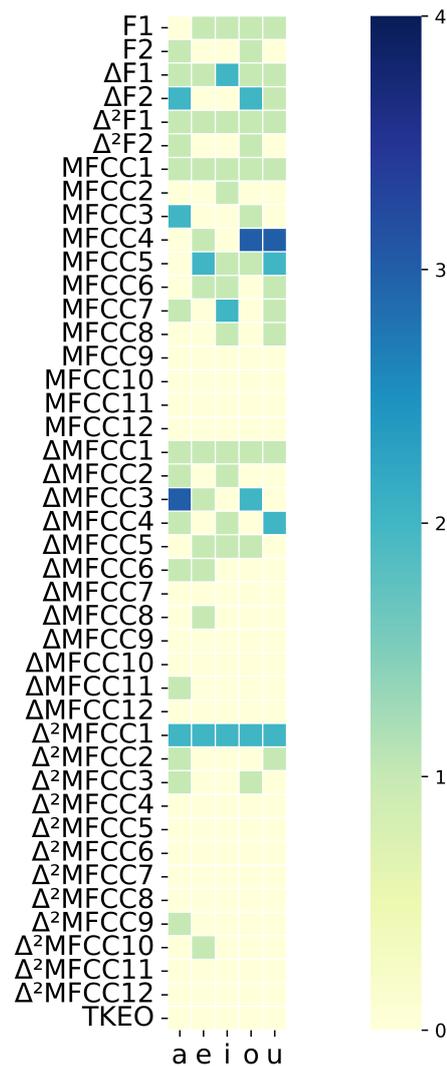
Figure 4.16: Heatmap of articulation-related features extracted in protocol #3 and selected by the AUC_{ROC} method distributed along vowels. Color intensity represents the number of functionals per feature.

0.80 for all vowels (see Table C.9), whilst most of the electrophysiological and mechanical features achieved $\text{AUC}_{\text{ROC}} < 0.75$. Only two features in Experiment #1 (maximum of DFA from RIH in water$_{10}$, and minimum of ZC from the same muscles in saliva), and one feature in Experiment #2 (standard deviation of E_{D5} from LSH in yogurt$_5$), achieved $\text{AUC}_{\text{ROC}} = 0.80$.

Another interesting finding is that in features extracted from sustained vowels in phonation and articulation, the most retrieved functional was the standard deviation. Thus, as expected, the variability of acoustic parameters in patients with dysphagia was higher than in healthy individuals.

Agreement between Experiments #3 and #4 is remarkable in all speech dimensions because the first one was carried out in pre-swallowing recordings, whilst the second one was performed using post-swallowing tasks. Additionally, in Experiment #3 the features were selected with hypothesis tests, whilst in Experiment #4 the selection method was the

AUC$_{ROC}$. This indicates the generalization capability of the obtained results in terms of speech-related biomarkers of dysphagia. There was just one important difference between Experiments #3 and #4: vowel triangle-related features were retrieved in the first experiment but not in the second one. This could be an effect of the swallowing tasks that, even though they do not produce a change in the discrimination capability (at least observable given the consistency of the results), they appear to affect the distribution of the formants in healthy individuals, maybe due to some prandial material after subsequent deglutitions [Grov 07].

These biomarkers were tested in uni-modal classification scenarios. The performance measures outperformed the achieved with no feature selection in most cases (see Section 6.4.1). This is an indicator of the convenience to use the proposed biomarkers since they provide information on different swallowing and speech tasks, different acquisition channels for sEMG and Acc, and different speech dimensions.

Finally, the feature-related functionals describe static electrophysiological patterns rather than dynamics. This could hide temporal characteristics related to the sequentiality of the swallowing process. However, such analyses require other strategies that imply a completely different approach, and further works in this way should be implemented. Notwithstanding, the outcomes indicate that the spatial distribution of the biomarkers from both populations is separable, and the screening capability between healthy individuals and patients with dysphagia is more dependent on the biomarkers than on the models used to build some separation surface in a high dimensional space.

Chapter 5

Machine and Deep Learning algorithms in swallowing evaluation

5.1. Overview of automatic algorithms in the context of swallowing

5.1.1. sEMG-based models

Several models have been proposed to classify gestures or movements using sEMG signals [Nazm 16, Liu 14], such as Artificial Neural Networks (ANN), fuzzy models, hybrid neural networks, Support Vector Machines (SVM), Decision Trees (DT), and Bayesian models [Yous 14]. The main advantage of these methods is that they are not affected by human fatigue, emotional states, or habituation [Suba 12], i.e., they are systematic. Most of the works that model myoelectric behavior are in the field of myocontrol, mainly dealing with large muscles in upper or lower limbs [Duan 16, Phin 13]. In swallowing assessment, only a few simple statistical models have been used [Cons 18].

Although sEMG provides information on the timing and amplitude patterns in oral, pharyngeal and esophageal diseases [Vaim 09], automatic evaluation of swallowing-related sequences by using sEMG signals has not been addressed so far. The sequential activation pattern has shown a degree of population variance and high intersubject variability that difficult to draw a unique model of muscular recruitment [Dell 18, Perl 99]. Thus, despite the sequentiality pattern is an indicator of healthy/unhealthy swallows, few works have been focused on this kind of analysis via sEMG based on automated algorithms.

One work used a fuzzy logic-based system and characterized sEMG signals from submental muscles and sound signals from a microphone for automatic discrimination of dysphagia severity in patients with myasthenia gravis [Hsu 13]. The authors reported an accuracy of 82.2%, overcoming other strategies reported in the literature like those based on k-means, DT and ANN. Another study introduced a method where background, oral and pharyngeal phases were detected in multichannel sEMG recordings [Rold 18]. Time and frequency domain features were extracted to train a SVM and ANN, with an accuracy of about 90%. Finally, another study reported two classification schemes based on classical Machine Learning (MaL) and Deep Learning (DL) methods to model sEMG recordings collected from 8 healthy young males. Signals from supra- and infrahyoid muscles using a sensor array were considered. The classification accuracy of effortful swallows was higher

with the DL scheme [Suzu 20]. In this way, the lack of work oriented to the automatic discrimination between healthy and dysphagic individuals using sEMG is remarkable.

5.1.2. Acc-based models

Different models applied to cervical auscultation signals have been proposed from the classical point of view: linear discriminant analysis, neural networks, probabilistic neural networks, *k*-nearest neighbors, support vector machines, Bayesian classification, fuzzy c-means, among others [Lee 06, Lee 11, Shir 14, Dudi 18b, Sazo 10, Inou 18, Sejd 13, Spad 09, Nikj 11, Sanc 18].

The pioneering work of Lazareck and Moussavi in 2004 [Laza 04] opened the field of computational deglutition using Acc signals. In a small database of healthy and young dysphagic individuals, they used discriminant analysis to classify healthy and dysphagic swallows with an overall accuracy of 94.8 %. In the next years, many works were intended to segment and filter Acc recordings, to obtain a clear profile of real swallows instead of unrelated events [Sejd 09, Lee 09a]. During the last decade, the increase in the number of works that applied ML to Acc signals is noticeable. For instance, a reputation-based classifier was used to detect safe and unsafe swallows, with acceptable performance measures (accuracy = 80.48±5.0 %, sensitivity = 97.1±2 %, and specificity = 64±8.8 %) [Nikj 11]. Best results were obtained with Bayes classification and wavelet packet features (accuracy = 94.6 %, sensitivity = 92.5 %, and specificity = 95.6 %) [Sejd 13]. A Linear Discriminant Analysis for the detection of abnormal swallows, retrieved a sensitivity of 90.4 % and a specificity of 60.0 % [Stee 19]. Bayesian classifiers, which have been used to classify aspirations in normal swallowing in people with dysphagia, achieved an accuracy of 90 % [Dudi 18b]. More recently, classical ML algorithms were used to discriminate swallows from healthy and dysphagic individuals with Acc signal and a microphone [Dono 21b]; authors obtained impressive sensitivity, specificity, and accuracy with logistic regression (99 %), but the results can be optimistic due to the small and unbalanced database, in addition to the bias that swallows by demand could introduce into the results.

Aforementioned models have been used mostly to classify normal and abnormal swallows, but they could be unhelpful in the consulting room. Classification between aspirated and non-aspirated patients could be more useful but the capability to detect aspirations with cervical auscultation remains unclear.

More recently, DL architectures have been performed for the classification and detection of the hyoid bone movement [Mao 19], detection of upper esophageal sphincter [Khal 20b] and laryngeal vestibule closure/opening [Dono 22a]. Such models have been implemented to track different swallowing-related movements using videofluoroscopy images and Acc signals [Khal 20a, Dono 20]. The best results have been achieved for esophageal sphincter closure detection [Dono 22b]. Additionally, deep belief networks were used for healthy and non-healthy swallow detection [Dudi 18a], with good performance (accuracy = 91.3 %, sensitivity = 85.7 %, and specificity = 96.9 %). Even though the difference with the performance obtained in [Sejd 13] does not seem remarkable, the importance of this result is the generalization capability of the work by Dudik et al., because of the number of individuals and the control group.

In summary, the Acc signal has been widely approached for swallowing evaluation, but the models implemented for analysis are far to be used in the consulting room because

they have achieved a modest (and even contradictory) equilibrium between sensitivity and specificity. Furthermore, the clinical validation of the aforementioned results is still open.

5.1.3. Speech-based models

As mentioned in Section 3.3.1, the literature in computational paralinguistics is extensive, e.g. works oriented to the automatic assessment of dysarthria, Parkinson, dementia, or speech in patients with cochlear implants or cleft lip and palate [Vasq 18, Pere 22, Aria 21, Vasq 20]. This is a topic with academic increasing interest because it supports the development of computer-aided tools for diagnosis [Vasq 20]. There are recent works oriented to automatic pathology detection by speech analysis [Barr 20, Moha 20, Vasq 21].

Several automation strategies have been implemented, such as classical machine learning methods (e.g. linear regression, SVM, Bayesian classifiers, hidden Markov models, dynamic time warping, random forest, gaussian mixture models, and neural networks) [Vasq 18, Trav 17, Much 17, Bhat 17, Ijit 17, Gill 17, Vyas 16]. More recently, deep learning approaches have been applied such as convolutional, recurrent and combined convolutional-recurrent neural networks, autoencoders, or gated multimodal units [Aria 21, Vasq 20, Alos 21].

Nevertheless, they have performed in dysphagia-related applications in a limited number of papers: one intended to classify wet voice with support vector machines on the utterance of "aeiou" prior to and immediately following swallowing, with promising accuracy but with a limited database [Ipin 18]; another paper evaluated the performance of logistic regression, decision trees, random forest, support vector machines, Gaussian mixture models and extreme gradient boost to detect dysphagia severity (oral vs. non-oral feeding) and risk of respiratory complications in post-stroke patients [Park 22]. An additional paper trained a multilayer perceptron-based integrated classifier, using a support vector machine and an adaptive boosting internally [Zhao 22]. This scheme was used to detect dysphagia from throat vibrators using speech analysis. By the way, Table 1.2 shows that. In this manner, the application of computational para-linguistics in dysphagia has a lot of research opportunities in order to answer questions such as: are the speech recordings good descriptors of swallowing disorders?; does the speech have related biomarkers with discrimination capability between healthy individuals and patients with dysphagia?; are there differences between algorithms to classify healthy and dysphagic individuals? These questions were addressed in the current thesis.

5.2. Machine learning algorithms implemented in this thesis

A ML algorithm learns from data [Good 16]. The term "machine" refers to an automated system that could be implemented, for example, in software [Guyo 08] The following definition of learning has become the reference in the ML field: "a computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E" [Mitc 97]. In ML, the task T could be classification, regression, transcription, translation, anomaly detection, imputation, synthesis and sampling, and denoising, among

others [Good 16]. In the current thesis, the task T is the classification of healthy and dysphagic individuals, the experience E is the *training data* (see Section 5.3 for details), and the performance measure P is the AUC[1]. The learning model can be seen as the function:

$$f : \mathcal{X} \to \mathcal{Y} \tag{5.1}$$

The function f transforms the objects of $\mathcal{X}$ to the set of target values $\mathcal{Y}$, also called *labels* [Gero 22]. In this way, the learning algorithm is asked to produce f, which usually depends on *adaptive parameters* [Guyo 08]. These parameters are obtained by training processes using a sequence of data $\mathcal{D}$ defined in the space $\mathcal{X} \times \mathcal{Y}$ [Guyo 08]:

$$\mathcal{D} = \{\langle \mathbf{x}^{(1)}, y_1 \rangle, \langle \mathbf{x}^{(2)}, y_2 \rangle, \ldots, \langle \mathbf{x}^{(m)}, y_m \rangle\} = \langle \mathbf{X}, Y \rangle \tag{5.2}$$

When $y = f((X))$, the model assigns a numeric code or label $y^{(i)}$ to an input described by the n-dimensional vector $\mathbf{x}^{(i)}$ [Good 16]; this input is the feature space or feature observations of the *i*-th individual (described in Section 3.4). This work focuses on bi-class classification, i.e. $|\mathcal{Y}| = 2$. For convenience, it is often assumed that $\mathcal{Y} = \{-1, +1\}$, $\mathcal{Y} = \{0, 1\}$ or $\mathcal{Y} = \{1, \ldots, c\}$ [Guyo 08] For the *i*-th sample, that is, for the *i*-th subject in the database, the label was:

$$y^{(i)} = \begin{cases} 0, & \text{if healthy} \\ 1, & \text{if dysphagic} \end{cases} \tag{5.3}$$

Since the training data of the subjects are labeled by the desired solutions, this thesis addresses a *supervised* learning problem[2]. The goal is to build the function f that maximizes the performance P.

5.2.1. Classical machine learning algorithms

The most important supervised ML algorithms are [Gero 22]: k-Nearest Neighbors, Logistic Regression, Support Vector Machines, Decision Trees, Random Forests, and Neural Networks. These algorithms were implemented in this thesis, in addition to Linear Discriminant Analysis and Extreme-Gradient Boosting. The fundamentals of these algorithms are explained next.

Linear Discriminant Analysis (LDA)

The simplest representation of linear discriminant functions is [Bish 06]:

$$y(\mathbf{x}) = \mathbf{w}^{\mathbf{T}}\mathbf{x} + b, \tag{5.4}$$

where $\mathbf{w}$ is a weight vector, and b is the bias term. The classification is performed as follows [Bish 06]:

---

[1]Accuracy, F_1 score, precision, sensitivity, and specificity were also computed, but the maximum AUC was the optimization criterion to learn models (further details are provided in Section 5.3)

[2]Neither unsupervised learning, in which there are no labels for the data, nor semi-supervised learning, are not covered in this document.

$$y = \begin{cases} \mathcal{C}_1, & \text{if } f(\mathbf{x}) \geq 0 \\ \mathcal{C}_2, & \text{otherwise} \end{cases} \tag{5.5}$$

The decision boundary, i.e. the n-dimensional curve that retrieves the minimum probability of misclassification, is defined by the relation $f(\mathbf{x}) = 0$ (see Figure 5.1)
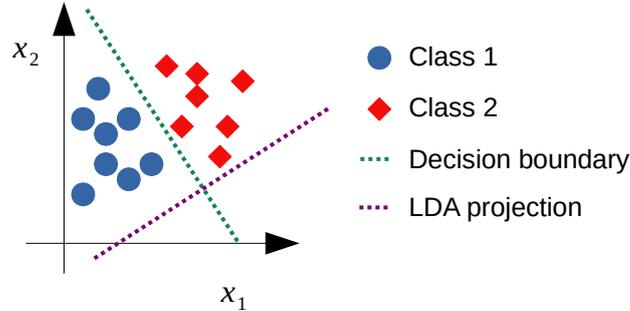
Figure 5.1: Representation of a linear decision boundary for a bi-class classification problem, assuming only two features x_1 and x_2. The LDA projection represents $\mathbf{w}$ that maximizes the between-class variance while minimizing the within-class one.

The decision boundary is an $(n\text{-}1)$ dimensional hyperplane within an n dimensional feature space [Bish 06]. The vector $\mathbf{w}$ is perpendicular to the decision boundary. In case such hyperplane does not exist, the problem would be linearly inseparable [Guyo 08]. One way to determine the linear discriminant function is by Fisher's linear discriminant. It is based on the means computed over the two classes:

$$\mu_\mathbf{1} = \frac{1}{m_1} \sum_{i \in \mathcal{C}_1} \mathbf{x}^{(i)}, \qquad \mu_\mathbf{2} = \frac{1}{m_2} \sum_{i \in \mathcal{C}_2} \mathbf{x}^{(i)}, \tag{5.6}$$

where $\mu_\mathbf{k}$ and m_k are the mean vector and number of points of the class $\mathcal{C}_k$, respectively [Bish 06], and $\mathbf{x}^{(i)}$ is the feature vector of the *i*-th sample (individual). The goal is to maximize the distance between the projected averages μ_1 and μ_2, as follows [Guyo 08]:

$$\mu_2 - \mu_1 = \mathbf{w}^T(\mu_\mathbf{2} - \mu_\mathbf{1}) \tag{5.7}$$

The final Fisher criterion is [Guyo 08]:

$$J(\mathbf{w}) = \frac{(\mu_2 - \mu_1)^2}{s_2^2 + s_1^2}, \tag{5.8}$$

where s_k^2 is the within-class variance of the class $\mathcal{C}_k$, and $(\mu_2 - \mu_1)^2$ is the between-class variance. So, the Fisher criterion maximizes the ratio of the between-class variance to the within-class variance [Bish 06], and it is usually written in matrix form, as follows:

$$J(\mathbf{w}) = \frac{\mathbf{w}^\mathbf{T} \boldsymbol{\Sigma}_\mathbf{B} \mathbf{w}}{\mathbf{w}^\mathbf{T} \boldsymbol{\Sigma}_\mathbf{W} \mathbf{w}}, \tag{5.9}$$

where $\boldsymbol{\Sigma}_\mathbf{B}$ is the between-class covariance matrix to maximize, given by:

$$\Sigma_{\mathbf{B}} = (\mu_{\mathbf{2}} - \mu_{\mathbf{1}})(\mu_{\mathbf{2}} - \mu_{\mathbf{1}})^{\mathbf{T}}, \tag{5.10}$$

and $\Sigma_{\mathbf{W}}$ is the within-class covariance matrix to minimize, given by:

$$\Sigma_{\mathbf{W}} = \sum_{j \in \mathcal{C}_1} (\mathbf{x}_j - \mu_{\mathbf{1}})(\mathbf{x}_j - \mu_{\mathbf{1}})^T + \sum_{j \in \mathcal{C}_2} (\mathbf{x}_j - \mu_{\mathbf{2}})(\mathbf{x}_j - \mu_{\mathbf{2}})^T \tag{5.11}$$

Eq. 5.9 can be represented as:

$$\max_{\mathbf{w}} \mathbf{w}^{\mathbf{T}} \Sigma_{\mathbf{B}} \mathbf{w} \quad \text{subject to} \quad \mathbf{w}^{\mathbf{T}} \Sigma_{\mathbf{W}} \mathbf{w} = 1 \tag{5.12}$$

Likewise in PCA, this is a generalized eigenvalue problem suitable to handle via SVD, where $\mathbf{w}$ is given by the largest value of $\Sigma_{\mathbf{W}}^{-1}\Sigma_{\mathbf{B}}$ [Hast 09]. In this way, the problem can be expressed with Lagrange multipliers:

$$\max_{\mathbf{w}} \{ \mathbf{w}^{\mathbf{T}} \Sigma_{\mathbf{B}} \mathbf{w} - \lambda \mathbf{w}^{\mathbf{T}} \Sigma_{\mathbf{W}} \mathbf{w} \} \tag{5.13}$$

Although LDA is a classification algorithm, it is also used as a feature selection method since it learns the most discriminative axes between classes [Gero 22], as can be seen in the purple line in Figure 5.1.

Logistic Regression

Logistic regression is used for both regression and classification. A general opinion established that logistic regression has fewer assumptions, and is safer and more robust than LDA, but both models produce similar results [Hast 09]. In binary classification applications, it is used to estimate the probability that an instance belongs to a class [Gero 22]. The estimated probability is computed as follows [Gero 22, Bish 06]:

$$p(y = 1|\mathbf{x}; \boldsymbol{\theta}) = h_\theta(\mathbf{x}) = \sigma\left(\boldsymbol{\theta}^T \cdot \mathbf{x}\right) = \frac{1}{1 + e^{-\boldsymbol{\theta}^T \cdot \mathbf{x}}}, \tag{5.14}$$

where $\boldsymbol{\theta}$ is the vector of parameters, and $\sigma(\cdot)$ is the sigmoid function, also called logistic function. The output of $\sigma(\cdot)$ will be always between 0 and 1. If the probability is greater than 0.5, the classifier assigns one class, otherwise, another class is assigned [Gero 22]:

$$\hat{y} = \begin{cases} 1, & \text{if } h_{\boldsymbol{\theta}}(\mathbf{x}) \geq 0.5 \\ 0, & \text{if } h_{\boldsymbol{\theta}}(\mathbf{x}) < 0.5 \end{cases} \tag{5.15}$$

The training process consists in finding the parameters $\boldsymbol{\theta}$ such that the probability be high for $y = 1$ and low for $y = o$. The loss function (error) of one single sample, i.e. one individual, is given by:

$$Loss\left(h_{\boldsymbol{\theta}}(\mathbf{x}), y\right) = \begin{cases} -\log\left(h_{\boldsymbol{\theta}}(\mathbf{x})\right), & \text{if } y = 1 \\ -\log\left(1 - h_{\boldsymbol{\theta}}(\mathbf{x})\right), & \text{if } y = 0 \end{cases} \tag{5.16}$$

Thus, if the actual label is $y = 1$ and $h_{\boldsymbol{\theta}}(\mathbf{x}) \to 0$, the cost will be high (bad classification), but if $h_{\boldsymbol{\theta}}(\mathbf{x}) \to 1$, the cost will be low (good classification). In contrast, if the actual label is $y = 0$ and $h_{\boldsymbol{\theta}}(\mathbf{x}) \to 1$, the cost will be high (bad classification), but if $h_{\boldsymbol{\theta}}(\mathbf{x}) \to 0$,
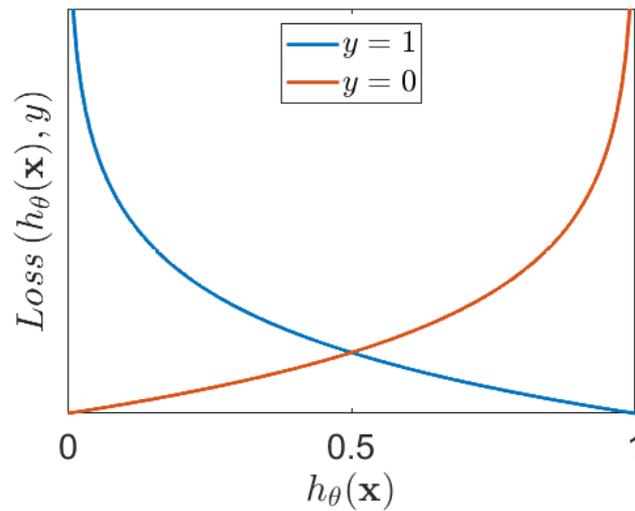
Figure 5.2: Behavior of the loss function in the logistic regression.

the cost will be low (good classification). Figure 5.2 illustrates this behavior. Eq. 5.16 can be re-written as follows:

$$Loss\left(h_{\boldsymbol{\theta}}(\mathbf{x}), y\right) = -y \log\left(h_{\boldsymbol{\theta}}(\mathbf{x})\right) - (1-y) \log\left(1 - h_{\boldsymbol{\theta}}(\mathbf{x})\right) \tag{5.17}$$

The cost function is the mean of the loss functions of all instances, as follows [Gero 22]:

$$\begin{aligned} J(\boldsymbol{\theta}) &= \frac{1}{m} \sum_{i=1}^{m} Loss\left(h_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}), y^{(i)}\right) \\ &= -\frac{1}{m} \sum_{i=1}^{m} y^{(i)} \log\left(h_{\boldsymbol{\theta}}(\mathbf{x}^{(i)})\right) + (1-y^{(i)}) \log\left(1 - h_{\boldsymbol{\theta}}(\mathbf{x}^{(i)})\right) \end{aligned} \tag{5.18}$$

The goal is to find the values of $\boldsymbol{\theta}$ that minimize $J(\boldsymbol{\theta})$, but it is nonlinear in $\boldsymbol{\theta}$ and it has not closed-form [Good 16, Hast 09]. However, $J(\boldsymbol{\theta})$ is convex, so there are techniques to find the global minimum, such as the *Gradient Descent*. The formula to update the parameters is given as follows[Bish 06]:

$$\boldsymbol{\theta}_j^{(\tau+1)} = \boldsymbol{\theta}_j^{(\tau)} - \eta \frac{\partial}{\partial \boldsymbol{\theta}_j} J(\boldsymbol{\theta}), \tag{5.19}$$

where $\boldsymbol{\theta}_j^{(\tau)}$ is the *j*-th parameter computed in the τ-th iteration, η is the *learning rate* (recommended to be small [Bish 06]), and the partial derivative is given as follows [Gero 22]:

$$\frac{\partial}{\partial \boldsymbol{\theta}_j} J(\boldsymbol{\theta}) = \frac{1}{m} \sum_{i=1}^{m} \left(h_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}) - y^{(i)}\right) x_j^{(i)}, \tag{5.20}$$

where $x_j^{(i)}$ is the *j*-th feature of the *i*-th individual.

In order to control the over-fitting risk, regularization over the cost function is recommended, which produces a parameter shrinkage [Bish 06] Note that Eq. 5.18 is data dependent. The regularization adds an error term that depends on the parameters $\boldsymbol{\theta}$. The general regularized cost function is given by [Bish 06]:

$$J_{reg}(\boldsymbol{\theta}) = J(\boldsymbol{\theta}) + \frac{\lambda}{2} \sum_{i=1}^{n} |\theta_i|^q, \tag{5.21}$$

where λ is the regularization coefficient that controls the relative importance of $J(\boldsymbol{\theta})$, and q determines the type of regularization. If excessive large λ, there will be the risk of underfitting, but if $\lambda \to 0$ there will be risk of overfitting [Good 16]. If $q = 0$, no regularization is applied; if $q = 1$, the so-called LASSO[3] or L1-norm regularization is applied, which produces some parameters to be driven to zero; and if $q = 2$, ridge or L2-norm regularization is applied, which produces smaller penalization than L1. Since L2 regularization is a good default [Gero 22], it was used in this thesis.

Artificial Neural Networks (ANN)

Many problems cannot be modeled properly with logistic regression; so an extension of this model is performed to capture non-linear relationships, a technique known as feed-forward neural network, also called multilayer perceptron [Guyo 08], onward ANN. The basic unit of an ANN is the neuron, represented by a node (see Figure 5.3a).

(a) A neuron   **(b)** Four-layer fully connected ANN
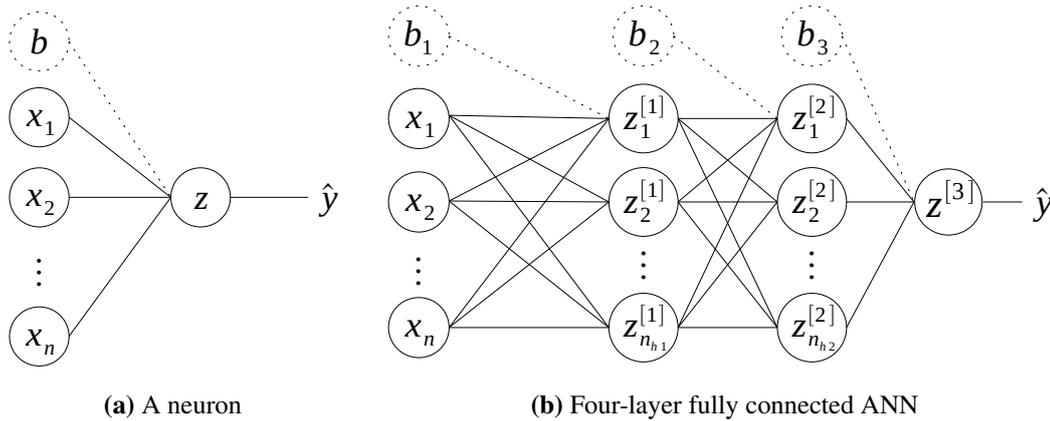
Figure 5.3: Computational representation of a single neuron and an ANN fed with n features

The output of the neuron is highly related to the decision boundaries in LDA and logistic regression:

$$z = \mathbf{w^T}\mathbf{x} + b, \tag{5.22}$$

where z is a scalar. The weight vector is given $\mathbf{w} = [w_1, w_2, \ldots w_n]$, where w_i denotes the connection of the i-th input to the node. An *activation function* $\phi(\cdot)$ is applied, and the final output is given as follows:

$$\hat{y} = \phi(z) \tag{5.23}$$

The most implemented activation function are the logistic (sigmoid) one, hyperbolic tangent (tanh), and Rectified Linear Unit (ReLU [4]). The basic ANN, has one input layer

---

[3]Acronym for Least Absolute Shrinkage and Selection Operator regression [Gero 22]
[4]$\text{ReLU}(x) = \max(0, x)$

with the features, one or more intermediate layers, known as *hidden layers*, and one output layer (see Figure 5.3b). When all neurons of a layer are connected to all neurons of the next one, the ANN is called as *fully connected* network [Huds 00]. The hidden layer contains multiple neurons in parallel, each representing a vector-to-scalar function [Good 16]. The governing equations for the l-th are:

$$\mathbf{z}^{[l]} = \mathbf{W}^{[l]}\mathbf{a}^{[l-1]} + \mathbf{b}^{[l]} \tag{5.24}$$

$$\mathbf{a}^{[l]} = \phi^{[l]}\left(\mathbf{z}^{[l]}\right) \tag{5.25}$$

The weight matrix is $\mathbf{W}^{[l]} = \left(w_{ij}^{[l]}\right) \in \mathbb{R}^{n^{[l]} \times n^{[l-1]}}$, where $w_{ij}^{[l]}$ represents the connection between the i-th node of layer $l-1$, and the j-th node of the l-th layer [Huds 00], and $n^{[l]}$ is the number of neurons in the l-th layer. Furthermore, $\mathbf{b}^{[l]}$ is the bias term of the l-th layer; $\mathbf{a}^{[l-1]}$ and $\mathbf{a}^{[l]}$ are the input and output of the l-th layer, respectively; and $\phi^{[l]}$ is the activation function of the l-th layer. In this way, for an L-layer ANN, $\mathbf{a}^{[0]} = \mathbf{x}$ and $\mathbf{a}^{[L]} = \hat{\mathbf{y}}$. The weights of the input layer indicate the contribution of each feature; however, the weights of the other layers are difficult to interpret [Huds 00].

The initial weights should be small to avoid $\mathbf{z}$ becoming extremely large or small, especially for DNN. The training process of the network involves two stages in counterflow [Wyth 93]: the forward propagation or activation (see Eq. 5.24), and the backward propagation of error (backpropagation). The backpropagation is a type of non-linear Gradient Descent [Huds 00]. Similar to Eq. 5.17, the error function of any layer can be defined as the cross-entropy cost:

$$J = -\frac{1}{m}\sum_{i=1}^{m} y^{(i)} \log\left(a^{[L](i)}\right) + (1 - y^{(i)}) \log\left(1 - a^{[L](i)}\right), \tag{5.26}$$

where m is the number of training samples and $a^{[L](i)}$ represents the estimated output of the last layer L. The backpropagation governing equations are the gradients of the cost with respect to the activations, weights, and biases:

$$\frac{\partial J}{\partial \mathbf{z}^{[l]}} = \left[\mathbf{W}^{[l+1]}\right]^T \frac{\partial J}{\partial \mathbf{z}^{[l+1]}} \odot \phi'^{[l]}\left(\mathbf{z}^{[l]}\right) \tag{5.27}$$

$$\frac{\partial J}{\partial \mathbf{W}^{[l]}} = \frac{\partial J}{\partial \mathbf{z}^{[l]}}\left[\mathbf{a}^{[l-1]}\right]^T \tag{5.28}$$

$$\frac{\partial J}{\partial \mathbf{b}^{[l]}} = \frac{\partial J}{\partial \mathbf{z}^{[l]}}, \tag{5.29}$$

where $\odot$ denotes the Hadamard product, and $\phi'^{[l]}$ is the derivative of the activation function used in the l-th layer. The derivative in the output layer is given by $\frac{\partial J}{\partial \mathbf{z}^{[L]}} = \mathbf{a}^{[L]} - \mathbf{y}$. At each iteration, gradients can be computed over the entire training samples (*Batch Gradient Descent*), over a fraction of the training samples (*MiniBatch Gradient Descent*), or over one random sample (Stochastic Gradient Descent - SGD) [Good 16]. The perceptron learning rule to update the weights and biases for each iteration is quite similar to Eq. 5.19 [Gero 22, Huds 00, Wyth 93]:

$$\mathbf{W}^{[l]} = \mathbf{W}^{[l]} - \eta \frac{\partial J}{\partial \mathbf{W}^{[l]}} \tag{5.30}$$

$$\mathbf{b}^{[l]} = \mathbf{b}^{[l]} - \eta \frac{\partial J}{\partial \mathbf{b}^{[l]}}, \tag{5.31}$$

where η is the learning rate. Selection of the proper η is not trivial since it has an important effect on model performance [Good 16]. Subsequently, computationally efficient algorithms with adaptive learning rate have been introduced for optimization of the weights and biases such as gradient descent with momentum, AdaGrad, RMSProp, and Adam[5]. Further details are provided in [King 14]. In this thesis, SGD and Adam were used for learning.

Support Vector Machines (SVM)

The SVM classifier is intended to find linear or highly nonlinear boundaries in the feature space [Guyo 08], even with large (or infinite) dimension [Hast 09]. SVM does not provide posterior probabilities [Bish 06], that is $p(y|\mathbf{x})$. The main concept of the SVM is the *hyperplane* which separates different classes with maximal margin [Guyo 08]. The hyperplane is defined in a similar way as the linear discriminant function:

$$\mathbf{w^T x} + b = 0, \tag{5.32}$$

where $\mathbf{w}$ is perpendicular to the hyperplane. The margin is the distance between the hyperplane and the closest training point of any class [Guyo 08]. Otherwise, the margin is defined as the distance between the *support vectors*, which are the hyperplanes $\mathbf{w^T x} + b = \pm 1$ (see Figure 5.4). Thus, the goal is to maximize the minimum distance between support vectors and the hyperplane [Guyo 08]:

$$\max_{\mathbf{w},b} \quad \min\{||\mathbf{x} - \mathbf{x}^{(i)}|| \quad : \quad \mathbf{w^T x} + b = 0, \quad i = 1, \ldots, m\} \tag{5.33}$$

The distance between two parallel hyperplanes defined by $\mathbf{w^T x} + b_1 = 0$ and $\mathbf{w^T x} + b_2 = 0$ is equal to $|b_1 - b_2|/||\mathbf{w}||$; thus, the distance between the support vectors, i.e. the margin, is equal to $2/||\mathbf{w}||$ [Guyo 08]. The optimization problem requires maximizing this distance, that is, to maximize $||\mathbf{w}||^{-1}$, which is equivalent to minimizing $||\mathbf{w}||^2$, and the following optimization is proposed[Bish 06]:

$$\min_{\mathbf{w},b} \frac{1}{2}||\mathbf{w}||^2 \quad \text{subject to} \quad y^{(i)}(\mathbf{w^T x} + b) \geq 1 \quad i = 1, \ldots, m \tag{5.34}$$

Notwithstanding, if data are not linearly separable, the construction of optimal hyperplanes is not possible [Guyo 08]. This is solved with the *soft margin* hyperplane [Cort 95]. The goal is to separate the training set with a minimal number of errors. For non-negative variables ξ_i, equation 5.34 becomes in [Cort 95]:

$$\min_{\mathbf{w},b} \frac{1}{2}||\mathbf{w}||^2 + C \sum_{i=1}^{m} \xi_i \quad \text{subject to} \quad y^{(i)}(\mathbf{w^T x} + b) \geq 1 - \xi_i \quad i = 1, \ldots, m, \tag{5.35}$$

---

[5]Acronym for ADAptive Moment estimation
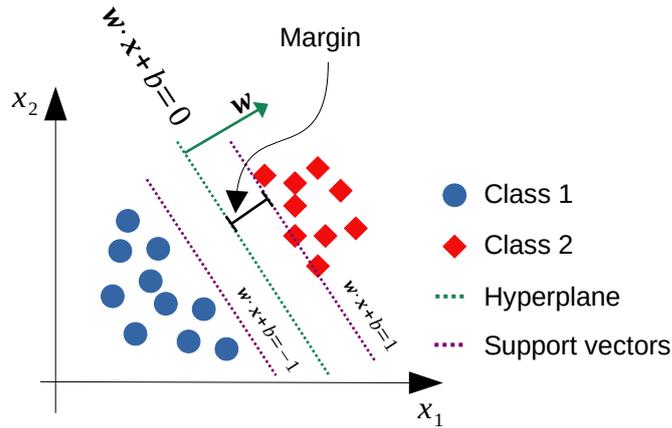
Figure 5.4: Representation of a SVM hyperplane for a bi-class classification problem, assuming only two features. The vector $\mathbf{w}$ is perpendicular to the hyperplane.

where $C > 0$ is a constant that controls the balance between training accuracy and margin width [Guyo 08].

Equation 5.34 illustrates a quadratic programming problem that can be solved by the following Lagrangian function [Bish 06, Guyo 08]:

$$\mathcal{L}(\mathbf{w}, b, \alpha) = \frac{1}{2}||\mathbf{w}||^2 - \sum_{i=1}^{m} \alpha_i\{y^{(i)}(\mathbf{w}^\mathbf{T}\mathbf{x} + b) - 1\}, \tag{5.36}$$

where $\alpha_i \geq 0$ are Lagrange multipliers. This function minimizes with respect to $\mathbf{w}$ and b, but maximized with respect to α[Bish 06]. The minimization of $\mathcal{L}(\mathbf{w}, b, \alpha)$ leads to the following conditions[Bish 06]:

$$\sum_{i=1}^{m} \alpha_i y^{(i)} = 0 \quad \mathbf{w} = \sum_{i=1}^{m} \alpha_i y^{(i)} \mathbf{x}^{(i)} \tag{5.37}$$

With these conditions combined with the Karush-Kuhn-Thucker ones (see [Gord 12] for details), the decision function can be written as follows [Guyo 08]:

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^{m} \alpha_i y^{(i)} \mathbf{x}^T \mathbf{x}_i + b\right) \tag{5.38}$$

One way to learn models efficiently is by the use of the so-called *kernel trick*. It allows to use convex optimization techniques for nonlinear functions of $\mathbf{x}$ [Good 16], and it is explained in the following lines.

Kernel trick: This technique expresses the learning algorithm in terms of dot products between training examples. A transformation of the original space into a feature function $\phi(\mathbf{x})$ is performed. The function $k(\mathbf{x}, \mathbf{x}^{(i)}) = \phi(\mathbf{x}) \cdot \phi(\mathbf{x}^{(i)})$ is called *kernel* [Good 16]. This transforms the feature space into another where linear separation exists. In this way, Equation 5.38 becomes in:

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^{m} \alpha_i y^{(i)} k(\mathbf{x}, \mathbf{x}^{(i)}) + b\right) \tag{5.39}$$

The most popular kernel functions are [Hast 09]:

- d-th Degree polynomial: $k(\mathbf{x}, \mathbf{x}') = \left[1 + \mathbf{x}^T \cdot \mathbf{x}'\right]^d$

- Gaussian[6]: $k(\mathbf{x}, \mathbf{x}') = \exp\left(-\gamma ||\mathbf{x} - \mathbf{x}'||^2\right)$, where γ is the inverse of the standard deviation of the kernel

- Sigmoid: $k(\mathbf{x}, \mathbf{x}') = \tanh\left(\gamma \left[\mathbf{x}^T \cdot \mathbf{x}'\right] + \theta\right)$

In this thesis, the linear approach as well as the RBF and sigmoid kernels were used.

k-Nearest Neighbors (kNN)

kNN is a family of non-parametric techniques intended for classification or regression [Good 16]. kNN classifies a test sample in a n-dimensional space according to the class of the k training samples that are closest to it. These training samples are known as *nearest neighbors*, defined in the set $N_k(\mathbf{x})$. The best k can be estimated via cross-validation [Guyo 08].

Different measures are used to compute how close the neighbors are, such as the Minkowski distance, the Canberra measure, or the Chebyshev function. The Minkowski distance is given by [Guyo 08]:

$$D_M^\alpha(\mathbf{x}, \mathbf{x}') = \sqrt[\alpha]{\sum_{i=1}^{n} |x_i - x_i'|^\alpha} \tag{5.40}$$

In this thesis, D_M^α with $\alpha = 2$ was used, which is the Euclidean distance. Thus, the predicted output can be estimated as follows [Hast 09]:

$$\hat{y}(x) = \frac{1}{k} \sum_{x^{(i)} \in N_k(\mathbf{x})} y^{(i)}, \tag{5.41}$$

where $N_k(\mathbf{x})$ is the neighborhood of x defined by the closest points $x^{(i)}$ [Hast 09]. In this way, the kNN algorithm finds the closest points in the n-dimensional space and averages their outputs.

Trees and Forests

Decision trees (DT) are hierarchical models with tree nodes described by logical conditions based on single features [Guyo 08], in which a sequence of questions are made until a *leaf node* is achieved (see Figure 5.5) [Hart 00]. They are capable to fit into complex datasets [Gero 22], and perform an internal feature selection [Hast 09]. This algorithm has high interpretability (for small feature spaces) and provides information about the relevance of a particular feature, which is advantageous for several applications such as medical diagnosis [Guyo 08]. Actually, after the explosion of deep learning models, DT was the most popular algorithm in data mining [Hast 09]. DT is a recursive process useful to return important features even if they are only relevant in a small region of the feature

---

[6]Also known as Radial Basis Function (RBF)

space [Guyo 08], and they work well in many cases, including those with few data preparation [Gero 22]. A DT model is characterized by the number of nodes, leaves, branches, branch length, tree depth, and class labels, among others [Guyo 08])
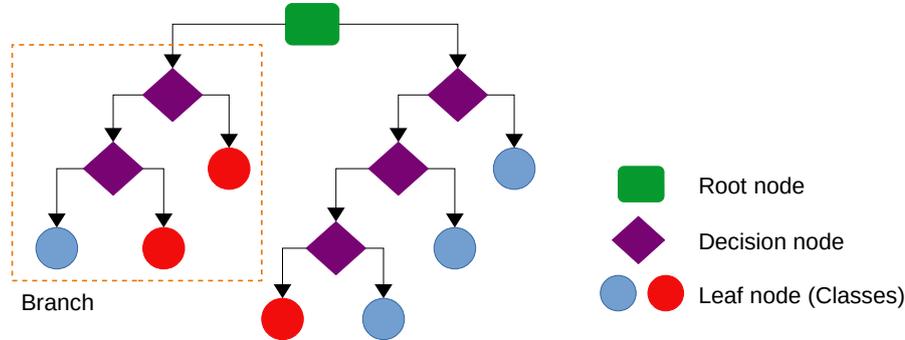
Figure 5.5: Representation of a decision tree.

One key point in the DT is to find the best split [Chen 16b]. The most common split selection criterion is the *impurity*. For a split s and node N, it is given by [Guyo 08, Hast 09]:

$$\Delta I(s, N) = I(N) - \sum_i p_i I\left(N_i^s\right),\qquad(5.42)$$

where I is an impurity measure, N_i^s is the i-th subnode of N resulting from split s, and p_i is the probability to fall into this node. One common impurity measure is the Gini index, given as follows [Guyo 08]:

$$I_G(N) = 1 - \sum_{y \in \mathcal{Y}} \left[p\left(y|N\right)\right]^2\qquad(5.43)$$

where $p\left(y|N\right)$ is the ratio of instances of the class $\mathcal{Y}$ among the training instances in the node N [Gero 22]. For a node with all samples belonging to the class $\mathcal{Y}$, the impurity will be zero. Otherwise, the Shannon entropy and log-loss (see Eq. 5.17) are also used as impurity measures [Guyo 08]. However, they do not provide important differences and Gini impurity is a good default [Gero 22]: even though it tends to isolate the most frequent class in its own branch, it is slightly faster to compute. Thus, the Gini impurity was used in this thesis.

The construction of DT models has been addressed with different methods. The CART [7] method is the most used to train DT, and it is actually the method implemented in the Scikit-Learn library for Python [Gero 22]. The CART method chooses a specific feature x_j and a threshold th_j. Afterwards it searches for the pair (x_j, th_j) that produces the purest nodes [Gero 22]. The cost function after the split is given as follows [Gero 22]:

$$J(x_j, th_j) = \frac{m_{left}}{m} I_G^{left} + \frac{m_{right}}{m} I_G^{right},\qquad(5.44)$$

where I_G^{node} is the impurity of the specific node after the split, i.e. right or left, m_{node} is the number of instances in such node, and m is the number of training samples. This

---

[7]Acronym for Classification And Regression Tree

process is applied recursively until it reaches the maximum depth or if it cannot find a split that reduces the impurity [Gero 22]. The limitations of DT are [Gero 22, Hast 09]: they tend to have orthogonal decision boundaries since they partition the feature space in a set of rectangles, and they are very sensitive to small variations in the training set, so the generalization capability is limited.

On the other hand, a Random Forest classifier (RF) is just an ensemble of DT (see Figure 5.6), and it is capable to handle multi-class problems directly [Gero 22]. It applies many DT and obtains the class predicted by all of them. Finally, the predicted class is such that gets the majority of votes, therefore this procedure is known as *majority voting* [Gero 22]. The voting scheme can be *hard* (when the class is assigned to the most voted by the *weak* classifiers), or *soft* (predictions are made as a vote by the trees in the forest, based on their weighted probabilities [Hart 00], i.e. the class with the highest probability across the trees is assigned). Soft voting was implemented in this thesis. RF is usually performed using the same tree with different -and random- subsets of the training set, such as one sample can appear many times in the subsets for each classifier, i.e. with replacement. This ensemble method is known as *bootstrap aggregating*, or simply *bagging* [Gero 22]. Even though RF can be trained without replacement (known as pasting), in this thesis the bagging method was implemented since it is the most preferred for RF.
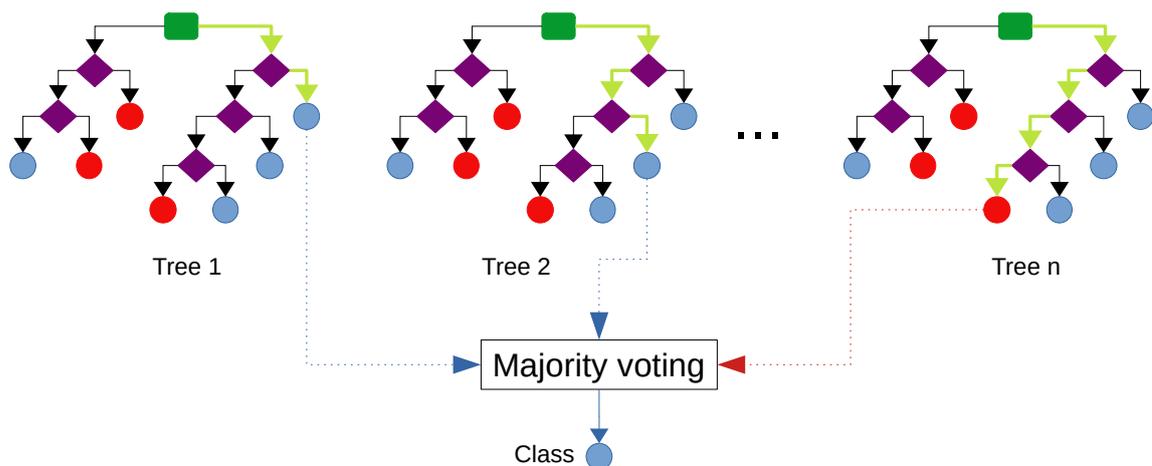
Figure 5.6: Representation of a random forest with decision trees. The final predicted class is obtained by -hard- majority voting.

Finally, another algorithm based on decision trees is known as Extreme Gradient Boosting (XGBoost). It is a computationally efficient and scalable for classification. It has shown very good very good performance in different famous machine learning and data mining challenges during the last decade [Chen 16b]. This method is based on the *Gradient Boosting* algorithm, which adds sequential predictors to an ensemble, each one correcting its predecessor trying to fit the new predictor to the residual errors made by the predecessor [Gero 22]. Most existing implementations of the XGBoost, e.g. Scikit-Learn for Python (used in this thesis) and GBM for R, are based on the *exact greedy algorithm* [Chen 16b], which splits until it achieves the highest score using the first and second order gradient statistics on the loss function. See [Chen 16b] for further details.

5.2.2.   Gated-Multimodal Units (GMU)

Deep Learning (DL) models are the state-of-the-art in many applications such as image classification, speech recognition, generative art, etc [LeCu 15]. When an ANN has two or more hidden layers, it is called *Deep Neural Network* (DNN) [Gero 22]. Thus, DL is based on DNN with different configurations or architectures, such as convolutional neural networks (CNN) and recurrent/recursive neural networks (RNN) [Good 16]. CNNs are inspired in the organization of the cat's visual cortex [Miot 18]; they are intended mostly for image analysis by the use of convolutional filters, they are typically composed by more than ten convolutional and pooling layers, and require large datasets to be trained [Miot 18]. Otherwise, RNNs are specialized in sequence data and are composed of one network that performs the same task for every element of the sequence, while keeping dependencies on the previous computations [Miot 18], i.e. they have memory. However, they are not intended when the interpretability of a phenomenon is a paramount point, such as in swallowing and dysphagia. DL overcomes the performance of ML algorithms only for big datasets due to their data-driven orientation [Mukh 21, LeCu 15] (see Figure 5.7).
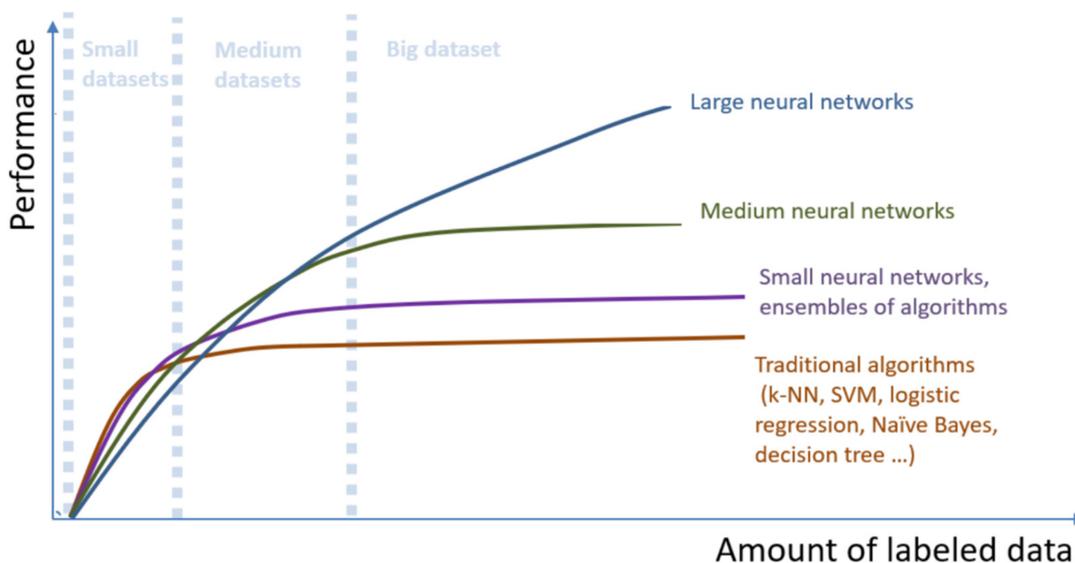
Figure 5.7: Performance of different machine learning algorithms depending of the amount of available data. Taken from [Mukh 21].

For cases where there are more than one source of information, i.e., two or more modalities, a different kind of architecture for multi-modal problems based on RNN was proposed, namely *Gated Multimodal Units* (GMU) [Arev 20]. Like other architectures such as LSTM [8] or GRU[9], GMUs are composed by gated recurrent networks. GMU combines multiple inputs in a single operation, i.e. for information fusion, rather than to assume temporal dependencies among data [Arev 20]. GMU combines the concepts of *early fusion* (feature sets of multiple sources of information are combined to feed a classification algorithm) and *late fusion* (combines the output of different classifiers, one per modality,

---

[8]Long-Short Term Memory
[9]Gated Recurrent Units

and makes a prediction by consensus, e.g. majority voting) [Arev 17]. Figure 5.8 illustrates the GMU architecture for a bi-modal approach and for three or more modalities.

(a) The model to use with more than two modalities.

(b) A simplification for the bimodal approach.
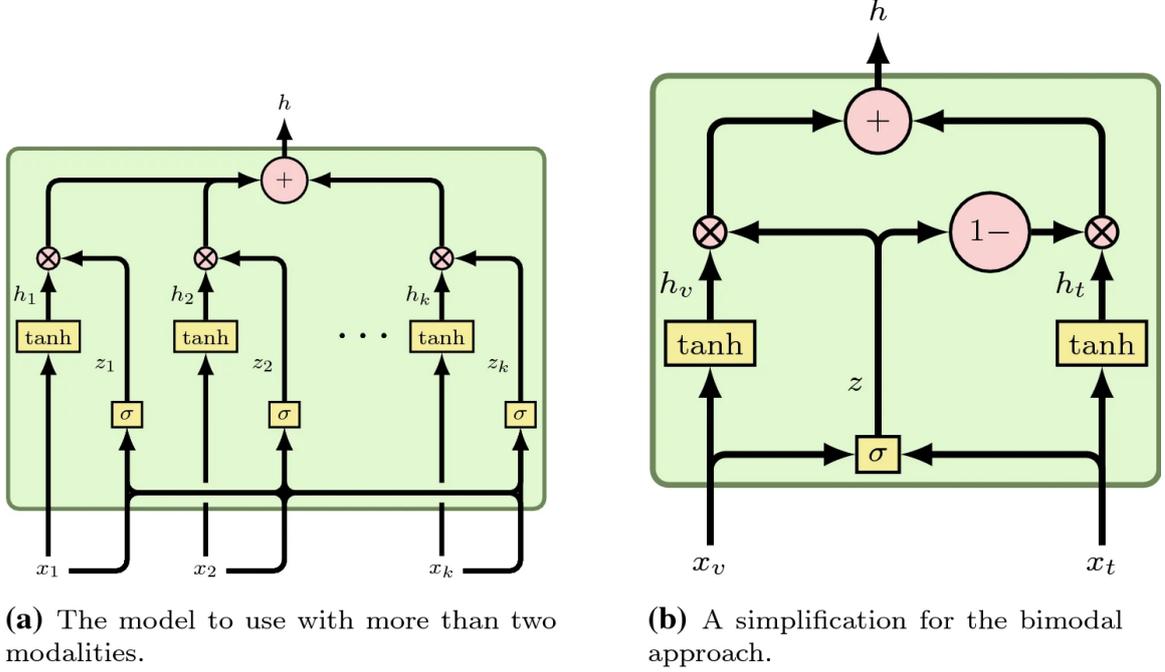
Figure 5.8: Scheme of GMU. a) Architecture for three or more modalities. b) Bi-modal approach. Taken from [Arev 17].

Let $\mathbf{x}_k \in \mathbb{R}^{n_k}$ a column vector of the k-th modality. For the bimodal approach, the GMU extracts hidden representations for each modality as follows [Arev 20]:

$$h_1 = \tanh\left(\mathbf{w}_1 \mathbf{x}_1^T\right) \tag{5.45}$$
$$h_2 = \tanh\left(\mathbf{w}_2 \mathbf{x}_2^T\right), \tag{5.46}$$

where $\mathbf{w}_k$ is the -learnable- weight vector for the k modality. An additional parameter is computed as follows:

$$z = \sigma\left(\mathbf{w}_z \left[\mathbf{x}_1, \mathbf{x}_2\right]^T\right), \tag{5.47}$$

where $[.,.]$ is the concatenation operator and $\mathbf{w}_z \in \mathbb{R}^{n_1+n_2}$. The output activation is given by [Arev 20]:

$$h = z \odot h_1 + (1 - z) \odot h_2 \tag{5.48}$$

Thus, the GMU tunes how each modality affects the output. When more than two modalities are available, the internal connections of the GMU differ:

$$h_k = \tanh\left(\mathbf{w}_k \mathbf{x}_k^T\right) \tag{5.49}$$

$$z = \sigma\left(\mathbf{w}_z\left[\mathbf{x}_1, \ldots, \mathbf{x}_{k_{max}}\right]^T\right) \tag{5.50}$$

$$h = \sum_{k=1}^{k_{max}} z_k \odot h_k \tag{5.51}$$

In this thesis, the following combinations of modalities were applied: {sEMG, Speech}, {sEMG, Acc}, {Acc, Speech}, and {sEMG, Acc, Speech}. The binary cross-entropy was used as a loss function and the Adam optimization was also implemented. The GMU-based models were trained using the validation methods described in Section 5.3. Ten epochs were used to train the models.

5.2.3.   Hyperparameters tuning of classification algorithms

The hyperparameters of each classifier were tuned in a grid search, in order to control over-fitting and to improve the classification performance. The lines below illustrate parameters range for each classification strategy.

- SVM: $C \in \{10^{-4}, 10^{-3}, ..., 10^3, 10^4\}$; $\gamma \in \{10^{-4}, 10^{-3}, ..., 10^3, 10^4\}$; kernels: linear, RBF and sigmoid (with $\theta = 0$).

- ANN: hidden layer size $\in \{10, (10, 10), (10, 50), 50, (50, 10), (50, 50), (50, 100), 100, (100, 50), (100, 100)\}$; $\alpha \in \{10^{-4}, 10^{-3}, ..., 10^3, 10^4\}$; activation functions: ReLU, tanh, logistic; and solvers: Adam and SGD.

- RF: number of trees $N \in \{5, 10, 20, 30, 50, 100\}$; and maximum depth $D \in \{2, 5, 10, [0]20, 30, 50, 100\}$

- Logistic regression: penalty parameter $C \in \{10^{-4}, 10^{-3}, ..., 10^3, 10^4\}$; and regularization norms: L1, L2.

- LDA: solvers $\in \{$singular value decomposition, least squares solution$\}$

- DT: maximum depth $D \in \{2, 5, 10, 20, 30, 50, 100\}$

- XGBoost: maximum depth $D \in \{2, 5, 10, 20, 30, 50, 100\}$; and ratio of negative/positive classes $R \in \{1, 10, 25, 50, 75, 99, 100, 1000\}$.

- GMU: hidden layer size $\in \{64, 128, 256, 512\}$; learning rate $\in \{10^{-3}, 10^{-2}, 10^{-1}\}$.

5.3.   Model validation

One of the most used methods for model selection and hyperparameter optimization is the minimization of the cross-validation (CV) estimate of the generalization performance [Wain 21]. The *flat* k-fold CV splits the database into k chunks of approximately the same size; it takes $k - 1$ chunks, namely training sets, to fit the model and optimize the hyperparameters. The resulting model is tested with the remaining chunk, and the

classification error is obtained. Thus, k iterations are performed, and the generalization performance is the average of the k obtained performances, one per chunk [Wain 21]. Notwithstanding, this procedure introduces bias and could be very optimistic for small-size databases [Vaba 19]. To overcome this limitation, the nested CV was introduced. The nested CV is similar to the *flat* one, with an additional iteration level, i.e. a nested CV with k_{inner} chunks is performed inside the main CV with k_{outer} chunks. In the outer chunks, the hyperparameters are tuned independently and used to train the model with the inner chunks. So, the outer CV minimizes the inner CV estimate of generalization performance [Wain 21], and whilst inner loops are used for hyperparameter optimization and model training, outer loops are used for error estimation. The nested approach produces an almost unbiased and robust estimate of the performance [Vaba 19].

The CV is often applied under a *stratified* scheme, an approach that guarantees in each partition almost the same class ratio of the entire database [Berr 19]. For the case of this thesis it refers to the same proportion of healthy and dysphagic subjects in each partition. It aims to provide a sample proportion as an unbiased estimate of the population one [Berr 19]. For stratified CV, $k = 10$ partitions are recommended [Berr 19].

Consequently, two stratified nested CV (snCV) schemes for model validation were applied, with $k_{inner} = 5$ and $k_{outer} = 10$. For experiments #1, #2, and #3, no test sets were implemented and the entire database was divided into training and validation sets (see Figure 5.9).
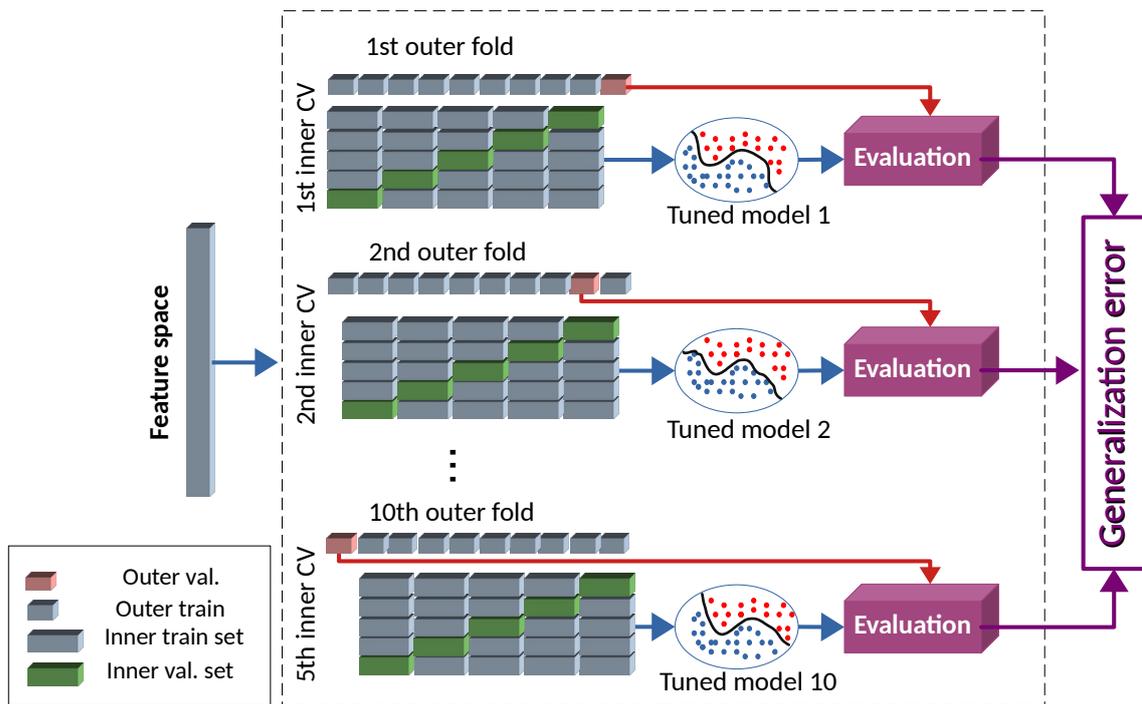
Figure 5.9: Stratified nested cross-validation without test set for Experiments #1, #2 and #3.

For the case of experiment #4, a separate test set was considered. Thus the database was divided into a train-validation (train-val) and a test set. The train-val set was the input of the snCV. The hyperparameters of the ten tuned models of the outer partitions were

retrieved. A new model with the mode of such hyperparameters was tested with the test set, and the generalization performance, not the average of the partitions, was obtained (see Figure 5.10). No volunteer of the test set was used in the train-val stage. Different proportions of test/train-val were evaluated to figure out how sensible was the model to the partition size: 10%/90%, 15%/85%, 20%/80%, 25%/75%, 30%/70%, 35%/65%, and 40%/60%. Each test set was selected randomly. This procedure was repeated five times with different -and random- composition of each test, aiming to avoid biased results.
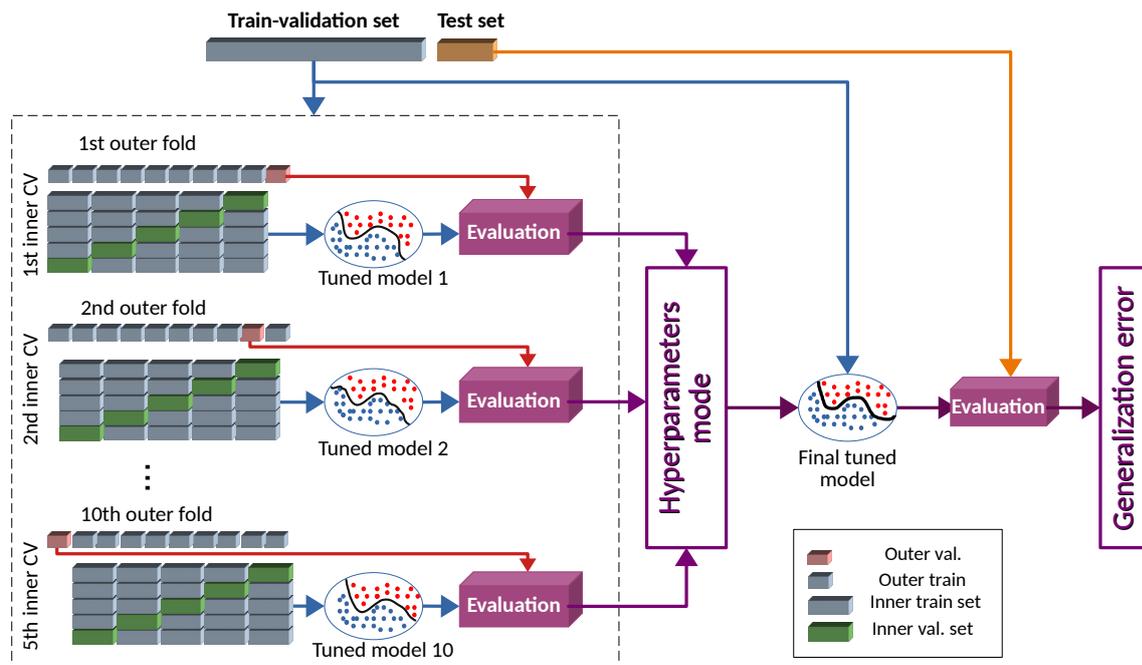
Figure 5.10: Stratified nested cross-validation with a test set for experiment #4.

5.3.1.   Performance measures

The classifiers were evaluated by the use of different performance measures that provide specific information about the capability of discrimination between the two populations assessed. Each measure is described next, using the definitions of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) provided in Section 3.4. The following measures are recommended to be reported not only in machine learning-related works but also in clinical diagnostic accuracy studies [Cohe 16].

- AUC: Equilibrium between TPR and FPR. Described in Section 3.4. Even though it was used also as a feature selection method, in the validation stage the AUC was used to describe the discrimination capability of the classifier instead of the feature space[10]. This was the selection criterion for hyperparameters tuning in the snCV schemes.

---

[10]To avoid confusion in this document, AUC_{ROC} denotes the feature selection method, and AUC denotes the performance measure of the classifier.

- Accuracy: a measure of the numbers of individuals that were correctly classified, defined as $acc = \frac{TP+TN}{TP+TN+FP+FN}$. This is not a good measure for unbalanced databases.

- Precision: also called as *confidence* in data mining is a measure of predicted positives that are actually positives [Powe 11]. Computed as $precision = \frac{TP}{TP+FP}$

- Sensitivity: also called *recall* in machine learning, is a measure of the real positives that were classified as positives [Powe 11]. Computed as $se = \frac{TP}{TP+FN}$

- Specificity: also called as *inverse recall* in machine learning, is a measure of the real negatives that were classified as negatives [Powe 11]. Computed as $sp = \frac{TN}{TN+FP}$

- F_1-score: it is the harmonic mean of the sensitivity and precision, computed as $F_1 = 2 \times \frac{se \times precision}{se+precision}$

Chapter 6

Classification experiments

As mentioned in Section 3.4, the experiments were conducted in different stages of the research and with different databases, so different combinations of classifiers were performed for each experiment. As experiments were carried out, the best methods in terms of interpretability and simplicity were selected. Table 6.1 summarizes the machine learning algorithms used for each experiment.

Table 6.1: Summary of feature selection methods and classification algorithms used for each experiment.

Experiment	SVM	XGBoost	kNN	ANN	RF	LR	LDA	DT	GMU
Experiment #1	●	●		●					
Experiment #2	●	●	●	●					
Experiment #3	●			●	●	●	●	●	
Experiment #4	●	●	●						●

6.1. Experiment #1: Automatic detection of dysphagia using electrophysiological biomarkers

In this experiment, the discrimination capability of classification algorithms for dysphagia screening was evaluated. The following hypothesis was proposed: there are differences between sEMG recordings of healthy individuals and patients with dysphagia, that can be detected in a new representation domain, i.e. a feature space.

Figure 6.1 shows an example of a healthy and a dysphagic sEMG recordings during dry swallowing (saliva). This differential behavior of piecemeal deglutition in patients using different consistencies was evidenced in several subjects, which motivates the use of machine learning models for identifying dysphagia-related patterns.

6.1.1. Classification using individual features

Table 6.2 shows the classifiers that achieved AUC≥0.75 when considering individual features. Note that such results are in the line of those presented in the section 4.1: time

Figure 6.1: Example of multi-channel sEMG generated from one dysphagic patient and one healthy control. The amplitude is given in micro-volts. RM and LM: right and left masseter; RSH and LSH: right and left suprahyoid; RIH and LIH: right and left infrahyoid muscles.

domain features achieved the highest discrimination capability between healthy and dysphagic sEMG recordings.

Table 6.2: Classification results obtained with individual features. Only classifiers with AUC≥0.75 are shown. Values≥0.80 are highlighted in bold.

Feature	Task	Classifier	AUC	F1	Accuracy	Precision	Sensitivity
FR	W20	SVM	0.76 ± 0.16	0.74 ± 0.18	0.74 ± 0.16	0.82 ± 0.18	0.75 ± 0.27
iEMG	W5	SVM	0.77 ± 0.19	0.76 ± 0.21	0.76 ± 0.20	0.83 ± 0.22	0.77 ± 0.26
LOG	**Saliva**	**SVM**	$\mathbf{0.82 \pm 0.17}$	$\mathbf{0.79 \pm 0.19}$	$\mathbf{0.81 \pm 0.17}$	$\mathbf{0.88 \pm 0.18}$	$\mathbf{0.75 \pm 0.23}$
		ANN	$\mathbf{0.80 \pm 0.16}$	$\mathbf{0.76 \pm 0.20}$	$\mathbf{0.79 \pm 0.16}$	$\mathbf{0.92 \pm 0.17}$	$\mathbf{0.68 \pm 0.24}$
	W5	ANN	0.77 ± 0.18	0.77 ± 0.17	0.76 ± 0.18	0.84 ± 0.21	0.73 ± 0.20
		XGBoost	0.77 ± 0.15	0.80 ± 0.16	0.78 ± 0.15	0.77 ± 0.17	0.87 ± 0.22
	Y5	ANN	0.76 ± 0.18	0.76 ± 0.19	0.75 ± 0.20	0.84 ± 0.22	0.77 ± 0.26
		SVM	0.75 ± 0.17	0.77 ± 0.16	0.74 ± 0.19	0.81 ± 0.22	0.83 ± 0.22
RMS	W10	ANN	0.76 ± 0.15	0.77 ± 0.16	0.76 ± 0.17	0.82 ± 0.20	0.82 ± 0.24

Wx: x mL of water; Y5: 5 mL of yogurt.

This evaluation also provides information related to swallowing tasks and classifiers. SVM and ANN were the most accurate in saliva recordings using the log detector (LOG). Their precision was close to 0.90 in both cases. For water$_5$ and yogurt$_5$, good results were also obtained with both classifiers. Thus, the capability of discrimination seems to be more related to the feature space and swallowing task rather than the classifier, indicating that those features seem to be good candidates to create a set of promising biomarkers to evaluate dysphagia. Another pattern that was observed is that the smaller the volume, the higher the accuracy.

6.1.2. Classification per muscle group

Different feature sets were considered in this scenario: the entire feature set, features selected with PCA, mRMR, and AUC_{ROC}, as well as features grouped by domain. This scenario was computed separately by swallowing task and the evaluation intended to determine if there are muscle groups with better discrimination capability for dysphagia evaluation than others. Table 6.3 shows the best classification results (only AUC$\geq$0.80). The column named "Feature set" refers to the different groups that were created to perform the experiments.

It is remarkable that masseter and suprahyoid muscles yielded good results, in contrast to infrahyoid that achieves AUC values smaller than 0.80 (not shown in the table). Another relevant finding is that water achieved the highest performance, specifically 20 mL for masseter and 10 mL for suprahyoid (AUC$\geq$0.85 - Table 6.3). Actually, these tasks were the most prevalent with *good* results in terms of AUC. Otherwise, saliva in suprahyoid muscles and yogurt$_{20}$ in masseter achieved AUC 0.80$\leq$AUC$\leq$0.84.

Table 6.3: Classification results obtained for different muscle groups, feature sets, and classifiers. Only cases with AUC$\geq$0.80 are shown. Classifiers with AUC$\geq$0.85 are highlighted in bold.

Muscles	Task	Classifier	Feature set	AUC	F1	Accuracy	Precision	Sensitivity
Masseter	W5	SVM	PCA	0.81 ± 0.17	0.83 ± 0.12	0.80 ± 0.16	0.88 ± 0.19	0.83 ± 0.17
	W10	ANN	None	0.81 ± 0.15	0.82 ± 0.15	0.82 ± 0.14	0.88 ± 0.16	0.85 ± 0.24
		SVM	Time-Freq	0.82 ± 0.12	0.82 ± 0.16	0.84 ± 0.12	0.92 ± 0.11	0.80 ± 0.26
	W20	ANN	**NLD**	$\mathbf{0.85 \pm 0.17}$	$\mathbf{0.87 \pm 0.13}$	$\mathbf{0.85 \pm 0.16}$	$\mathbf{0.83 \pm 0.17}$	$\mathbf{0.93 \pm 0.13}$
		SVM	**AUC_{ROC}**	$\mathbf{0.88 \pm 0.14}$	$\mathbf{0.88 \pm 0.13}$	$\mathbf{0.88 \pm 0.14}$	$\mathbf{0.91 \pm 0.14}$	$\mathbf{0.88 \pm 0.18}$
			None	0.80 ± 0.16	0.80 ± 0.15	0.78 ± 0.18	0.87 ± 0.21	0.80 ± 0.21
	Y20	ANN	NLD	0.83 ± 0.26	0.85 ± 0.22	0.84 ± 0.25	0.90 ± 0.21	0.82 ± 0.24
			Time-Freq	0.84 ± 0.13	0.85 ± 0.14	0.85 ± 0.12	0.88 ± 0.17	0.85 ± 0.19
Suprahyoid	Saliva	ANN	PCA	0.83 ± 0.19	0.85 ± 0.14	0.84 ± 0.17	0.92 ± 0.16	0.82 ± 0.19
		XGBoost	AUC_{ROC}	0.82 ± 0.13	0.85 ± 0.12	0.83 ± 0.12	0.85 ± 0.15	0.88 ± 0.18
	W5	XGBoost	Freq	0.83 ± 0.12	0.86 ± 0.14	0.85 ± 0.11	0.84 ± 0.13	0.93 ± 0.20
			mRMR	0.83 ± 0.15	0.86 ± 0.16	0.85 ± 0.14	0.87 ± 0.14	0.90 ± 0.21
	W10	ANN	AUC_{ROC}	0.82 ± 0.19	0.77 ± 0.29	0.80 ± 0.22	0.82 ± 0.33	0.78 ± 0.32
			mRMR	$\mathbf{0.86 \pm 0.10}$	$\mathbf{0.86 \pm 0.11}$	$\mathbf{0.86 \pm 0.09}$	$\mathbf{0.95 \pm 0.10}$	$\mathbf{0.82 \pm 0.19}$
		SVM	**AUC_{ROC}**	$\mathbf{0.86 \pm 0.17}$	$\mathbf{0.87 \pm 0.16}$	$\mathbf{0.86 \pm 0.16}$	$\mathbf{0.94 \pm 0.13}$	$\mathbf{0.87 \pm 0.22}$
			mRMR	0.82 ± 0.15	0.83 ± 0.13	0.82 ± 0.14	0.91 ± 0.14	0.78 ± 0.18
			None	0.81 ± 0.16	0.83 ± 0.14	0.82 ± 0.14	0.85 ± 0.19	0.85 ± 0.19
			Time	0.82 ± 0.16	0.85 ± 0.11	0.82 ± 0.14	0.84 ± 0.17	0.90 ± 0.15
		XGBoost	Time	0.82 ± 0.14	0.84 ± 0.12	0.81 ± 0.14	0.82 ± 0.18	0.90 ± 0.15
	W20	XGBoost	Time	0.81 ± 0.18	0.82 ± 0.19	0.82 ± 0.19	0.82 ± 0.20	0.88 ± 0.24
	Y20	ANN	mRMR	0.80 ± 0.16	0.81 ± 0.14	0.79 ± 0.16	0.85 ± 0.19	0.80 ± 0.16

Wx: x mL of water; Y20: yogurt$_{20}$.

None of the feature selection techniques or feature sets significantly outperformed the others, so there were no conclusive results in this regard. The AUC_{ROC} selection approach was used with the SVM classifier in two of the four instances where AUC$\geq$0.85 (highlighted in bold in Table 6.3). The other two cases featured ANN models, one of which had features chosen with mRMR (suprahyoid, water$_{10}$), and the other of which had features

chosen using only nonlinear features (masseter, water$_{20}$). Despite Table 6.3 displays four classifiers with features selected by mRMR or AUC$_{ROC}$, and three with the entire feature space or with all time domain features, it is not possible to conclude that this result is a predictor of which feature space should be used. Similarly, SVM and ANN seem to be the most suitable, and XGBoost only exhibited AUC$\geq$0.80 in suprahyoid muscles.

On the other hand, some classifiers had sensitivity or precision values higher than 0.90. The highest sensitivity was only attained with water intakes, even when the classifiers with the highest precision showed no pattern in terms of the muscle group, feature set, or swallowing task. In contrast, the XGBoost produced a sensitivity of 0.90 in three instances; all of them involved suprahyoid muscles and water, despite not being the most accurate classifier in these studies.

6.1.3. Classification per swallowing task

Features extracted from the three muscle groups (six channels) were considered simultaneously to assess the discrimination capability of swallowing tasks. Table 6.4 summarizes the classification outcomes with AUC$\geq$0.80. SVM had much fewer occurrences, whilst the XGBoost and ANN appeared most frequently. The three results highlighted in bold were the ones with AUC$\geq$0.85: saliva with time-related features, and water$_{10}$ with ANN considering two feature sets: without selection and time domain features. The most frequent feature sets were those selected with AUC$_{ROC}$ and mRMR; actually, all swallowing tasks with AUC$\geq$0.80 involved at least one of these selection methods. No improvement in classification accuracy was observed with the combination of swallowing tasks.

Table 6.4: Classification results obtained with all muscle groups. Only classifiers with AUC$\geq$0.80 are shown. Classifiers with AUC$\geq$0.85 are highlighted in bold.

Task	Feature set	Classifier	AUC	F1	Accuracy	Precision	Sensitivity
Saliva	AUC$_{ROC}$	XGBoost	0.82 ± 0.13	0.87 ± 0.09	0.83 ± 0.12	0.82 ± 0.16	0.97 ± 0.10
	mRMR	XGBoost	0.81 ± 0.15	0.82 ± 0.13	0.81 ± 0.15	0.89 ± 0.16	0.78 ± 0.18
	PCA	XGBoost	0.82 ± 0.13	0.84 ± 0.14	0.83 ± 0.12	0.83 ± 0.17	0.88 ± 0.17
	Time	**ANN**	$\mathbf{0.86 \pm 0.10}$	$\mathbf{0.83 \pm 0.11}$	$\mathbf{0.85 \pm 0.10}$	$\mathbf{0.97 \pm 0.10}$	$\mathbf{0.75 \pm 0.16}$
		SVM	0.81 ± 0.13	0.80 ± 0.13	0.81 ± 0.12	0.91 ± 0.14	0.78 ± 0.23
W5	AUC$_{ROC}$	ANN	0.81 ± 0.21	0.82 ± 0.20	0.81 ± 0.20	0.86 ± 0.22	0.85 ± 0.24
	Freq	XGBoost	0.81 ± 0.18	0.86 ± 0.17	0.83 ± 0.16	0.81 ± 0.16	0.93 ± 0.19
	mRMR	ANN	0.81 ± 0.20	0.73 ± 0.32	0.80 ± 0.23	0.85 ± 0.32	0.72 ± 0.37
		XGBoost	0.83 ± 0.18	0.83 ± 0.18	0.83 ± 0.17	0.90 ± 0.15	0.80 ± 0.22
W10	AUC$_{ROC}$	XGBoost	0.82 ± 0.14	0.84 ± 0.12	0.82 ± 0.14	0.80 ± 0.18	0.93 ± 0.13
	mRMR	ANN	0.82 ± 0.21	0.85 ± 0.16	0.82 ± 0.21	0.87 ± 0.21	0.88 ± 0.18
		XGBoost	0.83 ± 0.13	0.86 ± 0.11	0.84 ± 0.12	0.82 ± 0.15	0.93 ± 0.13
	None	**ANN**	$\mathbf{0.86 \pm 0.16}$	$\mathbf{0.84 \pm 0.17}$	$\mathbf{0.84 \pm 0.17}$	$\mathbf{0.95 \pm 0.14}$	$\mathbf{0.77 \pm 0.20}$
	PCA	ANN	0.84 ± 0.16	0.84 ± 0.14	0.82 ± 0.18	0.87 ± 0.20	0.87 ± 0.16
	Time	**ANN**	$\mathbf{0.86 \pm 0.14}$	$\mathbf{0.86 \pm 0.13}$	$\mathbf{0.84 \pm 0.17}$	$\mathbf{0.88 \pm 0.19}$	$\mathbf{0.90 \pm 0.15}$
Y10	mRMR	XGBoost	0.82 ± 0.18	0.84 ± 0.18	0.83 ± 0.18	0.86 ± 0.18	0.87 ± 0.22
Cracker	AUC$_{ROC}$	ANN	0.82 ± 0.17	0.80 ± 0.29	0.84 ± 0.15	0.76 ± 0.29	0.87 ± 0.31
All tasks	None	ANN	0.82 ± 0.13	0.79 ± 0.16	0.82 ± 0.14	0.88 ± 0.18	0.77 ± 0.21

Wx: x mL of water; Y10: yogurt$_{10}$.

6.1.4.   Analysis of discrimination capability of sEMG

This is the first study that uses ML algorithms on multichannel sEMG to classify individuals as healthy or dysphagic. Despite other works have used ML models in swallowing evaluation, they have been mainly oriented to swallow detection [Suzu 20, Cons 17, Schu 14], as well as onset detection and segmentation [Rieb 19, McNu 21]. Other works are more descriptive in terms of how some features, e.g. duration and amplitude, vary according to the type of bolus [Watt 15], or age [Wang 15]. This point is highly related to the novelty of this experiment.

Contribution of individual features

Section 4.1.2 indicated that the most suitable electrophysiological biomarkers are VAR, RMS, iEMG, LOG, WL, DASDV, WAMP, MYOP, TKEO, and MNP. The classification results confirmed this: the individual features that achieved the highest performance were also in the set of features selected by the AUC_{ROC}: iEMG, LOG and RMS (see Table 6.2).

Classification results obtained with individual features can be compared only with two works: [Miya 20] used a SVM applied on swallowing sounds in 27 healthy young controls and 143 patients with dysphagia, achieving modest results (F1 = 78.9%, accuracy = 77%, precision = 73.7%, and recall = 87%). Otherwise, [Dono 21b] used cervical accelerometry and sounds to differentiate between swallows from 51 healthy people and 20 patients with suspected dysphagia. They achieved impressive classification results with classical ML algorithms: accuracy and specificity of 99%, and sensitivity of 100%, but the results could be biased and optimistic because of the unbalanced database.

Contribution of muscle groups and swallowing tasks

When the classifiers were fed with features grouped by muscles instead of individual features, the performance improved. While the highest AUC value obtained with individual features was 0.82 and there were only two cases with AUC≥0.80, the classification on muscle groups returned 21 cases with AUC≥0.80, and four cases with AUC≥0.85. This behavior was expected because the complex nature of the swallowing process should not be well explained by only one single feature.

Bearing in mind that infrahyoid muscles participate in all swallowing phases [Li 17], one could expect that such muscles contribute to the classification performance. Surprisingly, in contrast to masseters and suprahyoid muscles, infrahyoid ones did not retrieve *good* results in terms of AUC. Furthermore, the classification scenario with all muscle groups simultaneously did not outperform the results obtained with separated muscles in any of the evaluated performance measures (see Tables 6.3 and 6.4). However, Figure 4.2 shows some relevance of time-domain features extracted from infrahyoid muscles, mainly for the intake of 5 and 10 mL of liquids (water and yogurt). Therefore, it is not convenient to neglect the contribution of such muscles in the analysis of the swallowing process.

The intake of water was the most frequent task with *good* performance in the classification scenario per muscle group. Water tests are commonly used and have demonstrated their suitability in clinical practice for bedside examination in dysphagia [Carn 08]. Also, saliva-related tasks retrieved *good* performance in two cases with suprahyoid muscles. Subsequently, saliva and water should be included in an automatic evaluation protocol

based on sEMG, because they have shown differential activation patterns in literature: the duration of the suprahyoid activity has been reported as longer for saliva than for water swallows (5 and 10 mL) [Perl 99]. The saliva-related pattern is quite different and requires more muscle activity. Thus, saliva swallowing is shorter in duration than water, but higher in amplitude [Enge 12, Zhu 17]. This is in line with the results per muscle group, in which saliva retrieved good discrimination capability in suprahyoid but not in masseter (Table 6.3). Furthermore, water$_5$ retrieved high sensitivity values in suprahyoid (≥ 0.90). Such muscles have also shown high statistical discrimination capability between patients with dysphagia with multiple sclerosis and healthy controls [Alfo 13]. It was also observed in this work that saliva and water could be well modeled with the proposed approach, especially when using ANN or SVM as classifiers (Table 6.2).

On the other hand, yogurt retrieved good performance for 20 mL in masseter and suprahyoid, as well as 10 mL using all muscle groups. Variations in liquid thickness have shown also differential activation patterns [Enge 12]. The intake of water requires less muscle activity and has a shorter duration than solids or thick liquids. Also, the contraction strength increases when the liquid volume or density increases too [Zhu 17]. Thus, this task should not be disregarded for the automatic screening protocol.

The classification performance using the combination of muscle groups confirmed many observations found with individual groups: water$_5$ and water$_{10}$ were the most frequent swallowing tasks with AUC≥ 0.80; saliva also retrieved similar results. The occurrence of good results with crackers is marginal. The best classification performances were obtained also with time-domain features. Additionally, the performance was found to be dependent on the consistency and on the feature domain but not on the feature selection method or classifier.

Finally, despite the AUC was the criterion for hyperparameter tuning, sensitivity plays also a key role in the development of diagnostic tests. In this case, scenarios of separated muscle groups and their combination showed some cases with high sensitivity (≥ 0.90, in Tables 6.3 and 6.4). This suggests that the obtained modeling schemes are suitable to the development of automatic screening tests.

6.2. Experiment #2: Automatic detection of dysphagia using electrophysiological and mechanical biomarkers

The aim of this experiment was to study the potential use of multi-modal information based on multi-channel sEMG and tri-axial Acc signals for the automatic classification of individuals with normal swallowing and patients with functional oropharyngeal dysphagia. Besides, the capability of discrimination of each muscle group and acceleration axis was also studied.

6.2.1. Uni-modal and bimodal classification scenarios

One feature space was created per type of signal and swallowing task. Additionally, the early fusion of features was considered through the concatenation of feature spaces produced by each type of signal individually. In this way, each consistency was evaluated under three scenarios (sEMG, Acc, sEMG+Acc), as well as four classifiers (SVM, ANN,

XG Boost, and kNN), for a total of 84 schemes, following the same methodology as described in Figure 5.9. Due to space limitations, only the best results per swallowing task and scenario are shown in Table 6.5. The most remarkable finding is that the fusion of signals retrieved the highest classification performance in all swallowing tasks, which suggests that sEMG and Acc information are complementary. When evaluated individually, the Acc signals yielded better results than sEMG in all tasks.

Table 6.5: Summary of the best classification results achieved by swallowing task and signal-related scenario. Best results are highlighted in bold.

Task	Scenario	Classifier	AUC	F1	Accuracy	Precision	Sensitivity
	sEMG	XGBoost	0.73 ± 0.11	0.72 ± 0.13	0.72 ± 0.11	0.74 ± 0.19	0.78 ± 0.24
W5	Acc	XGBoost	0.78 ± 0.11	0.77 ± 0.14	0.79 ± 0.11	0.81 ± 0.17	0.77 ± 0.20
	Acc+sEMG	**SVM**	$\mathbf{0.86 \pm 0.15}$	$\mathbf{0.80 \pm 0.29}$	$\mathbf{0.86 \pm 0.15}$	$\mathbf{0.85 \pm 0.30}$	$\mathbf{0.78 \pm 0.32}$
	sEMG	XGBoost	0.71 ± 0.16	0.69 ± 0.18	0.71 ± 0.16	0.72 ± 0.21	0.73 ± 0.24
W10	Acc	XGBoost	0.83 ± 0.07	0.82 ± 0.08	0.82 ± 0.08	0.78 ± 0.15	0.90 ± 0.15
	Acc+sEMG	**XGBoost**	$\mathbf{0.83 \pm 0.07}$	$\mathbf{0.82 \pm 0.08}$	$\mathbf{0.81 \pm 0.09}$	$\mathbf{0.78 \pm 0.16}$	$\mathbf{0.93 \pm 0.13}$
	sEMG	XGBoost	0.71 ± 0.16	0.71 ± 0.13	0.69 ± 0.17	0.66 ± 0.19	0.87 ± 0.21
W20	Acc	kNN	0.78 ± 0.26	0.69 ± 0.37	0.80 ± 0.24	0.80 ± 0.40	0.63 ± 0.37
	Acc+sEMG	**kNN**	$\mathbf{0.82 \pm 0.21}$	$\mathbf{0.73 \pm 0.38}$	$\mathbf{0.84 \pm 0.18}$	$\mathbf{0.73 \pm 0.39}$	$\mathbf{0.73 \pm 0.39}$
	sEMG	SVM	0.72 ± 0.17	0.68 ± 0.20	0.72 ± 0.17	0.78 ± 0.20	0.67 ± 0.26
S	Acc	SVM	0.82 ± 0.12	0.79 ± 0.15	0.82 ± 0.12	0.88 ± 0.15	0.77 ± 0.21
	Acc+sEMG	**SVM**	$\mathbf{0.83 \pm 0.13}$	$\mathbf{0.82 \pm 0.15}$	$\mathbf{0.83 \pm 0.13}$	$\mathbf{0.88 \pm 0.16}$	$\mathbf{0.83 \pm 0.22}$
	sEMG	kNN	0.70 ± 0.15	0.68 ± 0.15	0.69 ± 0.15	0.76 ± 0.22	0.67 ± 0.21
Y5	Acc	XGBoost	0.82 ± 0.16	0.84 ± 0.14	0.82 ± 0.16	0.81 ± 0.19	0.90 ± 0.15
	Acc+sEMG	**XGBoost**	$\mathbf{0.87 \pm 0.12}$	$\mathbf{0.89 \pm 0.10}$	$\mathbf{0.87 \pm 0.12}$	$\mathbf{0.84 \pm 0.16}$	$\mathbf{0.97 \pm 0.10}$
	sEMG	XGBoost	0.66 ± 0.19	0.66 ± 0.19	0.66 ± 0.20	0.68 ± 0.25	0.72 ± 0.25
Y10	Acc	XGBoost	0.78 ± 0.17	0.77 ± 0.17	0.77 ± 0.17	0.74 ± 0.20	0.82 ± 0.19
	Acc+sEMG	**XGBoost**	$\mathbf{0.85 \pm 0.14}$	$\mathbf{0.85 \pm 0.14}$	$\mathbf{0.84 \pm 0.16}$	$\mathbf{0.82 \pm 0.20}$	$\mathbf{0.93 \pm 0.13}$
	sEMG	XGBoost	0.79 ± 0.15	0.73 ± 0.28	0.79 ± 0.13	0.67 ± 0.27	0.82 ± 0.32
Y20	Acc	ANN	0.80 ± 0.16	0.74 ± 0.28	0.82 ± 0.14	0.79 ± 0.31	0.73 ± 0.32
	Acc+sEMG	**XGBoost**	$\mathbf{0.84 \pm 0.20}$	$\mathbf{0.82 \pm 0.22}$	$\mathbf{0.84 \pm 0.20}$	$\mathbf{0.85 \pm 0.24}$	$\mathbf{0.82 \pm 0.23}$

Wx: x mL of water; Yx: x mL of yogurt; S: saliva.

Among the classifiers, XGBoost was the one that achieved the highest classification performance in most of the experiments, e.g., in all yogurt tasks and also in water$_{10}$. The SVM and kNN only showed the highest accuracy in one of the cases. The results obtained with ANN are not reported in Table 6.5 because it only showed good results for Acc signals in yogurt$_{20}$ but it was not the best one in any of the cases. Furthermore, the best classifiers within the same swallowing task stayed the same for uni-modal and multi-modal experiments, which suggests that the representation spaces created in this experiment are stable and not biased by swallowing tasks.

6.2.2. Analysis of discrimination capability of combined sEMG and Acc

To the best of my knowledge, this experiment was the first work exploring the combination of sEMG and Acc signals for automatic dysphagia screening. Obtained results are discussed below.

In most of the cases, Acc signals yielded higher AUC values than those achieved with sEMG signals, regardless of the swallowing task. When considering different tasks, some of the cases in which this performance difference is more clear in water$_{10}$ (0.71 with sEMG, 0.83 with Acc), saliva (0.72 with sEMG, 0.82 with Acc), yogurt$_5$ (0.70 with sEMG, 0.82 with Acc), and yogurt$_{10}$ as well (0.66 with sEMG, 0.78 with Acc). Additionally, the standard deviation was reduced when the kinematic signals were incorporated. This behavior was also observed in other performance measures like sensitivity. This could be related to the fact that the sEMG measures the electrophysiological dimension indirectly, since it does not allow to record isolated muscles or activation patterns within a muscle [Sejd 18, Step 12]. In contrast, the accelerometry directly records vibratory movements on the throat during swallowing [Sejd 18].

Besides the improvement when using Acc signals, the multi-modal approach produces better results, i.e., when merging information from Acc and sEMG. This is expected from the signal processing point of view because the fusion of multiple sensors, i.e., multi-modality, improves the quality of the information and the uncertainty reduction when creating a model [Hack 90]. This theoretical claim was confirmed by the results of this experiment, regardless of the swallowing tasks (see Table 6.5).

Furthermore, all the resulting AUC values when using the combination of sources were above 0.80. Unfortunately, it is difficult to compare such results with others reported in the literature, because: (1) there are no works where Acc and sEMG signals are combined for dysphagia screening; (2) most of the works that combine sEMG with other sensors are descriptive; (3) the multi-modal works in sEMG and Acc are focused on swallow detection, rather than the discrimination between healthy individuals with normal swallowing and patients with dysphagia. Among the few studies where multi-modal information is considered to model swallowing-related phenomena, [Schu 14] detected swallows from healthy and dysphagic individuals using sEMG and bio-impedance. When only considering healthy participants, the authors reported an accuracy of 96.6%, however, when considering the patients with dysphagia the accuracy drooped down to 84.5%. Similarly, [Hsu 13] reported an $acc = 82.6\%$ when discriminating different levels of dysphagia severity using sEMG and acoustic signals recorded with a microphone. Apart from highlighting the scarcity of related works, it is also necessary to stress the fact that the performance results found in the literature are comparable to the found here.

Why did Experiment #1 (only sEMG) retrieve AUC>0.8 (Table 6.4), whilst Experiment #2 in the uni-modal scenario with sEMG always retrieved AUC<0.8 (Table 6.5)? The most probable reason is the reduced number of sEMG channels used in Experiment #2 (3 vs. 6 in Experiment #1). The lack of masseteric and left infrahyoid channels seems to reduce dramatically the information, making it unsuitable to perform a proper dysphagia screening. This conclusion is supported by results obtained in Experiment #4, explained later.

6.3. Experiment #3: Automatic detection of dysphagia using acoustic biomarkers

In this experiment, the discrimination capability of speech dimensions for dysphagia screening was assessed using different classification algorithms. The obtained results are presented next.

6.3.1. Classification per speech dimension

Different classification models were optimized for each speech dimension, and features selected in sustained vowels, DDK, and monologues, were considered. The classification performance of the assessed models is summarized in Table 6.6.

Although the classification scheme was also performed on the whole feature space, without any feature selection (results not reported here), the best results were obtained with the features selected after the aforementioned statistical tests.

On the other hand, there was consistency in terms of the classifiers and their hyperparameters. The RF achieved the highest performance in phonation and articulation dimensions. In both cases, the mode of the maximum depth was $D = 2$, i.e. few partitions to retrieve information were required. The mode of the number of required estimators was high ($N = 50$ for phonation, $N = 100$ for articulation), but this did not affect bias or variance because the RF model did not overfit. Otherwise, LDA was the best model regarding the DDK and prosody dimensions, and the optimal solver for both cases was the singular value decomposition for all partitions in the nested cross-validation. These simple models strengthen the hypothesis that the evaluated speech dimensions, and particularly the selected features, have the capability to discriminate between healthy and dysphagic recordings. Figure 6.2 illustrates the normalized confusion matrices obtained with the best classifier for each speech dimension. Since the classification scheme was based on nested cross-validation, the values on each confusion matrix were the mean (normalized) value of true positives, false positives, true negatives, and false negatives per fold. It is highlighted that the articulation-related models outperformed the other dimensions in all metrics. This observation has been also reported in previous studies with patients suffering from Parkinson's disease [Oroz 16b, Vasq 18].

Thence, the articulation dimension was the most suitable to discriminate between healthy and dysphagic individuals. Otherwise, the best classifiers in phonation, DDK, and prosody dimensions, had comparable performance, but there were slight differences among classifiers. Additionally, when all the selected features from the three dimensions were combined, the results did not improve consistently in comparison to the articulation dimension. Even though the AUC was the same, its standard deviation was five units higher; furthermore, none of the other performance measures improved the results achieved by the articulation alone.

Finally, aiming to improve the results, a voting ensemble scheme was applied. In such scheme, the following classifiers were gathered: SVM, RF, LR, and LDA. Due to the high computational cost produced by the voting ensemble together with the grid search and nested cross-validation, the ANN classifier was disregarded. Additionally, since the RF is an ensemble of Decision Trees, the DT model was also omitted in the voting ensemble scheme. This approach improved the results obtained with individual models and dimen-

Table 6.6: Classification performance for healthy controls vs. patients in Experiment #3. The best results are shown in bold.

Dimension	Classifier	AUC	F1	Accuracy	Precision	Sensitivity
Phonation	SVM	0.70 ± 0.18	0.54 ± 0.39	0.64 ± 0.24	0.50 ± 0.38	0.63 ± 0.44
	ANN	0.59 ± 0.14	0.47 ± 0.33	0.52 ± 0.20	0.39 ± 0.30	0.66 ± 0.46
	RF	$\mathbf{0.80 \pm 0.10}$	$\mathbf{0.80 \pm 0.08}$	$\mathbf{0.80 \pm 0.09}$	$\mathbf{0.81 \pm 0.15}$	$\mathbf{0.82 \pm 0.10}$
	LR	0.74 ± 0.13	0.65 ± 0.27	0.71 ± 0.18	0.70 ± 0.29	0.66 ± 0.32
	LDA	0.72 ± 0.08	0.72 ± 0.09	0.74 ± 0.08	0.77 ± 0.15	0.71 ± 0.15
	DT	0.70 ± 0.12	0.69 ± 0.09	0.69 ± 0.13	0.74 ± 0.19	0.71 ± 0.17
Articulation	SVM	0.64 ± 0.19	0.51 ± 0.37	0.57 ± 0.24	0.43 ± 0.35	0.68 ± 0.47
	ANN	0.63 ± 0.19	0.46 ± 0.41	0.60 ± 0.24	0.40 ± 0.38	0.57 ± 0.50
	RF	$\mathbf{0.86 \pm 0.10}$	$\mathbf{0.86 \pm 0.10}$	$\mathbf{0.85 \pm 0.11}$	$\mathbf{0.84 \pm 0.18}$	$\mathbf{0.91 \pm 0.12}$
	LR	0.84 ± 0.16	0.78 ± 0.30	0.80 ± 0.24	0.75 ± 0.31	0.86 ± 0.33
	LDA	0.76 ± 0.16	0.76 ± 0.15	0.76 ± 0.15	0.77 ± 0.15	0.79 ± 0.24
	DT	0.81 ± 0.12	0.80 ± 0.13	0.80 ± 0.11	0.83 ± 0.20	0.83 ± 0.20
DDK	SVM	0.73 ± 0.20	0.71 ± 0.20	0.73 ± 0.20	0.75 ± 0.24	0.72 ± 0.23
	ANN	0.70 ± 0.21	0.67 ± 0.29	0.69 ± 0.23	0.66 ± 0.34	0.73 ± 0.32
	RF	0.77 ± 0.17	0.75 ± 0.19	0.77 ± 0.18	0.81 ± 0.23	0.74 ± 0.22
	LR	0.71 ± 0.20	0.64 ± 0.30	0.71 ± 0.22	0.68 ± 0.34	0.66 ± 0.34
	LDA	$\mathbf{0.78 \pm 0.15}$	$\mathbf{0.78 \pm 0.14}$	$\mathbf{0.78 \pm 0.15}$	$\mathbf{0.81 \pm 0.21}$	$\mathbf{0.79 \pm 0.18}$
	DT	0.69 ± 0.20	0.68 ± 0.21	0.69 ± 0.20	0.68 ± 0.25	0.71 ± 0.21
Prosody	SVM	0.69 ± 0.21	0.57 ± 0.36	0.66 ± 0.22	0.54 ± 0.36	0.64 ± 0.42
	ANN	0.58 ± 0.16	0.45 ± 0.35	0.52 ± 0.22	0.40 ± 0.34	0.61 ± 0.48
	RF	0.74 ± 0.15	0.71 ± 0.17	0.73 ± 0.14	0.73 ± 0.21	0.73 ± 0.18
	LR	0.69 ± 0.22	0.66 ± 0.29	0.65 ± 0.25	0.63 ± 0.36	0.79 ± 0.32
	LDA	$\mathbf{0.84 \pm 0.10}$	$\mathbf{0.83 \pm 0.12}$	$\mathbf{0.84 \pm 0.10}$	$\mathbf{0.81 \pm 0.15}$	$\mathbf{0.87 \pm 0.17}$
	DT	0.79 ± 0.09	0.77 ± 0.11	0.77 ± 0.11	0.80 ± 0.21	0.79 ± 0.12
All dim.	SVM	0.69 ± 0.21	0.68 ± 0.28	0.65 ± 0.24	0.61 ± 0.33	0.85 ± 0.31
	ANN	0.82 ± 0.19	0.72 ± 0.39	0.78 ± 0.26	0.66 ± 0.38	0.80 ± 0.42
	RF	0.84 ± 0.11	0.83 ± 0.14	0.83 ± 0.11	0.85 ± 0.17	0.88 ± 0.23
	LR	0.86 ± 0.15	0.81 ± 0.30	0.82 ± 0.24	0.76 ± 0.31	0.88 ± 0.32
	LDA	0.60 ± 0.17	0.58 ± 0.18	0.61 ± 0.17	0.64 ± 0.24	0.60 ± 0.23
	DT	0.81 ± 0.14	0.82 ± 0.11	0.80 ± 0.14	0.82 ± 0.21	0.86 ± 0.12
	Ensemble	$\mathbf{0.91 \pm 0.10}$	$\mathbf{0.90 \pm 0.11}$	$\mathbf{0.90 \pm 0.11}$	$\mathbf{0.88 \pm 0.17}$	$\mathbf{0.93 \pm 0.11}$
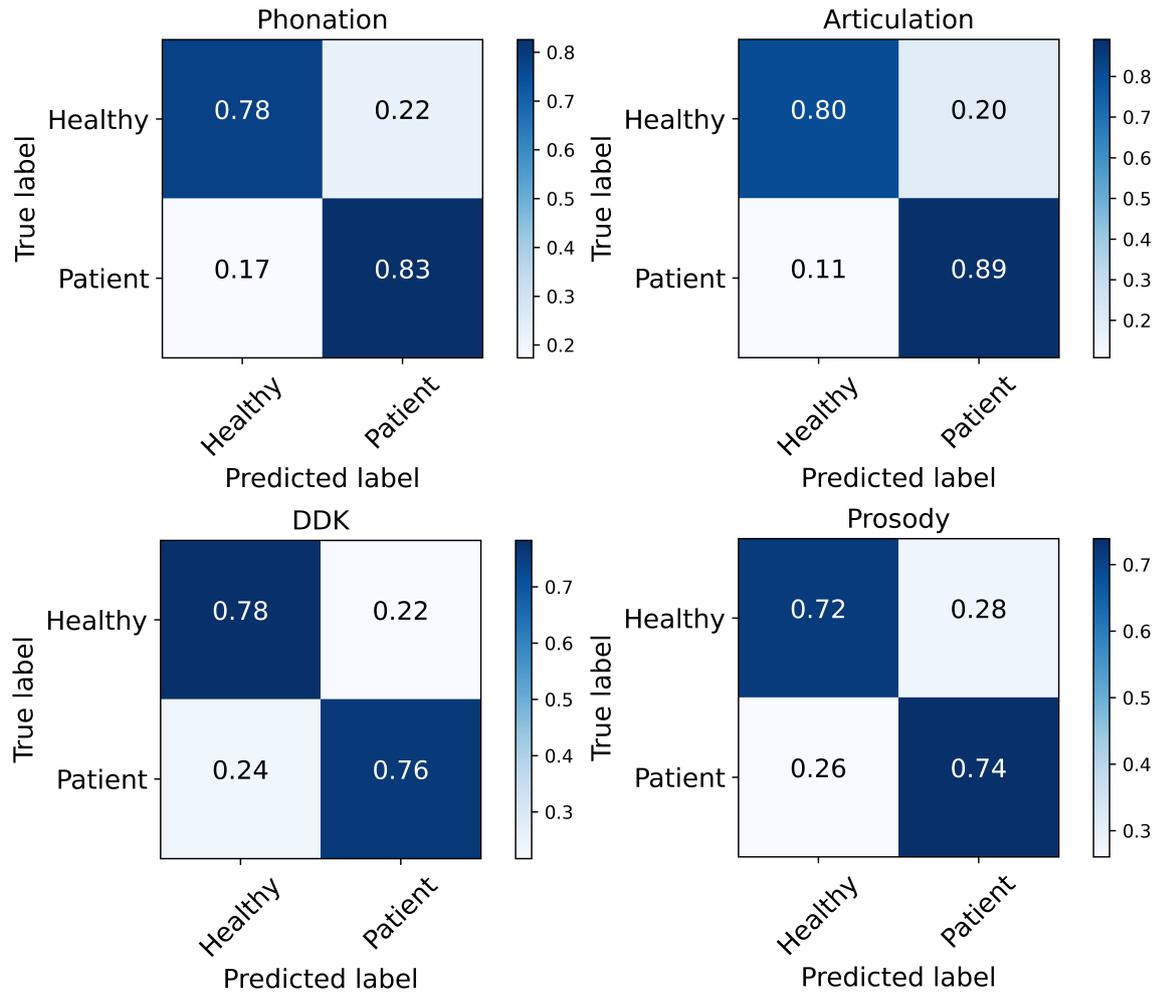
Figure 6.2: Normalized confusion matrices obtained with the best model of each speech dimension.

sions assessed separately, as the performance of all measures shows (see Table 6.6). The mode of hyperparameters was given by: SVM: $C = 10$, $\gamma = 0.01$, sigmoid kernel; RF: $N = 10$, $D = 2$; LR: $C = 10$, L2 norm; and LDA: SVD solver. Parameters of RF and LDA were similar to those obtained by the other models per speech dimension, with the exception of the number of trees which was smaller, as expected with a voting ensemble. Figure 6.3 shows the normalized confusion matrix obtained with the voting ensemble scheme.

6.3.2. Analysis of discrimination capability of speech

Classification results suggest that the studied groups exhibit well-defined patterns in terms of speech dimensions such that allow the automatic discrimination between patients with dysphagia and healthy controls. So, dysphagia produces quantifiable voice changes regardless of the leading clinical condition.

Regarding the implemented classifiers, RF behaved better than DT in all cases, except for prosody. It could be quite obvious since the random forest consists in an ensemble of independent and uncorrelated decision trees, so random forests tend to outperform decision
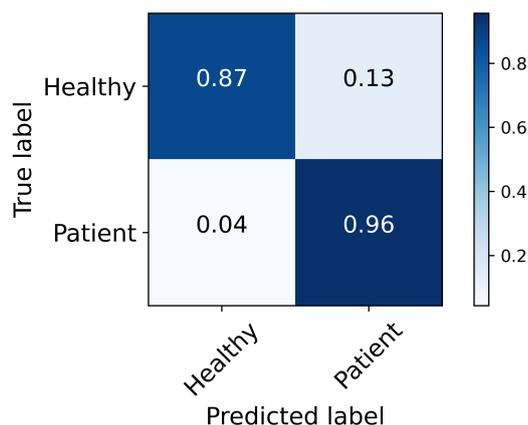
Figure 6.3: Normalized confusion matrix obtained with the voting ensemble scheme.

trees. Actually, the RF was the best model for phonation and articulation, whilst LDA had the higher performance for DDK and prosody (see the Table 6.6). Hence, the nature of the analyzed signals did not require complex algorithms to be modeled; it indicates that features used for the characterization of speech dimensions were appropriate to discriminate healthy and dysphagic states in simple ways. Otherwise, the articulation-related AUC excels because RF, LR and DT achieved excellent discrimination capability in this dimension. Additionally, the sensitivity of the RF in such dimension was 91%. It would be due to the high separability driven by the vowel triangle, as well as the first derivative of the formants or, in other words, by the distribution of the first and second formants in both groups. This likely indicates that the capability to control the position of the tongue during speech production is a clear biomarker of dysphagia produced by neurological conditions. On the other hand, in phonation and prosody, only the RF and LDA, respectively, achieved excellent results in terms of AUC. For DDK, even with the best classifier (LDA), acceptable results were achieved (see Table 6.6).

One important aspect to note is the high standard deviation of some models and metrics. For instance, the standard deviation of the results obtained with some of the classifiers was higher than 20%, particularly in SVM and ANN. This is a limitation of the reported models, but it was addressed in Experiment #4 by increasing the database in order to reduce bias, and with the execution of other speech tasks, in particular the reading of a predefined text.

Finally, when features selected on each dimension were combined, classification results did not improve considerably in terms of AUC, except for ANN (between 58% and 70% in individual dimensions vs. 82% in all dimensions combined). However, the sensitivity was more consistent than in the individual speech dimensions, but with high standard deviation. Notwithstanding, the classification strategy based on ensemble voting improved all the previous results in terms of all performance measures (AUC, F1, accuracy and sensitivity $\geq$90%, see Table 6.6). Additionally, the standard deviation using this scheme decreased or remained equal than in the best results of individual speech dimensions. This strategy seems to be the most suitable to implement dysphagia screening from speech recordings. In this way, although the articulation dimension showed better results individually, all speech dimensions contribute to characterize the dysphagic patient. This would help in improving the diagnosis and quantification of dysphagia by considering simple, cheap, and fast

speech tasks that can be carried out to complement traditional methods used in dysphagia screening.

6.4. Experiment #4: Multi-modality for dysphagia detection

Previous experiments showed that non-invasive biosignals approaches are potentially suitable for dysphagia screening. Also, Experiment #2 suggested that increasing the number of sources of information outperformed the automatic classification of healthy and dysphagic individuals. Subsequently, in this Experiment, different multi-modal classification scenarios were analyzed. Results are presented next.

6.4.1. Classification effect of selected biomarkers

Once different biomarkers of dysphagia were proposed in electrophysiological, mechanical, and acoustic dimensions (see Section 4.4), it is appropriate to determine if the selection actually improved or not the classification result. In this way, the feature selection effect was evaluated, but only in uni-modal scenarios. Using the stratified nested cross-validation scheme shown in Figure 5.10, repeated five times, the different classification algorithms were evaluated with and without feature selection, with the AUC_{ROC} as criterion. Figure 6.4 summarizes the effect for sEMG in Protocols #1 and #2 (onwards $sEMG_{P1}$ and $sEMG_{P2}$, respectively), Acc, and speech, for different test set cardinality. Tables D.1, D.2, D.3, D.4, D.5, D.6, and D.7 of Appendix D, show in detail the effect in performance measures of the feature selection for different test set partitions.

The behavior was dependent on the classifier and the test set size. In Speech, the performance was the same with and without feature selection regardless of the test set size and classifier, with subtle variations in standard deviation. For the other signals, the feature selection tended to improve the classification performance, especially in XGboost and kNN. Only the linear SVM tended to have a larger AUC without feature selection, especially for $sEMG_{P2}$ and Acc in almost all test set partitions.

Note the good classification results achieved in terms of the AUC, especially for speech, with different classifiers. These uni-modal classification outcomes with feature selection outperformed all the best results obtained in Experiments #2 and #3, but were slightly smaller than in Experiment #1. For instance, with a test set size of 20%, the performance measures {AUC, F1, accuracy, precision, sensitivity, specificity[1]}, achieved the following results:

- $sEMG_{P1}$: $\{0.86 \pm 0.14, 0.86 \pm 0.13, 0.84 \pm 0.17, 0.88 \pm 0.19, 0.90 \pm 0.15,\text{-}\}_{Exp\#1}$ with ANN and time domain features in water$_{10}$ vs. $\{0.85 \pm 0.04, 0.82 \pm 0.05, 0.85 \pm 0.04, 0.95 \pm 0.08, 0.73 \pm 0.06, 0.96 \pm 0.06\}_{Exp\#4}$ with kNN.

- $sEMG_{P2}$: $\{0.79 \pm 0.15, 0.73 \pm 0.28, 0.79 \pm 0.13, 0.67 \pm 0.27, 0.82 \pm 0.32,\text{-}\}_{Exp\#2}$ with XGBoost, features selected with mRMR in yogurt$_{20}$ vs. $\{0.86 \pm 0.05, 0.85 \pm 0.05, 0.86 \pm 0.06, 0.89 \pm 0.13, 0.82 \pm 0.05, 0.90 \pm 0.13\}_{Exp\#4}$ with sigmoid SVM.

---

[1]For Experiments #1, #2, and #3 the specificity was not computed.
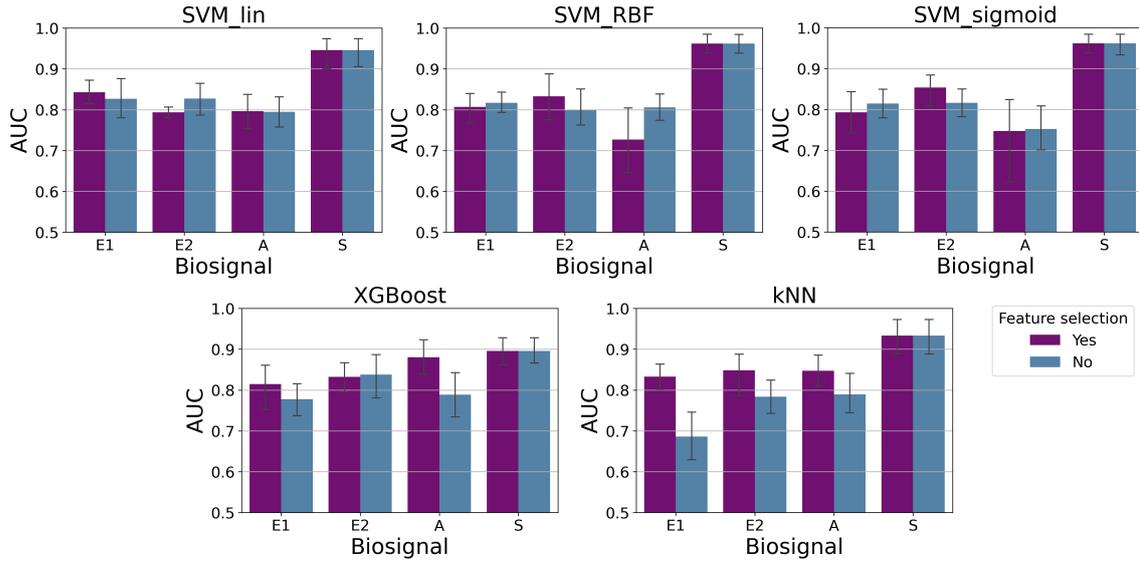
Figure 6.4: Effect of feature selection (AUC_{ROC}) in classical classifiers. Results are shown with a test set size of 25% of the database. E1: sEMG in protocol #1; E2: sEMG in protocol #2; A: accelerometry; S: speech.

- Acc: $\{0.82 \pm 0.16, 0.84 \pm 0.14, 0.82 \pm 0.16, 0.81 \pm 0.19, 0.90 \pm 0.15,-\}_{Exp\#2}$ with XGBoost, features selected with mRMR in yogurt$_5$ vs. $\{0.87 \pm 0.05, 0.87 \pm 0.05, 0.87 \pm 0.05, 0.86 \pm 0.04, 0.88 \pm 0.08, 0.86 \pm 0.05\}_{Exp\#4}$ with XGBoost.

- Speech: $\{0.91 \pm 0.10, 0.90 \pm 0.11, 0.90 \pm 0.11, 0.88 \pm 0.17, 0.93 \pm 0.11,-\}_{Exp\#3}$ with an ensemble of classifiers, features selected with hypothesis tests in all speech dimensions vs. $\{0.95 \pm 0.02, 0.95 \pm 0.02, 0.95 \pm 0.02, 1.00 \pm 0.00, 0.90 \pm 0.04, 1.00 \pm 0.00\}_{Exp\#4}$ with kNN.

The standard deviation provided in Experiment #4 was clearly smaller than in the other experiments, suggesting more reliable models.

6.4.2. Effect of the test set size in the results

The effect of the test set size in the classification performance was also evaluated in this Experiment, varying from 10% to 40%. In the remaining experiments (#1, #2, and #3), this evaluation was not performed due to the limited number of individuals, an issue addressed by Experiment #4 by increasing the database. Bearing in mind that the feature selection tended to improve the uni-modal classification results, henceforth outcomes will be shown only with feature selection. Figure 6.5 shows barplots of the AUC achieved by classification algorithms using different partitions. Detailed results are provided in Appendix D.

It was not observed a trend related to the increase in the database size. Even though some particularities, it's not possible to determine a pattern. In some cases, the reduction of the AUC is more evident, for instance in Acc classified with RBF SVM, and it is natural to expect some kind of decrease because it is hardest to fit a classifier with less training data. The acid test is the exposure of the trained models to the test set and, in this case, the experiments achieved AUC≥ 0.80 in several combinations of classifiers and test sizes.
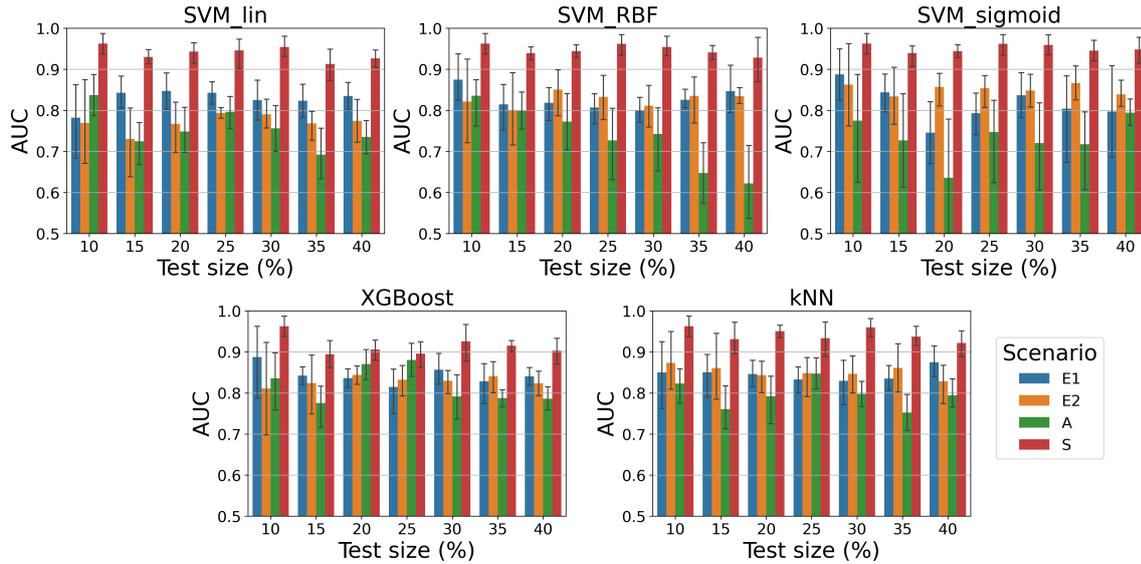
Figure 6.5: Effect of the test size on the AUC from classical algorithms, in uni-modal scenarios considering feature selection. E1: sEMG$_{P1}$; E2: sEMG$_{P2}$; A: accelerometry; S: speech.

This is an indicator of the outcomes' strength: the classification results do not seem to be highly dependent on the test size, in other words, results were not optimistic and classifiers did not exhibit overfitting.

6.4.3. Classical multi-modal classification

The following bimodal and trimodal scenarios were assessed:

- sEMG$_{P1}$ + Acc

- sEMG$_{P1}$ + Speech

- sEMG$_{P2}$ + Acc

- sEMG$_{P2}$ + Speech

- Acc + Speech

- sEMG$_{P1}$ + Acc + Speech

- sEMG$_{P2}$ + Acc + Speech

In order to confirm the generalization of the uni-modal-related findings, the effect of the test set size was also evaluated in the multi-modal scenarios. Figure 6.6 shows barplots of the AUC achieved by classification algorithms using different partitions in such scenarios. Like uni-modal, multi-modal scenarios did not show performance dependencies; in other words, classification was not improved or worsened with dependence on the increase of the test set size. Detailed results for each partition size are shown in Appendix E.
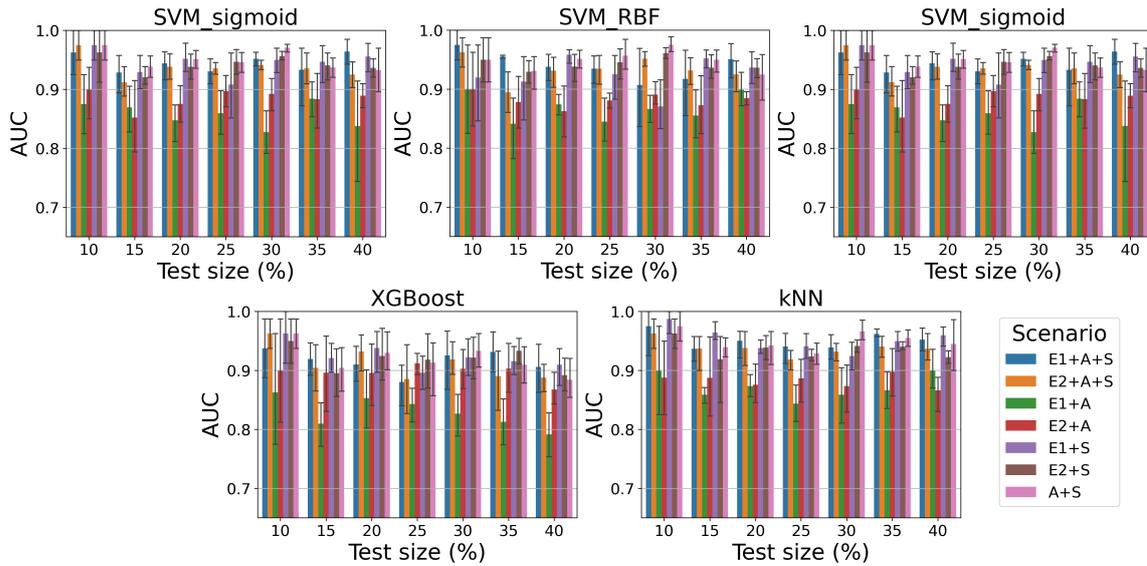
Figure 6.6: Effect of the test size on the AUC from classical algorithms, in multi-modal scenarios considering feature selection. E1: sEMG in protocol #1; E2: sEMG in protocol #2; A: accelerometry; S: speech.

Tables E.1, E.2, E.3, E.4, E.5, E.6, and E.7 show several data hard to analyze. In this way, it is convenient to choose a single partition with feature selection, to compare multi- and uni-modality in Experiment #4.

Table 6.7 shows the AUC achieved by the classical algorithms in different scenarios using feature selection and 20% of the database. In terms of the AUC, the performance was outstanding for almost all of the multi-modal scenarios, with the exception of $sEMG_{P1}$ + Acc, and $sEMG_{P2}$ + Acc. The lowest performances were obtained with uni-modal configurations. The latter scenarios also retrieved the lowest sensitivities and specificities (details in Appendices D and E). Even though some configurations retrieved sensitivity or specificity equal to 1.00, in no case both measures achieved such value. However, several bi- and trimodal scenarios reported a balance between sensitivity and specificity, in particular for trimodal scenarios, e.g. $sEMG_{P1}$ + Acc Speech in kNN (AUC = 0.95 ± 0.03, sensitivity = 0.90 ± 0.06, and specificity = 1.00 ± 0.00), or $sEMG_{P2}$ + Acc Speech in linear SVM (AUC = 0.94 ± 0.02, sensitivity = 0.91 ± 0.03, and specificity = 0.97 ± 0.04). Another interesting scenario is $sEMG_{P1}$ + Speech, which achieved the highest value for two performance measures with different classifiers, e.g. with RBF SVM (**AUC** = 0.96 ± 0.02, **sensitivity** = 0.94 ± 0.03, and specificity = 0.97 ± 0.04), sigmoid SVM (**AUC** = 0.95 ± 0.03, sensitivity = 0.90 ± 0.06, and **specificity** = 1.00 ± 0.00), and with XG-Boost (**AUC** = 0.94 ± 0.05, sensitivity = 0.91 ± 0.03, and **specificity** = 0.96 ± 0.06). In general, bimodal scenarios with speech achieved performance measures higher than 0.90

Additionally, the fact that all classifiers achieved outstanding performance measures (≥ 0.90) in some scenario, is remarkable. Actually, Figure 6.6 and Appendix E show that all classifiers achieved, for certain partitions and scenarios, measures equal to 1.00. There were also cases that retrieved all performance measures equal or greater than 0.95: in 10%, $sEMG_{P1}$ + Acc + Speech (RBF SVM and kNN), $sEMG_{P2}$ + Acc + Speech (all classifiers), $sEMG_{P1}$ + Speech (sigmoid SVM, XGBoost and kNN), and Acc + Speech (linear and

Table 6.7: AUC achieved by classical machine learning algorithms and a test set size of 20%. The best results per classifier are highlighted in bold.

Modality	Scenario	SVM_lin	SVM_RBF	SVM_sigmoid	XGBoost	kNN
Unimodal	E1	0.85 ± 0.06	0.82 ± 0.05	0.75 ± 0.10	0.84 ± 0.03	0.85 ± 0.04
	E2	0.77 ± 0.08	0.85 ± 0.07	0.86 ± 0.05	0.84 ± 0.03	0.84 ± 0.05
	A	0.75 ± 0.07	0.77 ± 0.09	0.64 ± 0.19	0.87 ± 0.05	0.79 ± 0.08
	S	**0.94 ± 0.03**	0.94 ± 0.02	0.94 ± 0.02	0.91 ± 0.03	**0.95 ± 0.02**
Bimodal	E1+A	0.90 ± 0.05	0.87 ± 0.02	0.85 ± 0.04	0.85 ± 0.06	0.87 ± 0.02
	E2+A	0.85 ± 0.06	0.86 ± 0.06	0.88 ± 0.04	0.90 ± 0.07	0.88 ± 0.04
	E1+S	**0.94 ± 0.03**	**0.96 ± 0.02**	**0.95 ± 0.03**	**0.94 ± 0.05**	0.94 ± 0.02
	E2+S	**0.94 ± 0.02**	0.94 ± 0.03	0.94 ± 0.03	0.92 ± 0.06	0.94 ± 0.03
	A+S	**0.94 ± 0.02**	0.95 ± 0.02	**0.95 ± 0.02**	0.93 ± 0.04	0.94 ± 0.04
Trimodal	E1+A+S	**0.94 ± 0.03**	0.94 ± 0.03	0.94 ± 0.03	0.91 ± 0.04	**0.95 ± 0.03**
	E2+A+S	**0.94 ± 0.02**	0.93 ± 0.04	0.94 ± 0.03	0.93 ± 0.04	0.94 ± 0.04

E1: $sEMG_{P1}$; E2: $sEMG_{P2}$; A: accelerometry; S: speech.

sigmoid SVM, XGBoost and kNN); in 15%, $sEMG_{P1}$ + Acc + Speech (RBF SVM), and $sEMG_{P1}$ + Speech (kNN); and in 30%, $sEMG_{P2}$ + Acc + Speech (linear SVM), $sEMG_{P2}$ + Speech (linear, sigmoid and RB SVM), and Acc + Speech (sigmoid and RBF SVM, and kNN).
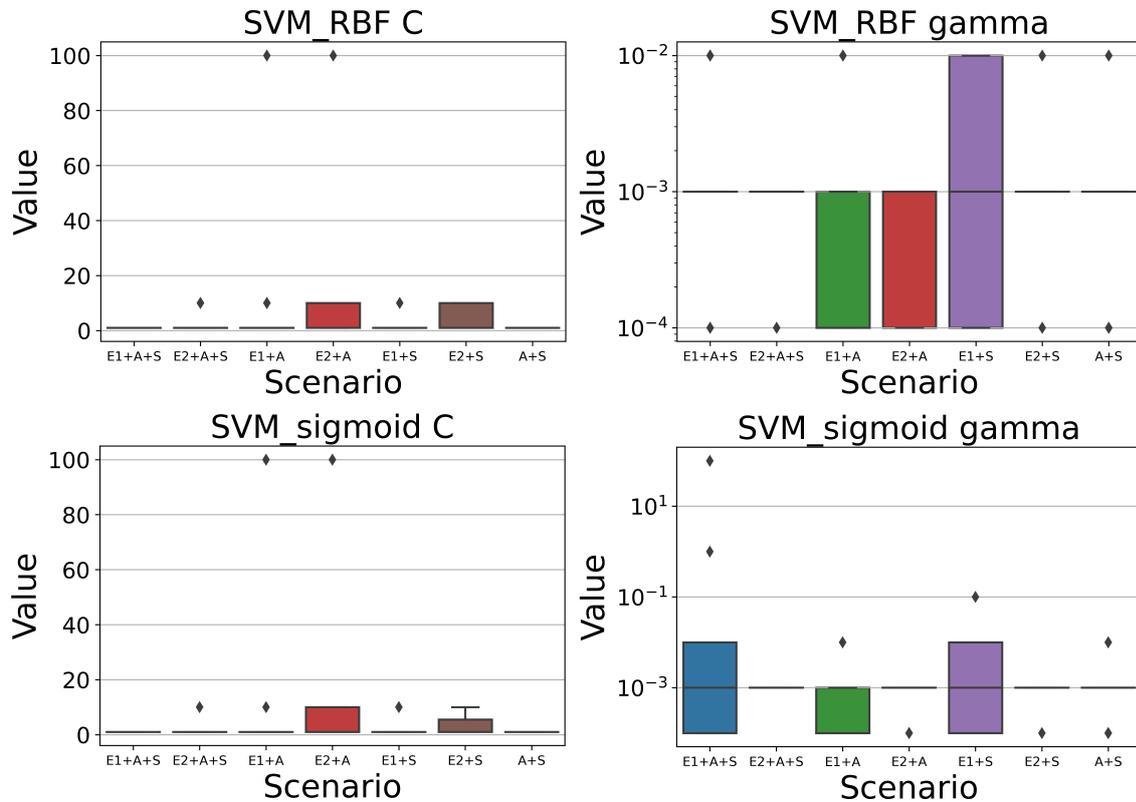
Figure 6.7: Boxplot of the mode of hyperparameters of SVM with sigmoid and RBF kernels. Scenarios with only horizontal lines indicate a lack of variation in hyperparameter values. The y-axis of γ is shown in logarithmic scale, for visualization purposes.

One final remark about classical classifiers: the mode of hyperparameters of the linear SVM, sigmoid SVM, and XGBoost, was very stable for all test set partitions and classification scenarios. The most frequent values, i.e. the mode of the modes, were $C = 1$ for the three SVM kernels and $\gamma = 10^{-3}$ for the RBF and sigmoid kernels. There is not a rule of thumb about what are the ideal C and γ, but low values of C can lead to wider margins but more margin violations, whilst small γ values make the bell-shape of the kernel wider with a smoother decision boundary, i.e., if a model is overfitting, reducing of C or γ can regularize the model [Gero 22]. Linear SVM retrieved $C = 1$ for all scenarios regardless of the partition. These values of C and γ indicate that, despite the good obtained performance, the models seem to generalize well and they don't seem to be overfitted. The RBF and sigmoid SVM had comparable behavior of the hyperparameter C (see Figure 6.7), and the most visible variation was retrieved by scenarios sEMG$_{P2}$ + Acc, and sEMG$_{P2}$ + Speech in both cases. Even though the behavior of γ was different in both kernels, as mentioned, the most frequent value was $\gamma = 10^{-3}$ for both types of model.

Like the hyperparameter C for SVM, the XGBoost-related hyperparameters were very stable. The most frequent maximum depth was $D = 2$, and the ratio of negative/positive classes was $R = 1$. It has sense because the classes were perfectly balanced. In contrast, the number of neighbors of the kNN was very variable in different scenarios and partitions (see Figure 6.8). It is problematic but can be explained by the fact that the kNN has poor tolerance to noise, to highly interdependent and redundant attributes, contrarily to SVM and decision trees-based algorithms [Osis 17].
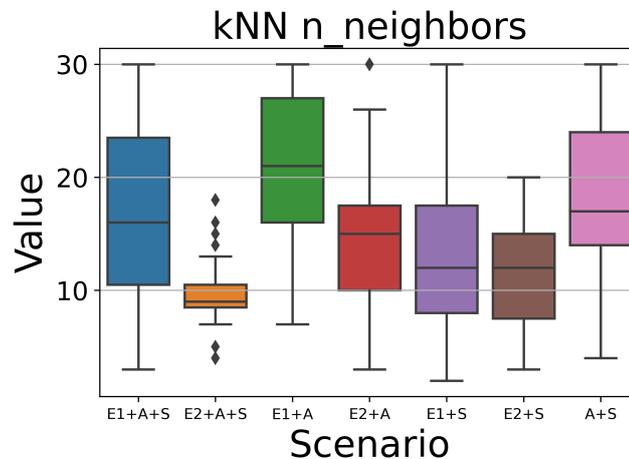
Figure 6.8: Boxplot of the mode of hyperparameters of kNN.

6.4.4. Deep multi-modal classification with GMU

Bearing in mind the results obtained in the previous section, three bimodal and one trimodal scenarios were evaluated with the biomarkers selected in Section 4.4:

- sEMG$_{P1}$ + Speech
- sEMG$_{P1}$ + Acc
- Acc + Speech

> - sEMG$_{P1}$ + Acc + Speech

sEMG$_{P2}$ was not tested to avoid ignoring information about masseters.

These models retrieved better results than the classical ones. Table 6.8 shows the classification outcomes achieved by each scenario using different test set partitions. Like with the classical algorithms, no dependence on the test set size was observed, even though with 10% the bimodal scenario of sEMG$_{P1}$ + Speech retrieved perfect classification. For the other partitions, the best performances were comparable, and from 15% to 40% there was not a clear decrease. The performance was outstanding in general. There were cases with some performance measure equal to 1.00, like the specificity of the trimodal scenario in 15%, and the precision in the bimodal scenario Acc + Speech (in 20% and 25%). The lowest classification result was obtained always with the bimodal scenario sEMG$_{P1}$ + Acc, which is in the same line as previous results, i.e. the speech was the biosignal with the highest performance followed by sEMG$_{P1}$. But this does not mean that the result is poor; in contrast, the AUC was equal to or higher than 0.88 regardless of the test set partition for the aforementioned scenario.

On the other hand, the trimodal scenario or the bimodal sEMG$_{P1}$ + Speech achieved the highest classification performance within the different test set sizes, which strengthens the conclusions about biomarkers and classical classifiers.

Table 6.8: Best classification performances of the GMU-based deep learning architectures, for bimodal and trimodal scenarios. Best result per test set size are highlighted in bold.

Test size	Scenario	AUC	F1	Accuracy	Precision	Sensitivity	Specificity
10%	sEMG$_{P1}$ + Speeh	**1.00 ± 0.00**	**1.00 ± 0.00**	**1.00 ± 0.00**	**1.00 ± 0.00**	**1.00 ± 0.00**	**1.00 ± 0.00**
	sEMG$_{P1}$ + Acc	0.91 ± 0.07	0.91 ± 0.08	0.91 ± 0.07	0.92 ± 0.07	0.90 ± 0.10	0.92 ± 0.07
	Acc + Speech	0.99 ± 0.03	0.99 ± 0.03	0.99 ± 0.03	0.98 ± 0.05	1.00 ± 0.00	0.98 ± 0.06
	sEMG$_{P1}$ + Acc + Speech	0.96 ± 0.03	0.96 ± 0.04	0.96 ± 0.03	0.98 ± 0.05	0.95 ± 0.07	0.98 ± 0.06
15%	sEMG$_{P1}$ + Speeh	0.92 ± 0.05	0.92 ± 0.07	0.92 ± 0.05	0.95 ± 0.04	0.90 ± 0.14	0.95 ± 0.05
	sEMG$_{P1}$ + Acc	0.87 ± 0.03	0.86 ± 0.04	0.87 ± 0.03	0.90 ± 0.06	0.83 ± 0.10	0.90 ± 0.07
	Acc + Speech	0.95 ± 0.02	0.95 ± 0.02	0.95 ± 0.02	0.97 ± 0.04	0.93 ± 0.04	0.97 ± 0.05
	sEMG$_{P1}$ + Acc + Speech	**0.97 ± 0.05**	**0.96 ± 0.05**	**0.97 ± 0.05**	**1.00 ± 0.00**	**0.93 ± 0.09**	**1.00 ± 0.00**
20%	sEMG$_{P1}$ + Speeh	0.97 ± 0.02	0.97 ± 0.02	0.97 ± 0.02	0.99 ± 0.03	0.95 ± 0.03	0.99 ± 0.03
	sEMG$_{P1}$ + Acc	0.89 ± 0.02	0.89 ± 0.02	0.89 ± 0.02	0.94 ± 0.07	0.85 ± 0.07	0.94 ± 0.08
	Acc + Speech	0.98 ± 0.02	0.98 ± 0.02	0.98 ± 0.02	1.00 ± 0.00	0.96 ± 0.03	1.00 ± 0.00
	sEMG$_{P1}$ + Acc + Speech	0.96 ± 0.02	0.95 ± 0.02	0.96 ± 0.02	0.99 ± 0.03	0.92 ± 0.05	0.99 ± 0.03
25%	sEMG$_{P1}$ + Speeh	**0.98 ± 0.02**	**0.97 ± 0.02**	**0.98 ± 0.02**	**0.98 ± 0.03**	**0.97 ± 0.03**	**0.98 ± 0.03**
	sEMG$_{P1}$ + Acc	0.91 ± 0.04	0.91 ± 0.05	0.91 ± 0.04	0.94 ± 0.02	0.88 ± 0.08	0.94 ± 0.02
	Acc + Speech	0.97 ± 0.03	0.97 ± 0.03	0.97 ± 0.03	1.00 ± 0.00	0.94 ± 0.05	1.00 ± 0.00
	sEMG$_{P1}$ + Acc + Speech	0.97 ± 0.03	0.96 ± 0.03	0.97 ± 0.03	0.99 ± 0.02	0.94 ± 0.04	0.99 ± 0.02
30%	sEMG$_{P1}$ + Speeh	0.95 ± 0.03	0.95 ± 0.03	0.95 ± 0.03	0.98 ± 0.04	0.92 ± 0.05	0.98 ± 0.04
	sEMG$_{P1}$ + Acc	0.92 ± 0.03	0.91 ± 0.03	0.92 ± 0.03	0.96 ± 0.04	0.87 ± 0.05	0.97 ± 0.03
	Acc + Speech	0.95 ± 0.03	0.95 ± 0.03	0.95 ± 0.03	0.95 ± 0.03	0.96 ± 0.03	0.95 ± 0.03
	sEMG$_{P1}$ + Acc + Speech	**0.97 ± 0.02**	**0.97 ± 0.02**	**0.97 ± 0.02**	**0.99 ± 0.02**	**0.95 ± 0.03**	**0.99 ± 0.02**
35%	sEMG$_{P1}$ + Speeh	**0.96 ± 0.02**	**0.96 ± 0.02**	**0.96 ± 0.02**	**0.98 ± 0.03**	**0.94 ± 0.05**	**0.98 ± 0.03**
	sEMG$_{P1}$ + Acc	0.91 ± 0.04	0.91 ± 0.04	0.91 ± 0.04	0.93 ± 0.05	0.89 ± 0.05	0.93 ± 0.05
	Acc + Speech	0.95 ± 0.04	0.95 ± 0.04	0.95 ± 0.04	0.96 ± 0.05	0.93 ± 0.06	0.96 ± 0.04
	sEMG$_{P1}$ + Acc + Speech	0.95 ± 0.01	0.94 ± 0.01	0.95 ± 0.01	0.98 ± 0.02	0.91 ± 0.03	0.98 ± 0.02
40%	sEMG$_{P1}$ + Speeh	0.96 ± 0.01	0.95 ± 0.01	0.96 ± 0.01	0.99 ± 0.02	0.92 ± 0.03	0.99 ± 0.02
	sEMG$_{P1}$ + Acc	0.88 ± 0.03	0.87 ± 0.03	0.88 ± 0.03	0.96 ± 0.05	0.81 ± 0.04	0.96 ± 0.04
	Acc + Speech	0.97 ± 0.02	0.97 ± 0.02	0.97 ± 0.02	0.98 ± 0.03	0.95 ± 0.03	0.98 ± 0.03
	sEMG$_{P1}$ + Acc + Speech	**0.97 ± 0.01**	**0.96 ± 0.01**	**0.97 ± 0.01**	**0.99 ± 0.03**	**0.94 ± 0.03**	**0.99 ± 0.03**

Did the GMU architecture outperform the results obtained with classical classifiers? Yes, GMU achieved not only the highest performance measures, even though the differ-

ence was generally about 0.01 or 0.02, but also a smaller standard deviation than classical classifiers.

On the other hand, the hyperparameters of the GMU were even more stable than in SVM with different kernels or XGBoost. In SVM with different kernels and XGBoost, it was evident that the best performances were obtained with a close range of hyperparameters (see Figures 6.7 and 6.8). However, the mode of the hidden layer size was always equal to 64, and the mode of the learning rate was equal to 10^{-3} in the 99.99% of the assessed scenarios.

6.4.5. A concluding analysis of multi-modality

In the last experiment performed in this thesis, the results obtained in the previous ones were assessed with an increase in the database, and with a selection of biomarkers of dysphagia, in order to conclude what is the best strategy to perform a dysphagia screening procedure based on non-invasive biosignals.

Even though $sEMG_{P1}$ had more acquisition channels than $sEMG_{P2}$, retrieved information about the oral phase with the masseter muscles, had bilaterality, and had information about swallowing of solids, i.e. cracker, it exhibited a clear improvement of the classification results achieved by $sEMG_{P2}$ only with linear SVM. Nonetheless, bearing in mind that the acquisition protocol designed for Protocol #1 is more comprehensive than the one designed in Protocol #2, it is more suitable to assess the swallowing process. Their results were not comparable to bimodal or trimodal scenarios but, when a multi-modal scheme was applied, the classification performance was increased, especially when combined with speech. Actually, speech plays also a key role in the classification scenarios, regardless of the algorithm. It was also less affected by the test set size than $sEMG_{P1}$ and $sEMG_{P2}$. There were even cases with perfect performance, in particular with test set sizes of 10% and 15%.

On the other hand, $sEMG_{P2}$ achieved better classification results than Acc. Thus, even though the mechanical dimension has been widely explored during the last decade and has been reported as suitable to detect some kinematic events in videofluoroscopy [Dono 22a], this work suggests that electrophysiological measures such as sEMG, could improve the detection of events in an indirect way (out of the scope of the current thesis). Despite the aforementioned ideas, Experiments #2 and #4 agree in the fact that bimodal scenarios outperformed the classification results of the uni-modal ones.

Perhaps the most important contribution of Experiment #4 was to show that multi-modality improved the classification results achieved by uni-modal scenarios. And in such cases, the Acc helped to outperform the outcomes in both classical and deep learning schemes. This means that even though in uni-modal scenarios the Acc was not the most suitable biosignal to perform dysphagia screening, in multi-modal scenarios it should not be discarded. Additionally, bearing in mind that trimodal scenarios achieved, in general, the best results, it is not recommendable to ignore the information on the mechanical dimension of swallowing.

The fact that the classification was not severely affected by the cardinality of the test set, indicates that despite the difficulties (risk of underfitting) produced by the reduction of the training/validation dataset, classical classifiers with the proposed methodology had a trade-off between bias and variance.

SVM and XGBoost demonstrated their suitability to discriminate against groups of individuals. Even though kNN also demonstrated this capability in terms of performance measures, the high variance of hyperparameters and dependence on the classification scenario, suggest the use of other alternative algorithms, such as the aforementioned ones.

GMU demonstrated to be less sensitive to changes in the test size than the classical classifiers. Additionally, hyperparameters were very stable, more than the behavior evidenced in classical algorithms. Even though the amount of data collected is not small when compared with other related works, it is evident that the size of the dataset is very small to be considered to feed DL models. One way to avoid the training problem would be the use of transfer learning strategies but there are no models trained on similar data. DL models save the hand-crafted feature extraction process, but this would ignore the proposed biomarkers of dysphagia since they were computed as engineering features. Similarly, DL models are not intended to be used on small datasets.

Even though both schemes agree in the fact that the combination of three sources of information achieved the best classification results, the simplicity of classical models - SVM with different kernels and XGBoost - in terms of the number of trainable parameters (tens vs. millions) make the classical classifiers more affordable to perform dysphagia screening when using the information of electrophysiological, mechanical and acoustic swallowing dimensions. In this way, the best algorithm for automatic evaluation of dysphagia should be the simplest one but with a clearer interpretability. This does not mean that the use of other DL algorithms could not be used in the future to analyze the swallowing process, but in such case will be a loss of the interpretability provided by the biomarkers proposed in this thesis.

Chapter 7

Outlook

A conceptual definition of swallowing in terms of different dimensions was proposed in this thesis: electrophysiological, mechanical, acoustic, neurological, cardiorespiratory, symptomatic, and structural. From these, three dimensions were explored and evaluated by the acquisition of non-invasive biosignals: the electrophysiological by sEMG, the mechanical by Acc, and the acoustic by speech recordings. Each of them demanded the design of particular protocols, which led to the proposal of three different acquisition moments: before, during, and after the swallowing process. sEMG and Acc were suitable to be acquired during the swallowing process, so two protocols were designed for such signals: one using only sEMG with six electrodes placed in masseters, supra- and infrahyoid muscles, performing swallowing of different volumes of water and yogurt, saliva and cracker (namely Protocol #1); and another one using three sEMG electrodes in supra- and infrahyoid, as well as a triaxial accelerometer placed in the cricoid cartilage, performing the same swallowing tasks but excluding the cracker to avoid the mastication noise in the Acc signal (namely Protocol #2). The other protocol considered speech tasks (Protocol #3), i.e. sustained vowels, rapid repetitions of the syllables pa-ta-ka, reading of a text, and a free monologue in Spanish, before and after swallowing tasks. The first two protocols were designed together with Speech & Language Pathologists, Physicians, and Engineers. Protocol #3 was adopted from [Vasq 18], oriented to the evaluation of Parkinson's Disease patients, a common underlying condition leading to dysphagia.

The results of this thesis showed that the multi-modality, in particular using the three assessed biosignals, provides relevant information to discriminate between individuals with healthy swallowing and patients with dysphagia. In this regard, the three swallowing dimensions must be considered to perform a non-invasive procedure oriented to dysphagia screening. Notwithstanding, results obtained with the combination of swallowing tasks suggest that the design of an sEMG-based screening methodology could require fewer tasks and acquisition channels than those formulated originally. Bearing in mind that information of sEMG in Protocol #1 contains the same information in Protocol #2, it is recommended to unify such protocols, reducing the time demand and fatigue-like effects that could difficult the execution of swallowing tasks and post-swallowing speech tasks. Even though masseters retrieved good classification results in Experiment #1, masseters-related biomarkers in such Experiment only contributed in $yogurt_5$, and in Experiment #4 these muscles hardly contributed anything as biomarkers. In this way, sEMG and Acc can be acquired simultaneously using four sEMG channels (bilateral supra- and infrahyoid muscles),

in combination with a tri-axial accelerometer at the level of the cricoid cartilage. Furthermore, the cracker can be eliminated from the protocol, because generate artifacts produced by the masseteric activity not only in other sEMG signals but also in Acc. Additionally, this activity did not provide a high amount of information in terms of biomarkers nor classification performance. Moreover, it is also recommended to eliminate 5 mL of yogurt because the density of such consistency makes it hard to swallow the exact volume, and there is always a loss of liquid adhered in the cup, reducing the control and the repeatability of this swallowing task. This reinforces the elimination of the masseters measurements from the acquisition protocol. On the other hand, the small volume of water should be kept because it tended to produce a more accurate classification than the largest volumes. However, large volumes should be also established in the protocol since they tend to produce piecemeal deglutition in patients with dysphagia.

Regarding the speech, pre-swallowing recordings (evaluated in Experiment #3), and post-swallowing recordings (evaluated in Experiment #4), are valuable for dysphagia screening. However, it is not appropriate to perform both simultaneously, because it is another important factor that contributes to time consumption and fatigue in patients. Bearing in mind that several swallowing tasks must be performed to acquire sEMG and Acc signals, the execution of repeated speech recordings is so toilful for patients, especially when they suffer from neurodegenerative diseases in advanced states. In this way, what would be better to acquire in one single assessment session? Both have differential characteristics. Pre-swallowing recordings allow capturing speech impairments without the effect of swallowing tasks, i.e., effects of the underlying disease. In contrast, the post-swallowing recordings measure the effect of such condition plus the cumulative effect of all performed swallowing tasks. Depending on the goal of the evaluation, it could be better to perform pre- or post-swallowing recordings but, in this case, the second one could be better to obtain information about the effects of the swallowing tasks performed.

On the other hand, the execution of free monologues should not be performed unless other co-variables be further analyzed deeply. Even though it worked well in the prosody analysis in Experiment #3, i.e. retrieved energy-related biomarkers and contributed to classification, this task has high variability and is dependent on several factors such as cultural, academic, personality-related, and the underlying disease and its state (more severe, more reduced duration). In contrast, the text that volunteers read is well-standardized and it is less prone to be affected by the mentioned factors. Notwithstanding, the free monologue is very interesting to be addressed in the study of speech impairments, and this is an open research field out of the scope of this thesis.

Another essential point is the good performance of the speech-related models in uni-modal scenarios. If the acoustic dimension of swallowing retrieved comparable classification results for dysphagia screening than the multi-modal approach, why not to discard both electrophysiological and mechanical dimensions, bearing in mind that it would reduce drastically the time consumption per session and the number of tasks, and it would also ease the acquisition protocol? Well, even though the aforementioned ideas are valid, they have deep flaws: 1) not all patients with dysphagia are verbal, i.e., there are patients with limited capabilities to perform speech tasks. Some patients cannot produce a single phoneme due to their clinical condition, e.g. individuals with late stages of progressive neurogenic conditions. Thus, the speech-related protocol would not be suitable for such kinds of patients, in contrast to electrophysiological and mechanical measurements; 2)

the acoustic dimension cannot be evaluated *during* the swallowing process, just before or after that. Subsequently, despite the information of this dimension demonstrated to have paramount importance to evaluate dysphagia, it should not be the only one to characterize patients with swallowing impairments properly. In other words, although important, the classification performance should not be the only criterion to have in mind in the development of a dysphagia screening procedure; other aspects such as the feasibility to perform an acquisition in patients with different etiologies and characteristics must be also assessed in the selection of the most suitable methodology.

One debatable aspect is the heterogeneous etiologies of the dysphagic patients in this study. One could argue that putting all patients with dysphagia in just one set labeled as "dysphagic" is a problematic "Total Evidence" approach, and division of the pathological set into different etiologies is the correct way to address the hypothesis proposed in this thesis. However, this argument faces two main problems: one methodological and another practical. The methodological issue is related to the fact that, if an AI approach is proposed to classify, for instance, patients with dysphagia produced by Parkinson's disease and by ALS, it would be very hard to figure out whether the model is discriminating the presence of dysphagia (the goal of this thesis) or by the swallowing related symptoms produced by the specific etiology. Note the subtle but important difference between these two ideas. The goal of this thesis was not to classify swallowing symptoms in specific neurogenic diseases, but it was to propose a screening approach for functional oropharyngeal dysphagia produced by different neurological conditions. As Roden and Altman claim, "*a universal definition and assessment tool would be useful in formalizing dysphagia evaluation*"[Rode 13]. Additionally, the heterogeneity of the database helps the generalization of results and reduces the risk of bias due to the kind of leading disease or condition. On the other hand, the practical reason is related to the difficulty to recruit patients with neuromuscular conditions. From the final database of 80 patients, only eight suffer from neuromuscular conditions (inflammatory myopathy, dermatomyosis, muscular dystrophy, and Rubinstein-Taybi syndrome). In this case, the number of volunteers with specific conditions to create clusters or sets would be insufficient to have significant outcomes.

Finally, the demonstrated hypothesis was that the multi-modality with three swallowing dimensions is suitable for dysphagia screening. Even though the findings are promising and signify a contribution to the screening phase of the dysphagia evaluation, two main aspects should be addressed in future works: (1) it is necessary to increase the database aiming to improve the generalization capability of the results, to have a big sample size for different neurogenic conditions, and to characterize biosignal-related differences according to each of them; and (2) to compare the obtained results with the reference instrumental methods, i.e. videofluoroscopy or fiberoptic endoscopy, aiming to correlate biosignals-related features with physiological changes directly observed during swallowing tasks, as well as to do epidemiological validations of the proposed method. Although each volunteer was evaluated by a speech and language pathologist and a neurologist, it is necessary to establish the clinical reliability of the method in comparison with the reference test. This consideration, even challenging, is being addressed, and -very- preliminary results are available in a recent publication [Suar 22].

Chapter 8

Summary

This thesis explored the use of non-invasive multi-modal biosignals to evaluate the automatic discrimination between healthy individuals and patients with functional oropharyngeal dysphagia produced by neurological etiologies. Thus, this work was an exploratory diagnostic test accuracy study, i.e. it was in a discrimination phase intended to characterize healthy and dysphagic individuals [Zhou 09]. This was a descriptive work performed on tens of volunteers with known health status evaluated by specialized clinical personnel and with a quantitative score (EAT-10). In this line, the work aimed to determine whether the multi-modal evaluation has any diagnostic value, rather than to pretend replacing the reference instrumental methods for diagnosis (VFSS or FEES). The main contributions of this work can be summarized as follows: 1) this is the first work that explores the combination of surface electromyography (sEMG), accelerometry (Acc), and speech recordings to perform dysphagia screening, using classical machine learning and multi-modal deep learning models; 2) several biomarkers for each biosignal were proposed, with interpretability in different feature domains. The most comprehensive analyses of sEMG and speech-related features were also published in the framework of this thesis; 3) different affordable acquisition protocols were proposed, suitable to perform objective evaluations in the consulting room; and 4) the conceptualization of swallowing dimensions was proposed, and three of them were assessed experimentally in this work: electrophysiological, mechanical, and acoustic.

The biomarkers were determined via AUC_{ROC} from features extracted in different domains, according to the evaluated signal. In sEMG and Acc, time, frequency, time-frequency, and non-linear dynamics domains were evaluated. In speech, the feature analysis was performed using the speech dimensions proposed in [Oroz 18] as a reference, i.e. phonation, articulation, diadochokinesia (DDK), and prosody.

Electrophysiological biomarkers were retrieved especially from supra- and infrahyoid muscles in the four explored domains: VAR, RMS, iEMG, LOG, WL, DASDV, and ZC from the time domain; MNF, MDF, and PKF from the frequency domain; E_A, $E_{D1]}$, E_{D2}, and W_{ent} from the time-frequency domain; and SampEn, LLE, HE, and DFA from the non-linear dynamics domain. All biomarkers achieved $0.70 \leq AUC_{ROC} \leq 0.80$. No significant swallowing task dependence was found.

In comparison, the mechanical biomarkers were retrieved from the three Acc axes, i.e. superior-inferior, anterior-posterior, and medial-lateral, but especially in the time domain: RMS, integral, and LOG. Additionally, the MNP was also retrieved in some swallowing

tasks (5 and 10 mL of water, saliva, and 20 mL of yogurt). Time-frequency and non-linear dynamics domains retrieved scarce features. Contrarily to the electrophysiological case, the number of mechanical biomarkers was small, and the AUC_{ROC} was also small (≤ 0.74). Therefore, the mechanical biomarkers exhibited limited discrimination capability by themselves. Moreover, unlike the electrophysiological biomarkers, the mechanical ones were retrieved mainly in saliva and 20 mL of yogurt. Thus, the evidence suggests that the Acc signals provided more information when the swallow effort is high.

Equally important were the acoustic biomarkers, retrieved from all of the evaluated speech dimensions: APQ, jitter, shimmer, F_0, ΔF_0, and logE extracted from the sustained vowels in the phonation dimension; F_1, ΔF_1, $\Delta^2 F_1$, ΔF_2, MFCC1, ΔMFCC1, Δ^2MFCC1, ΔMFCC4, and MFCC5 in the articulation dimension from sustained vowels; BBE1 and BBE2 (onset and offset), the onset of BBE3$\sim$17, and offset of BBE6, BEE7, and BBE[13$\sim$18] also in the articulation dimension but from readings of a pre-established text; syllables rate and logE from DDK tasks; and logE from a text in the prosody dimension. The AUC_{ROC} from these biomarkers achieved higher values than those extracted in electrophysiological and mechanical dimensions (≤ 0.86). The phonation and articulation evaluated from sustained vowels evidenced this good discrimination capability of the biomarkers by itself, in particular the jitter, shimmer, ΔMFCC1, and Δ^2MFCC1. Thus, even though the evaluated speech tasks retrieved suitable biomarkers, sustained vowels provided the highest amount of information not only in comparison with other speech tasks but also with the other swallowing dimensions. Pre- and post-swallowing recordings retrieved consistent biomarkers.

The results achieved in this work are relative to the classification of individual patients rather than the classification of individual swallows, which is one of the main advantages, bearing in mind that pathological individuals could have some normal swallows during the protocol because of the small volumes assessed. The complex nature of the swallowing process was evidenced by the number of muscle groups, accelerometry axes, and speech tasks required to perform good classification. Uni-modal and multi-modal scenarios with different combinations of biosignals were evaluated. The classification was performed with and without feature selection in order to determine the discrimination capability of the features selected as biomarkers when were used as input for different classifiers. Four experiments with different databases were performed and summarized next.

Experiment #1 evaluated the electrophysiological dimension with sEMG in 60 individuals. SVM, XGBoost, and ANN were used as classifiers and the models were optimized through a stratified nested cross-validation scheme without an external test set. The best classification result was obtained in saliva recordings using time features feeding an ANN: AUC = 0.86 ± 0.10, F1 = 0.83 ± 0.11, accuracy = 0.85 ± 0.10, precision = 0.97 ± 0.10, and sensitivity = 0.75 ± 0.16.

In Experiment #2 two uni-modal and one multi-modal scenarios were evaluated using the electrophysiological and mechanical dimensions. A reduced protocol of sEMG was applied in combination with Acc signals, acquired in 60 individuals. SVM, XG-Boost, kNN, and ANN were used as classifiers. Models were also optimized through a stratified nested cross-validation scheme without an external test set. The multi-modal scenario outperformed the uni-modal ones, and the best classification result was obtained in 5 mL of yogurt using features selected by mRMR feeding an XGBoost model: AUC = 0.87 ± 0.12, F1 = 0.89 ± 0.10, accuracy = 0.87 ± 0.12, precision = $0.84 \pm$

0.16, and sensitivity $= 0.97 \pm 0.10$. Even though the precision was lower than the obtained in Experiment #1, the sensitivity was importantly increased.

Another uni-modal scenario using pre-swallowing speech recordings was evaluated in Experiment #3 related to the acoustic dimension of swallowing. Sustained vowels, rapid repetitions of pa-ta-ka, and a free monologue were acquired previous to swallowing tasks in 92 individuals, oriented to the evaluation of speech dimensions, i.e. phonation, articulation, diadochokinesia, and prosody. SVM, ANN, RF, LR, LDA, and DT were used as classifiers and the models were also optimized through a stratified nested cross-validation scheme without an external test set. Features were selected by hypothesis tests. The best classification result using individual dimensions was obtained with articulation-related features and RF, but it was outperformed by an ensemble of all speech dimensions and classifiers: AUC $= 0.91 \pm 0.10$, F1 $= 0.90 \pm 0.11$, accuracy $= 0.90 \pm 0.11$, precision $= 0.88 \pm 0.17$, and sensitivity $= 0.93 \pm 0.11$.

The above mentioned experiments were performed with limited databases, which led to the modeling without an external test set. I am aware that this could lead to biased or limited results. Experiment (#4) addressed such limitations with the recruitment of 160 individuals matched by gender and age. Test sets were defined randomly but guaranteeing the same number of healthy and dysphagic individuals, and different test set cardinalities were tried, from 10% to 40% of the entire database. The same uni-modal scenarios of Experiments #1, #2, and #3 were also performed, as well as bi-modal and tri-modal scenarios, using SVM with linear, RBF, and sigmoid kernels, XGBoost, and kNN. In contrast with Experiment #3, in this case, post-swallowing speech recordings were analyzed. Uni-modal scenarios with sEMG and speech retrieved better results than in the previous experiments, even though the performance differed according to the test set size. However, this dependence had not a specific pattern. In general, the use of feature selection helped to improve the classification results. Additionally, bi-modal and tri-modal configurations outperformed the results obtained with uni-modal scenarios, with several cases in which all performance measures were higher than 0.95 (actually, there were cases with some performance measure equal to 1.00±0.00). The stability of the hyperparameters in SVM and XGBoost was highlighted among scenarios.

The final scheme applied in Experiment #4 used a GMU algorithm, a DL-based architecture intended for multi-modal problems. Thus, bi-modal and tri-modal scenarios were also evaluated with different test set sizes. Several configurations achieved high classification performance. The tri-modal scenario and bimodal with $sEMG_{P1}$ + Speech achieved measures ≥ 0.95 regardless of the test set size. The mentioned bimodal scenario achieved perfect classification performance (all measures equal to 1.00 ± 0.00). More stable performances were obtained with GMU than with classical classifiers, in terms of variations in the test set size. However, the number of trainable parameters in this algorithm - millions - could be cumbersome.

How are these results in comparison with the state of the art? Some studies have reported the use of other signals such as bi- or tri-axial accelerometry in the neck with microphones, namely high-resolution cervical auscultation, in order to classify healthy (safe, non-aspirated) and non-healthy (unsafe, aspirated) swallows. Even though this goal sounds similar to the proposed in this thesis, there is a slight but relevant difference: it is not the same to classify healthy and non-healthy swallows in one population (e.g. only patients with dysphagia) or even in just one individual, as to classify one individual as healthy

or non-healthy. Consequently, it is inconvenient to compare any performance measure of such research with the ones obtained in this work. Just two works are methodologically comparable: 1) Miyagi et al. (2020) used a highly unbalanced database with 27 healthy young controls and 143 patients with dysphagia. With an SVM applied on swallowing sounds, they achieved modest results in comparison with the ones obtained in this thesis: F1 = 78.9%, accuracy = 77%, precision = 73.7%, and recall = 87% [Miya 20]; and 2) Donohue, et al. (2021) reported the first study that used "HRCA (accelerometry based cervical auscultation and swallowing sounds) to differentiate between healthy swallows and swallows from people in a category of underlying disease that commonly results in dysphagia". They achieved impressive classification results with classical ML algorithms: accuracy and specificity of 99% and sensitivity of 100% [Dono 21b]. However, they also used an unbalanced database with 20 patients with dysphagia and 51 healthy controls, so the results could be biased and optimistic. Additionally, even in multi-modal (bi-modal), the contribution of each swallowing dimension was scarcely discussed, and electrophysiological or speech-related information was not acquired. Actually, to the best of my knowledge, there are no available models of dysphagia in the field of computational para-linguistics, with the exception of the study carried out by [Ipin 18], which focused on the detection of basal and viscosity states in six patients with Parkinson's disease, and such performed by [Zhao 22], with modest results and a clear misunderstanding of the results obtained in Experiment #3, with which they are compared. In summary, I believe this is one of the most comprehensive works in terms of automatic analysis of changes in swallowing dimensions in patients with dysphagia from an engineering point of view. The results of this work contribute to the state of the art in the objective analysis of dysphagia and have the potential to support the development of technology that can be effectively transferred to the consulting room.

To conclude, a proper tool for dysphagia evaluation should be reliable, non-invasive, radiation-free, inexpensive, simple to operate, and provide qualitative and quantitative information [Vaim 09]. In this way, the computational deglutition based on the integration and automatic analysis of Acc, sEMG, and speech signals could meet most of the established requirements, and provide solutions both in screening (characterization of functional causes of oropharyngeal dysphagia), as well as in follow-up activities. The multi-modal approach proposed in this thesis could support the gold standard methods for dysphagia diagnosis. Methodologies for the non-invasive evaluation of neurogenic or neuromuscular oropharyngeal dysphagia are relevant and useful in the healthcare environment and the proposed approach is potentially useful to determine specific therapy or to perform/document the follow-up of patients, helping to reduce the number of fluoroscopy/endoscopy sessions, especially in health systems of low- and middle-income countries with limited access to VFSS, FEES, or other instrumental tests.

Chapter 9

Publications emerged from the development of this thesis

This chapter contains publications with results obtained during the development of this thesis:

Roldan-Vasco, S., Restrepo-Uribe, J. P., Orozco-Duque, A., Suarez-Escudero, J. C., & Orozco-Arroyave, J. R. (2023). Analysis of electrophysiological and mechanical dimensions of swallowing by non-invasive biosignals. *Biomedical Signal Processing and Control*, 82, 104533.

Roldan-Vasco, S., Orozco-Duque, A., & Orozco-Arroyave, J. R. (2023). Swallowing disorders analysis using surface EMG biomarkers and classification models. *Digital Signal Processing*, 133, 103815.

Roldan-Vasco, S., Orozco-Duque, A., Suarez-Escudero, J. C., & Orozco-Arroyave, J. R. (2021). Machine learning based analysis of speech dimensions in functional oropharyngeal dysphagia. *Computer Methods and Programs in Biomedicine*, 208, 106248.

Suárez-Patiño, L. V., Orozco-Duque, A., Pérez-Giraldo, E., Roldán-Vasco, S., Suárez-Escudero, J. C., & Martínez-Moreno, L. (2022). Sincronización entre la videodeglución y la electromiografía de superficie en pacientes con afectación neurológica y síntomas de disfagia [Synchronization between videofluoroscopic swallowing study and surface electromyography in patients with neurological involvement presenting symptoms of dysphagia]. *Biomédica*, 42(4), 650-664.

Flórez-Gómez, A. F., Orozco-Arroyave, J. R., & Roldán-Vasco, S. (2022). Correlación entre espacios de características acústicas del habla y trastornos clínicos de la voz en pacientes con disfagia. [Correlation Between Speech-Related Feature Spaces and Clinical Voice Disorders in Patients with Dysphagia]. *TecnoLógicas*, 25(53).

Roldán-Vasco, S., Pérez-Giraldo, E., & Orozco-Duque, A. (2020). Scalogram-energy based segmentation of surface electromyography signals from swallowing related muscles.

Computer methods and programs in Biomedicine, 194, 105480.

Roldan-Vasco, S., Restrepo-Agudelo, S., Valencia-Martinez, Y., & Orozco-Duque, A. (2018). Automatic detection of oral and pharyngeal phases in swallowing using classification algorithms and multichannel EMG. *Journal of Electromyography and Kinesiology*, 43, 193-200.

The following conference papers were also published:

Restrepo-Uribe, J. P., Roldan-Vasco, S., Perez-Giraldo, E., Orozco-Arroyave, J. R., & Orozco-Duque, A. (2020). Electrophysiological and Mechanical Approaches to the Swallowing Analysis. In *Applied Computer Sciences in Engineering: 7th Workshop on Engineering Applications, WEA 2020, Bogota, Colombia, October 7–9, 2020, Proceedings 7* (pp. 281-290). Springer International Publishing.

Roldan-Vasco, S., Perez-Giraldo, E., & Orozco-Duque, A. (2018). Continuous wavelet transform for muscle activity detection in surface EMG signals during swallowing. In *Applied Computer Sciences in Engineering: 5th Workshop on Engineering Applications, WEA 2018, Medellín, Colombia, October 17-19, 2018, Proceedings, Part II 5* (pp. 245-255). Springer International Publishing.

Anhang

Appendix A

Statistical analysis of speech-related features in Experiment #3

This appendix includes statistical comparisons between features from healthy individuals and patients with dysphagia in the Experiment #3. The p-value was adjusted with Bonferroni correction. The effect size was estimated by η^2. Comparisons between healthy controls and patients are shown. Features extracted from sustained vowels that were used for classification are highlighted in bold.

Table A.1: Phonation related features extracted from the sustained vowels.

Feature	/a/		/e/		/i/		/o/		/u/	
	η^2	p-value	η^2	p-value	η^2	p-value	η^2	p-value	η^2	p-value
APQ_mean	0.026	1	0.017	1	0.000	1	0.183	p<0.01	0.107	0.141
PPQ_mean	0.043	1	0.009	1	0.013	1	0.066	1	0.116	0.089
Jitter_mean	**0.133**	**0.038**	**0.106**	**0.149**	**0.195**	**p<0.01**	**0.160**	**0.010**	**0.306**	**p<0.001**
Shimmer_mean	0.088	0.354	0.040	1	0.021	1	0.261	p<0.001	0.230	p<0.001
F_0_mean	0.000	1	0.001	1	0.011	1	0.006	1	0.013	1
ΔF_0_mean	0.000	1	0.006	1	0.009	1	0.004	1	0.094	0.265
$\Delta^2 F_0$_mean	0.018	1	0.011	1	0.021	1	0.022	1	0.000	1
logE_mean	0.000	1	0.017	1	0.011	1	0.198	p<0.01	0.125	0.057
APQ_std	0.017	1	0.015	1	0.000	1	0.158	0.011	0.023	1
PPQ_std	0.049	1	0.006	1	0.001	1	0.029	1	0.071	0.844
Jitter_std	**0.123**	**0.062**	**0.103**	**0.165**	**0.217**	**p<0.001**	**0.161**	**p<0.01**	**0.312**	**p<0.001**
Shimmer_std	0.012	1	0.042	1	0.022	1	0.205	p<0.01	0.179	p<0.01
F_0_**std**	**0.075**	**0.688**	**0.125**	**0.057**	**0.217**	**p<0.001**	**0.175**	**p<0.01**	**0.341**	**p<0.001**
ΔF_0_std	0.115	0.092	0.107	0.137	0.214	p<0.001	0.124	0.059	0.305	p<0.001
$\Delta^2 F_0$_std	0.101	0.183	0.126	0.054	0.222	p<0.001	0.146	0.020	0.305	p<0.001
logE_std	**0.051**	**1**	**0.193**	**p<0.01**	**0.170**	**p<0.01**	**0.246**	**p<0.001**	**0.270**	**p<0.001**
APQ_skew	0.001	1	0.001	1	0.008	1	0.017	1	0.162	p<0.01
PPQ_skew	0.008	1	0.005	1	0.013	1	0.001	1	0.006	1
Jitter_skew	0.083	0.473	0.037	1	0.054	1	0.126	0.054	0.138	0.030
Shimmer_skew	0.028	1	0.000	1	0.009	1	0.026	1	0.012	1
F_0_skew	0.087	0.381	0.055	1	0.174	p<0.01	0.114	0.099	0.152	0.015
ΔF_0_skew	0.002	1	0.008	1	0.076	0.672	0.045	1	0.188	p<0.01
$\Delta^2 F_0$_skew	0.070	0.922	0.005	1	0.003	1	0.032	1	0.025	1
logE_skew	0.001	1	0.035	1	0.035	1	0.115	0.094	0.019	1
APQ_kurt	0.000	1	0.000	1	0.024	1	0.029	1	0.138	0.030
PPQ_kurt	0.004	1	0.003	1	0.017	1	0.003	1	0.002	1
Jitter_kurt	0.073	0.754	0.032	1	0.051	1	0.118	0.080	0.119	0.078
Shimmer_kurt	0.031	1	0.000	1	0.007	1	0.037	1	0.018	1
F_0_kurt	0.035	1	0.014	1	0.036	1	0.185	p<0.01	0.179	p<0.01
ΔF_0_kurt	0.052	1	0.038	1	0.035	1	0.111	0.114	0.117	0.084
$\Delta^2 F_0$_kurt	0.060	1	0.027	1	0.020	1	0.135	0.035	0.101	0.188
logE_kurt	0.000	1	0.071	0.863	0.067	1	0.150	0.017	0.061	1

Table A.2: Mean of the articulation related features extracted from the sustained vowels.

Feature	/a/ η^2	/a/ p-value	/e/ η^2	/e/ p-value	/i/ η^2	/i/ p-value	/o/ η^2	/o/ p-value	/u/ η^2	/u/ p-value
F_1_mean	0.001	1	0.011	1	0.036	1	0.051	1	0.141	0.139
F_2_mean	0.013	1	0.097	1	0.113	0.551	0.103	0.891	0.179	0.022
ΔF_1_mean	**0.249**	**p<0.001**	**0.177**	**0.024**	**0.268**	**p<0.001**	**0.233**	**p<0.01**	**0.140**	**0.148**
ΔF_2_mean	0.263	p<0.001	0.099	1	0.220	p<0.01	0.135	0.188	0.174	0.028
$\Delta^2 F_1$_mean	0.012	1	0.003	1	0.047	1	0.031	1	0.003	1
$\Delta^2 F_2$_mean	0.005	1	0.001	1	0.009	1	0.023	1	0.082	1
MFCC1_mean	0.012	1	0.001	1	0.000	1	0.032	1	0.001	1
MFCC2_mean	0.021	1	0.036	1	0.088	1	0.000	1	0.011	1
MFCC3_mean	0.049	1	0.032	1	0.005	1	0.000	1	0.000	1
MFCC4_mean	0.096	1	0.005	1	0.015	1	0.345	p<0.001	0.215	p<0.01
MFCC5_mean	0.040	1	0.277	p<0.001	0.188	0.014	0.133	0.205	0.196	p<0.01
MFCC6_mean	0.022	1	0.123	0.344	0.042	1	0.067	1	0.151	0.086
MFCC7_mean	0.201	p<0.01	0.003	1	0.123	0.344	0.014	1	0.014	1
MFCC8_mean	0.034	1	0.048	1	0.041	1	0.024	1	0.103	0.915
MFCC9_mean	0.009	1	0.016	1	0.024	1	0.048	1	0.157	0.063
MFCC10_mean	0.045	1	0.035	1	0.030	1	0.001	1	0.003	1
MFCC11_mean	0.018	1	0.104	0.868	0.057	1	0.001	1	0.005	1
MFCC12_mean	0.007	1	0.001	1	0.000	1	0.001	1	0.005	1
ΔMFCC1_mean	0.017	1	0.020	1	0.003	1	0.024	1	0.003	1
ΔMFCC2_mean	0.244	p<0.001	0.126	0.290	0.212	p<0.01	0.030	1	0.016	1
ΔMFCC3_mean	0.022	1	0.006	1	0.017	1	0.042	1	0.084	1
ΔMFCC4_mean	0.030	1	0.000	1	0.010	1	0.034	1	0.023	1
ΔMFCC5_mean	0.059	1	0.003	1	0.010	1	0.007	1	0.021	1
ΔMFCC6_mean	0.125	0.298	0.104	0.868	0.056	1	0.292	p<0.001	0.250	p<0.001
ΔMFCC7_mean	0.041	1	0.002	1	0.001	1	0.017	1	0.002	1
ΔMFCC8_mean	0.119	0.418	0.122	0.354	0.070	1	0.049	1	0.000	1
ΔMFCC9_mean	0.000	1	0.056	1	0.010	1	0.205	p<0.01	0.136	0.177
ΔMFCC10_mean	0.125	0.298	0.000	1	0.077	1	0.195	0.010	0.205	p<0.01
ΔMFCC11_mean	0.121	0.374	0.016	1	0.016	1	0.009	1	0.119	0.418
ΔMFCC12_mean	0.015	1	0.000	1	0.001	1	0.030	1	0.005	1
Δ^2MFCC1_mean	0.181	0.019	0.108	0.702	0.069	1	0.186	0.015	0.258	p<0.001
Δ^2MFCC2_mean	0.155	0.069	0.035	1	0.134	0.199	0.109	0.666	0.190	0.012
Δ^2MFCC3_mean	0.141	0.139	0.008	1	0.003	1	0.211	p<0.01	0.162	0.050
Δ^2MFCC4_mean	0.032	1	0.048	1	0.004	1	0.042	1	0.112	0.582
Δ^2MFCC5_mean	0.001	1	0.026	1	0.020	1	0.010	1	0.094	1
Δ^2MFCC6_mean	0.019	1	0.006	1	0.000	1	0.001	1	0.007	1
Δ^2MFCC7_mean	0.009	1	0.000	1	0.001	1	0.022	1	0.027	1
Δ^2MFCC8_mean	0.008	1	0.016	1	0.027	1	0.008	1	0.006	1
Δ^2MFCC9_mean	0.110	0.648	0.043	1	0.000	1	0.017	1	0.057	1
Δ^2MFCC10_mean	0.003	1	0.005	1	0.014	1	0.050	1	0.024	1
Δ^2MFCC11_mean	0.017	1	0.141	0.139	0.058	1	0.056	1	0.080	1
Δ^2MFCC12_mean	0.025	1	0.055	1	0.034	1	0.073	1	0.001	1
TKEO_mean	0.003	1	0.005	1	0.015	1	0.014	1		

Table A.3: Standard deviation of the articulation related features extracted from the sustained vowels.

Feature	/a/		/e/		/i/		/o/		/u/	
	η^2	p-value	η^2	p-value	η^2	p-value	η^2	p-value	η^2	p-value
F_1_std	**0.059**	**1**	**0.271**	**p<0.001**	**0.304**	**p<0.001**	**0.316**	**p<0.001**	**0.318**	**p<0.001**
F_2_std	0.140	0.144	0.026	1	0.014	1	0.236	p<0.01	0.268	p<0.001
ΔF_1_std	**0.111**	**0.614**	**0.252**	**p<0.001**	**0.325**	**p<0.001**	**0.282**	**p<0.001**	**0.335**	**p<0.001**
ΔF_2_std	**0.202**	**p<0.01**	**0.031**	**1**	**0.083**	**1**	**0.242**	**p<0.01**	**0.292**	**p<0.001**
$\Delta^2 F_1$_std	**0.094**	**1**	**0.281**	**p<0.001**	**0.290**	**p<0.001**	**0.298**	**p<0.001**	**0.289**	**p<0.001**
$\Delta^2 F_2$_std	0.172	0.030	0.026	1	0.091	1	0.254	p<0.001	0.227	p<0.01
MFCC1_std	**0.107**	**0.761**	**0.221**	**p<0.01**	**0.221**	**p<0.01**	**0.231**	**p<0.01**	**0.280**	**p<0.001**
MFCC2_std	0.035	1	0.053	1	0.050	1	0.248	p<0.001	0.303	p<0.001
MFCC3_std	0.127	0.282	0.044	1	0.014	1	0.219	p<0.01	0.091	1
MFCC4_std	0.035	1	0.078	1	0.055	1	0.169	0.035	0.190	0.013
MFCC5_std	0.002	1	0.201	p<0.01	0.152	0.080	0.032	1	0.085	1
MFCC6_std	0.038	1	0.027	1	0.066	1	0.104	0.846	0.045	1
MFCC7_std	0.011	1	0.021	1	0.071	1	0.017	1	0.055	1
MFCC8_std	0.049	1	0.026	1	0.052	1	0.026	1	0.043	1
MFCC9_std	0.088	1	0.015	1	0.004	1	0.020	1	0.062	1
MFCC10_std	0.021	1	0.016	1	0.009	1	0.029	1	0.013	1
MFCC11_std	0.053	1	0.030	1	0.017	1	0.089	1	0.005	1
MFCC12_std	0.058	1	0.065	1	0.008	1	0.041	1	0.040	1
ΔMFCC1_std	**0.241**	**p<0.01**	**0.259**	**p<0.001**	**0.235**	**p<0.01**	**0.293**	**p<0.001**	**0.372**	**p<0.001**
ΔMFCC2_std	0.009	1	0.030	1	0.061	1	0.205	p<0.01	0.262	p<0.001
ΔMFCC3_std	0.049	1	0.007	1	0.001	1	0.094	1	0.044	1
ΔMFCC4_std	0.003	1	0.012	1	0.021	1	0.015	1	0.080	1
ΔMFCC5_std	0.028	1	0.099	1	0.027	1	0.020	1	0.116	0.480
ΔMFCC6_std	0.000	1	0.012	1	0.084	1	0.082	1	0.073	1
ΔMFCC7_std	0.022	1	0.003	1	0.012	1	0.034	1	0.035	1
ΔMFCC8_std	0.010	1	0.035	1	0.092	1	0.086	1	0.098	1
ΔMFCC9_std	0.013	1	0.015	1	0.012	1	0.020	1	0.039	1
ΔMFCC10_std	0.013	1	0.030	1	0.053	1	0.186	0.015	0.116	0.480
ΔMFCC11_std	0.040	1	0.049	1	0.074	1	0.141	0.139	0.070	1
ΔMFCC12_std	0.040	1	0.065	1	0.114	0.522	0.106	0.781	0.073	1
Δ^2MFCC1_std	**0.174**	**0.028**	**0.224**	**p<0.01**	**0.250**	**p<0.001**	**0.256**	**p<0.001**	**0.328**	**p<0.001**
Δ^2MFCC2_std	0.001	1	0.007	1	0.016	1	0.145	0.116	0.138	0.162
Δ^2MFCC3_std	0.004	1	0.003	1	0.000	1	0.018	1	0.001	1
Δ^2MFCC4_std	0.039	1	0.009	1	0.006	1	0.011	1	0.000	1
Δ^2MFCC5_std	0.047	1	0.020	1	0.001	1	0.001	1	0.033	1
Δ^2MFCC6_std	0.003	1	0.002	1	0.059	1	0.022	1	0.042	1
Δ^2MFCC7_std	0.044	1	0.000	1	0.001	1	0.008	1	0.006	1
Δ^2MFCC8_std	0.002	1	0.029	1	0.054	1	0.066	1	0.053	1
Δ^2MFCC9_std	0.001	1	0.004	1	0.010	1	0.000	1	0.011	1
Δ^2MFCC10_std	0.003	1	0.019	1	0.044	1	0.138	0.162	0.079	1
Δ^2MFCC11_std	0.017	1	0.034	1	0.084	1	0.120	0.385	0.032	1
Δ^2MFCC12_std	0.022	1	0.069	1	0.122	0.354	0.135	0.188	0.066	1
TKEO_std	0.015	1	0.006	1	0.021	1	0.011	1		

Table A.4: Skewness of the articulation related features extracted from the sustained vowels.

Feature	/a/		/e/		/i/		/o/		/u/	
	η^2	p-value	η^2	p-value	η^2	p-value	η^2	p-value	η^2	p-value
F_1_skew	0.045	1	0.008	1	0.004	1	0.027	1	0.026	1
F_2_skew	0.020	1	0.028	1	0.155	0.071	0.031	1	0.032	1
ΔF_1_skew	0.025	1	0.014	1	0.008	1	0.018	1	0.001	1
ΔF_2_skew	0.028	1	0.003	1	0.069	1	0.008	1	0.048	1
$\Delta^2 F_1$_skew	0.065	1	0.042	1	0.009	1	0.002	1	0.007	1
$\Delta^2 F_2$_skew	0.011	1	0.052	1	0.021	1	0.053	1	0.016	1
MFCC1_skew	0.014	1	0.001	1	0.004	1	0.003	1	0.038	1
MFCC2_skew	0.001	1	0.015	1	0.072	1	0.052	1	0.004	1
MFCC3_skew	0.003	1	0.033	1	0.021	1	0.077	1	0.147	0.103
MFCC4_skew	0.036	1	0.044	1	0.026	1	0.087	1	0.000	1
MFCC5_skew	0.129	0.244	0.012	1	0.005	1	0.064	1	0.095	1
MFCC6_skew	0.027	1	0.018	1	0.052	1	0.000	1	0.023	1
MFCC7_skew	0.035	1	0.088	1	0.053	1	0.095	1	0.053	1
MFCC8_skew	0.004	1	0.009	1	0.008	1	0.040	1	0.001	1
MFCC9_skew	0.015	1	0.006	1	0.000	1	0.001	1	0.000	1
MFCC10_skew	0.000	1	0.027	1	0.001	1	0.000	1	0.015	1
MFCC11_skew	0.009	1	0.001	1	0.000	1	0.046	1	0.018	1
MFCC12_skew	0.005	1	0.005	1	0.000	1	0.002	1	0.009	1
ΔMFCC1_skew	0.031	1	0.001	1	0.002	1	0.019	1	0.034	1
ΔMFCC2_skew	0.129	0.251	0.014	1	0.043	1	0.053	1	0.001	1
ΔMFCC3_skew	0.026	1	0.137	0.172	0.068	1	0.207	$p<0.01$	0.000	1
ΔMFCC4_skew	0.024	1	0.126	0.290	0.003	1	0.050	1	0.109	0.666
ΔMFCC5_skew	0.019	1	0.027	1	0.027	1	0.101	0.989	0.132	0.211
ΔMFCC6_skew	0.030	1	0.005	1	0.016	1	0.000	1	0.004	1
ΔMFCC7_skew	0.001	1	0.003	1	0.001	1	0.004	1	0.006	1
ΔMFCC8_skew	0.020	1	0.001	1	0.002	1	0.002	1	0.050	1
ΔMFCC9_skew	0.025	1	0.032	1	0.006	1	0.057	1	0.029	1
ΔMFCC10_skew	0.003	1	0.012	1	0.030	1	0.036	1	0.000	1
ΔMFCC11_skew	0.013	1	0.007	1	0.000	1	0.000	1	0.017	1
ΔMFCC12_skew	0.001	1	0.006	1	0.018	1	0.015	1	0.017	1
Δ^2MFCC1_skew	0.006	1	0.005	1	0.092	1	0.005	1	0.001	1
Δ^2MFCC2_skew	0.031	1	0.000	1	0.009	1	0.001	1	0.012	1
Δ^2MFCC3_skew	0.058	1	0.060	1	0.004	1	0.110	0.631	0.004	1
Δ^2MFCC4_skew	0.000	1	0.018	1	0.003	1	0.009	1	0.003	1
Δ^2MFCC5_skew	0.007	1	0.021	1	0.000	1	0.000	1	0.001	1
Δ^2MFCC6_skew	0.002	1	0.001	1	0.000	1	0.001	1	0.023	1
Δ^2MFCC7_skew	0.001	1	0.026	1	0.048	1	0.021	1	0.000	1
Δ^2MFCC8_skew	0.001	1	0.000	1	0.018	1	0.028	1	0.009	1
Δ^2MFCC9_skew	0.008	1	0.007	1	0.000	1	0.000	1	0.002	1
Δ^2MFCC10_skew	0.000	1	0.001	1	0.001	1	0.022	1	0.003	1
Δ^2MFCC11_skew	0.000	1	0.000	1	0.000	1	0.006	1	0.027	1
Δ^2MFCC12_skew	0.001	1	0.023	1	0.008	1	0.007	1	0.014	1
TKEO_skew	0.001	1	0.034	1	0.052	1	0.022	1	0.006	1

Table A.5: Kurtosis of the articulation related features extracted from the sustained vowels.

Feature	/a/		/e/		/i/		/o/		/u/	
	η^2	p-value	η^2	p-value	η^2	p-value	η^2	p-value	η^2	p-value
F_1_kurt	0.003	1	0.017	1	0.002	1	0.010	1	0.000	1
F_2_kurt	0.015	1	0.005	1	0.003	1	0.066	1	0.049	1
ΔF_1_kurt	0.000	1	0.066	1	0.020	1	0.021	1	0.006	1
ΔF_2_kurt	0.002	1	0.006	1	0.020	1	0.074	1	0.031	1
$\Delta^2 F_1$_kurt	0.000	1	0.037	1	0.027	1	0.005	1	0.008	1
$\Delta^2 F_2$_kurt	0.001	1	0.004	1	0.007	1	0.058	1	0.002	1
MFCC1_kurt	0.010	1	0.023	1	0.043	1	0.033	1	0.000	1
MFCC2_kurt	0.035	1	0.008	1	0.001	1	0.002	1	0.013	1
MFCC3_kurt	0.001	1	0.018	1	0.052	1	0.005	1	0.012	1
MFCC4_kurt	0.001	1	0.019	1	0.000	1	0.168	0.036	0.051	1
MFCC5_kurt	0.057	1	0.003	1	0.026	1	0.016	1	0.040	1
MFCC6_kurt	0.014	1	0.003	1	0.017	1	0.001	1	0.021	1
MFCC7_kurt	0.041	1	0.001	1	0.001	1	0.015	1	0.002	1
MFCC8_kurt	0.000	1	0.005	1	0.035	1	0.002	1	0.050	1
MFCC9_kurt	0.001	1	0.003	1	0.009	1	0.001	1	0.010	1
MFCC10_kurt	0.038	1	0.000	1	0.027	1	0.023	1	0.002	1
MFCC11_kurt	0.019	1	0.043	1	0.033	1	0.019	1	0.080	1
MFCC12_kurt	0.018	1	0.023	1	0.097	1	0.004	1	0.024	1
ΔMFCC1_kurt	0.013	1	0.012	1	0.049	1	0.008	1	0.001	1
ΔMFCC2_kurt	0.049	1	0.063	1	0.109	0.666	0.195	p<0.01	0.245	p<0.001
ΔMFCC3_kurt	0.154	0.073	0.126	0.290	0.005	1	0.178	0.023	0.060	1
ΔMFCC4_kurt	0.106	0.802	0.140	0.144	0.074	1	0.016	1	0.043	1
ΔMFCC5_kurt	0.027	1	0.089	1	0.022	1	0.158	0.059	0.201	p<0.01
ΔMFCC6_kurt	0.135	0.182	0.206	p<0.01	0.104	0.868	0.120	0.396	0.108	0.702
ΔMFCC7_kurt	0.043	1	0.041	1	0.001	1	0.055	1	0.098	1
ΔMFCC8_kurt	0.103	0.891	0.062	1	0.107	0.761	0.029	1	0.002	1
ΔMFCC9_kurt	0.064	1	0.067	1	0.001	1	0.000	1	0.037	1
ΔMFCC10_kurt	0.000	1	0.016	1	0.000	1	0.110	0.648	0.005	1
ΔMFCC11_kurt	0.017	1	0.002	1	0.001	1	0.055	1	0.014	1
ΔMFCC12_kurt	0.021	1	0.000	1	0.029	1	0.008	1	0.011	1
Δ^2MFCC1_kurt	0.001	1	0.001	1	0.018	1	0.011	1	0.031	1
Δ^2MFCC2_kurt	0.031	1	0.058	1	0.165	0.043	0.283	p<0.001	0.268	p<0.001
Δ^2MFCC3_kurt	0.098	1	0.118	0.430	0.005	1	0.158	0.059	0.010	1
Δ^2MFCC4_kurt	0.036	1	0.174	0.028	0.151	0.086	0.047	1	0.089	1
Δ^2MFCC5_kurt	0.051	1	0.097	1	0.013	1	0.056	1	0.115	0.494
Δ^2MFCC6_kurt	0.124	0.316	0.136	0.177	0.133	0.205	0.076	1	0.122	0.354
Δ^2MFCC7_kurt	0.020	1	0.045	1	0.001	1	0.059	1	0.064	1
Δ^2MFCC8_kurt	0.060	1	0.097	1	0.135	0.182	0.021	1	0.002	1
Δ^2MFCC9_kurt	0.038	1	0.055	1	0.012	1	0.000	1	0.031	1
Δ^2MFCC10_kurt	0.001	1	0.016	1	0.003	1	0.057	1	0.002	1
Δ^2MFCC11_kurt	0.053	1	0.052	1	0.001	1	0.070	1	0.006	1
Δ^2MFCC12_kurt	0.011	1	0.002	1	0.000	1	0.001	1	0.028	1
TKEO_kurt	0.015	1	0.052	1	0.051	1	0.001	1	0.101	1

Table A.6: Bark-band energies in monologues.

Features	Onset		Offset	
	η^2	p-value	η^2	p-value
BBE_1_mean	0.364	p<0.001	0.243	p<0.001
BBE_2_mean	0.137	0.014	0.001	1
BBE_3_mean	0.030	1	0.059	0.726
BBE_4_mean	0.055	0.910	0.066	0.495
BBE_5_mean	0.043	1	0.060	0.696
BBE_6_mean	0.041	1	0.057	0.805
BBE_7_mean	0.042	1	0.043	1
BBE_8_mean	0.022	1	0.034	1
BBE_9_mean	0.016	1	0.020	1
BBE_10_mean	0.013	1	0.023	1
BBE_11_mean	0.011	1	0.024	1
BBE_12_mean	0.007	1	0.024	1
BBE_13_mean	0.018	1	0.060	0.682
BBE_14_mean	0.049	1	0.092	0.131
BBE_15_mean	0.001	1	0.036	1
BBE_16_mean	0.012	1	0.039	1
BBE_17_mean	0.032	1	0.043	1
BBE_18_mean	0.021	1	0.044	1
BBE_1_std	0.503	p<0.001	0.401	p<0.001
BBE_2_std	0.402	p<0.001	0.260	p<0.001
BBE_3_std	0.093	0.124	0.044	1
BBE_4_std	0.018	1	0.036	1
BBE_5_std	0.044	1	0.113	0.047
BBE_6_std	0.073	0.339	0.203	p<0.001
BBE_7_std	0.034	1	0.180	p<0.01
BBE_8_std	0.042	1	0.167	p<0.01
BBE_9_std	0.047	1	0.196	p<0.001
BBE_10_std	0.045	1	0.187	p<0.01
BBE_11_std	0.044	1	0.196	p<0.001
BBE_12_std	0.059	0.726	0.210	p<0.001
BBE_13_std	0.090	0.148	0.186	p<0.01
BBE_14_std	0.066	0.495	0.124	0.026
BBE_15_std	0.078	0.263	0.060	0.696
BBE_16_std	0.056	0.856	0.078	0.276
BBE_17_std	0.066	0.506	0.054	0.928
BBE_18_std	0.088	0.159	0.079	0.251

Table A.7: DDK related features.

Feature	η^2	p-value
F_0_var	0.001	1
logE_mean	0.150	p<0.01
logE_var	0.200	p<0.001
logE_max	0.000	1
DDK_rate	0.129	p<0.01
DDK_mean	0.189	p<0.001
DDK_reg	0.151	p<0.001
Pause_rate	0.034	0.391
Pauses_mean	0.027	0.565
Pause_reg	0.004	1

Table A.8: Prosody related features.

Feature	η^2	p-value
F_0_max	0.010	1
F_0_mean	0.020	1
F_0_std	0.011	1
F_0_skew	0.024	1
F_0_kurt	0.020	1
E_max	0.003	1
logE_mean	0.095	0.031
logE_std	0.355	p<0.001
logE_skew	0.114	0.012
logE_kurt	0.018	1
V_rate	0.020	1
V_mean	0.020	1
V_std	0.002	1
V_skew	0.076	0.084
V_kurt	0.078	0.077
Sil_rate	0.028	1
Sil_mean	0.007	1
Sil_std	0.036	0.694
Sil_skew	0.041	0.523
Sil_kurt	0.042	0.495

Appendix B

Wavelet denoising

Since detail coefficients of the DWT, i.e. cD_j, have high frequency components, the decreasing of noise content in such coefficients has been implemented as a strategy of signal denoising. This procedure is based on an amplitude thresholding of cD_j [Phin 09]. Different ways to find such threshold are described next:

- Universal thresholding: It uses the fixed threshold $THR = \sqrt{2\log(N)}$, where N is the length of the signal $x[n]$ [Phin 09].

- Stein's Unbiased Estimate of Risk (SURE): This method was proposed by Stein (1981) [Stei 81]; it estimates and minimizes the risk for a particular threshold value [Phin 09]. Let Y be a random variable $N(0,1)$, and X another random variable $N(\mu, \sigma^2)$, that [Stei 81]:

$$X = \sigma_n Y + \mu$$

where μ is the mean of a multivariate normal distribution with the identity as covariance matrix, and usually $\sigma_n = \sigma/\sqrt{n}$ (n is the dimension of the random variables). Let also $\hat{\mu}(X) = \hat{\mu}$ be an estimator of μ, so the -squared error- loss is defined as follows [Wass 06]:

$$L(\hat{\mu}, \mu) = \sum_{i=1}^{n} (\hat{\mu}_i - \mu_i)^2 = ||\hat{\mu} - \mu||^2$$

Let the function $g(X) = \hat{\mu} - X$ that $g : \mathbb{R}^n \to \mathbb{R}^n$ [Wass 06]. Thus, Stein defines an unbiased estimator of the risk (or expected loss) of $\hat{\mu}$, as follows [Stei 81, Wass 06]:

$$\text{SURE}(\hat{\mu}) = n\sigma_n^2 + 2||g(X)||^2 + 2\sigma_n^2 \bigtriangledown g(X)$$

where $\bigtriangledown g(X) = \partial g(x_1, \cdots, x_n)/\partial x_i$.

- Heuristic SURE: It combines the universal and SURE thresholding methods [Phin 09].

- Minimax thresholding: It makes a minimax estimate of the threshold, and it was also proposed by Stein [Stei 81]. The minimax estimate of μ is given as follows:

$$\hat{\mu} = X + \triangledown \log f(X)$$

where $f(X)$ is related to $g(X)$ by the expression $g(X) = \triangledown \log f(X)$.

Once the threshold is retrieved, there are two ways to apply it on the detail coefficients, by hard and soft transformations:

$$cD_j = \begin{cases} \phi(cD_j) & \text{if } |cD_j| > THR_j \\ 0 & \text{otherwise} \end{cases} \tag{B.1}$$

where THR_j is the chosen threshold for the j-th detail coefficient, $\phi(cD_j) = cD_j$ for the hard thresholding, and $\phi(cD_j) = \text{sgn}(cD_j)(cD_j - THR_j)$ for the soft thresholding [Phin 09].

For sEMG and Acc signals, the following ranges of parameters were evaluated to maximize the SNR, according to [Phin 09],: mother wavelet (db2~db9, sym3~sym6, and coif2~coif5); number of decomposition levels (2~9); threshold selection rule (universal, minimax, the Stein's unbiased risk estimate - SURE, and a heuristic variant of SURE); type of threshold (hard/soft); and re-scaling with noise level determined from the first coefficient, level-dependent and without re-scaling.

Appendix C

AUC$_{ROC}$ of features in Experiment #4

This appendix includes the AUC$_{ROC}$ retrieved by each feature selected for Experiment #4, in Protocols #1 (only sEMG), #2 (sEMG+Acc), and #3 (post-swallowing speech). Only values ≥ 0.7 were included.

Table C.1: AUC$_{ROC}$ of the time domain features retrieved in Protocol #1 (only sEMG), Experiment #4.

Task	Muscle	VAR mean	VAR std	RMS mean	RMS std	iEMG mean	iEMG std	log max	log std	WL max	WL mean	WL std	DASDV max	DASDV std	ZC min	ZC mean	MYOP skew	MYOP kurt	TKEO mean	TKEO std
Water$_5$	LM																			
	RSH																			
	LSH																			
	RIH														0.74	0.71				
	LIH														0.72					
Water$_{10}$	RSH																			
	LSH																			
	RIH														0.71	0.70				
	LIH																			
Water$_{20}$	RM														0.72		0.71	0.71		
	LM																			
	RSH					0.70		0.71												
	LSH									0.73	0.70	0.75							0.71	
	RIH				0.72		0.72	0.72							0.78	0.74				
	LIH		0.70		0.70															
Saliva	RM																			
	RSH							0.70							0.71					
	LSH	0.71			0.73		0.75	0.70	0.74	0.71	0.73	0.74	0.71	0.73					0.74	0.72
	RIH														0.80					
	LIH														0.71					
Yogurt$_5$	RSH																			
	LSH																			
	RIH														0.73					
	LIH																			
Yogurt$_{10}$	RSH																			
	LSH							0.70												
	RIH	0.76	0.71	0.71	0.71	0.71	0.73								0.76					
	LIH																			
Yogurt$_{20}$	RSH																			
	LSH																			
	RIH	0.76	0.73		0.76		0.77								0.73					
	LIH		0.70		0.72		0.73	0.70												
Cracker	LM																			
	RSH																			
	LSH																			
	RIH														0.77					

RM: right masseter; LM: left masseter; RSH: right suprahyoid muscles; LSH: left suprahyoid muscles; RIH: right infrahyoid muscles; LIH: left infrahyoid muscles.

Table C.2: AUC$_{ROC}$ of the frequency domain features retrieved in Protocol #1 (only sEMG), Experiment #4.

Task	Muscle	FreqRatio min	MeanPow mean	MeanPow std	MeanFreq max	MeanFreq min	MeanFreq mean	MedianFreq max	MedianFreq min	MedianFreq mean	MedianFreq std	PeakFreq max	PeakFreq mean
Water$_5$	LM					0.71							
	RSH												
	LSH												
	RIH	0.75			0.71	0.72	0.74		0.71	0.73			0.75
	LIH	0.70											
Water$_{10}$	RSH												
	LSH												
	RIH	0.70			0.75	0.78	0.73	0.74	0.71	0.73			0.72
	LIH	0.71			0.71								
Water$_{20}$	RM												
	LM					0.70							
	RSH												
	LSH												
	RIH				0.76	0.77	0.77	0.78	0.75	0.76	0.71	0.72	0.75
	LIH			0.70	0.73	0.71	0.72	0.72				0.71	0.70
Saliva	RM								0.71				
	RSH					0.70							
	LSH		0.70										
	RIH					0.75	0.71		0.72				
	LIH					0.72							
Yogurt$_5$	RSH												
	LSH												
	RIH	0.76				0.75	0.72		0.72	0.71			0.73
	LIH				0.73			0.73					0.72
Yogurt$_{10}$	RSH												
	LSH												
	RIH	0.76	0.77	0.71	0.74	0.77	0.74	0.75	0.71	0.72			0.73
	LIH					0.71							
Yogurt$_{20}$	RSH												
	LSH												
	RIH	0.72	0.76	0.73	0.70	0.71	0.71	0.72					0.71
	LIH			0.70									
Cracker	LM												
	RSH												
	LSH												
	RIH					0.72			0.72				0.71

RM: right masseter; LM: left masseter; RSH: right suprahyoid muscles; LSH: left suprahyoid muscles; RIH: right infrahyoid muscles; LIH: left infrahyoid muscles.

Table C.3: AUC$_{ROC}$ of the time-frequency domain features retrieved in Protocol #1 (only sEMG), Experiment #4.

Task	Muscle	E_D1			E_D2			E_D3		E_D4			E_D5		E_A					Wav_entropy					
		max	mean	std	max	mean	std	max	mean	mean	skew	kurt	max	std	max	mean	std	skew	kurt	max	min	mean	std	skew	kurt
Water5	LM																								
	RSH																					0.70			
	LSH																						0.73		0.75
	RIH	0.74	0.71	0.72	0.75	0.71	0.72									0.74		0.72				0.73		0.74	
	LIH	0.73			0.71		0.70																		
Water10	RSH																						0.73		0.76
	LSH																						0.75		0.72
	RIH	0.73	0.73	0.71	0.74	0.73	0.72			0.72	0.79	0.79	0.70		0.70	0.73		0.71				0.71		0.72	
	LIH				0.75	0.71	0.72																		
Water20	RM																								
	LM																								
	RSH				0.71							0.70											0.73		0.75
	LSH										0.71	0.73											0.73		0.74
	RIH			0.70	0.77	0.76	0.75	0.74	0.74						0.74	0.76		0.75			0.74	0.77		0.75	
	LIH			0.72	0.76	0.74	0.75																		
Saliva	RM																								
	RSH																					0.72		0.73
	LSH																0.70						0.74		0.72
	RIH															0.72		0.72				0.72		0.74	
	LIH																								
Yogurt5	RSH												0.73	0.72									0.72		
	LSH																						0.76		
	RIH	0.71			0.72	0.71	0.70						0.74			0.75		0.73				0.73		0.73	
	LIH	0.70			0.77		0.73	0.71												0.70					
Yogurt10	RSH									0.71	0.71												0.70		0.70
	LSH																								
	RIH	0.71	0.72	0.71	0.71	0.72	0.71		0.70							0.74		0.72				0.74		0.70	
	LIH																								
Yogurt20	RSH																						0.74		0.71
	LSH									0.71	0.70														
	RIH	0.71			0.73	0.71	0.71																		
	LIH				0.72		0.71													0.73					
Cracker	LM																								
	RSH																						0.73		
	LSH									0.73	0.72	0.70						0.71					0.74		
	RIH																								

RM: right masseter; LM: left masseter; RSH: right suprahyoid muscles; LSH: left suprahyoid muscles; RIH: right infrahyoid muscles; LIH: left infrahyoid muscles.

Table C.4: AUC$_{ROC}$ of the nonlinear dynamics domain features retrieved in Protocol #1 (only sEMG), Experiment #4.

Task	Muscle	Shannon	Sample_ent					LZC		Lyapunov_largest_exp				Hurst_exp			Corr_dim	DFA			
		kurt	max	min	mean	std	kurt	min	mean	max	min	mean	std	max	skew	kurt	kurt	max	mean	skew	kurt
Water5	LM																				
	RSH																				
	LSH																				
	RIH				0.71					0.74		0.74						0.71	0.72		
	LIH													0.73				0.71			
Water10	RSH																				
	LSH										0.71				0.70	0.71					0.71
	RIH									0.73		0.73						0.73	0.71		
	LIH																				
Water20	RM																				
	LM																				
	RSH																				
	LSH					0.70											0.72				0.71
	RIH				0.71					0.75		0.76	0.71					0.80	0.75		
	LIH				0.71								0.70	0.73							
Saliva	RM																				
	RSH																				
	LSH		0.70		0.71																0.72
	RIH			0.71				0.71				0.70						0.75			
	LIH																				
Yogurt5	RSH																				
	LSH													0.72							
	RIH									0.73		0.72		0.70							
	LIH									0.75											
Yogurt10	RSH																				
	LSH																				
	RIH				0.71					0.75		0.73						0.75	0.70		
	LIH																				
Yogurt20	RSH																				
	LSH													0.73							
	RIH									0.72				0.70			0.72				
	LIH																				
Cracker	LM													0.72							
	RSH													0.73							
	LSH	0.71			0.73			0.72						0.73						0.73	
	RIH																	0.76			

RM: right masseter; LM: left masseter; RSH: right suprahyoid muscles; LSH: left suprahyoid muscles; RIH: right infrahyoid muscles; LIH: left infrahyoid muscles.

Table C.5: AUC$_{ROC}$ of the time domain features retrieved in Protocol #2 (sEMG+Acc), Experiment #4.

RSH: right suprahyoid muscles; LSH: left suprahyoid muscles; RIH: right infrahyoid muscles; AP: anterior-posterior axis; SI: superior-inferior axis; ML: medial-lateral axis.

Table C.6: AUC$_{ROC}$ of the frequency domain features retrieved in Protocol #2 (sEMG+Acc), Experiment #4.

Task	Channel	FreqRatio	MeanPow					MeanFreq		MedianFreq			PeakFreq
		std	max	mean	std	skew	kurt	min	kurt	min	mean	kurt	mean
Water$_5$	RSH												
	LSH		0.70	0.71	0.75								
	RIH			0.71				0.72					0.71
	AP				0.71								
Water$_{10}$	RSH												
	LSH		0.71		0.73								
	RIH		0.70										
	ML		0.70		0.71								
Water$_{20}$	RSH												
	LSH								0.70			0.70	
	RIH				0.70						0.71		0.72
	AP												
	ML												
Saliva	LSH		0.71	0.73	0.73			0.71					
	RIH												
	AP			0.70									
	SI				0.71								
	ML												
Yogurt$_5$	RSH												
	LSH			0.72	0.71								
	RIH			0.72	0.70			0.73					
	SI												
	ML	0.71											
Yogurt$_{10}$	RSH												
	LSH				0.73								
	RIH							0.75		0.73			
	ML												
Yogurt$_{20}$	RSH												
	LSH												
	RIH												0.70
	AP	0.72											
	SI					0.71	0.72						
	ML					0.73	0.73						

RSH: right suprahyoid muscles; LSH: left suprahyoid muscles; RIH: right infrahyoid muscles; AP: anterior-posterior axis; SI: superior-inferior axis; ML: medial-lateral axis.

Table C.7: AUC$_{ROC}$ of the time-frequency domain features retrieved in Protocol #2 (sEMG+Acc), Experiment #4.

Task	Channel	E_D1			E_D4			E_D5				E_D8			E_D9				E_A		Wav_entropy			
		max	mean	std	mean	skew	kurt	max	mean	std	skew	max	mean	std	min	mean	std	skew	mean	skew	mean	std	skew	kurt
Water$_5$	RSH							0.71		0.74												0.75		0.70
	LSH							0.73		0.74												0.70		
	RIH							0.74	0.75	0.72									0.74					
	AP																							
Water$_{10}$	RSH									0.71												0.70		
	LSH					0.71	0.74			0.71												0.73		
	RIH								0.72										0.73		0.71			
	ML															0.72								
Water$_{20}$	RSH							0.71	0.71	0.70												0.74		0.71
	LSH					0.74	0.75	0.78		0.73												0.76		0.72
	RIH	0.70							0.76		0.71								0.72	0.71				
	AP															0.72								
	ML												0.71			0.74		0.73						
Saliva	LSH																							
	RIH	0.73	0.71	0.72															0.71	0.72			0.70	
	AP																							
	SI																							
	ML																							
Yogurt$_5$	RSH																					0.70		
	LSH							0.74	0.76	0.80												0.71		
	RIH							0.73	0.76	0.73									0.70					
	SI																							
	ML											0.73	0.71	0.71										
Yogurt$_{10}$	RSH																					0.76		
	LSH					0.72	0.71	0.78	0.72	0.78												0.76		0.72
	RIH								0.75	0.71														
	ML																							
Yogurt$_{20}$	RSH				0.70	0.71	0.71															0.71		
	LSH																					0.72		
	RIH	0.72		0.72				0.77	0.72										0.70					
	AP												0.74		0.71	0.77	0.75		0.71	0.71			0.70	
	SI																							
	ML							0.71	0.71	0.72						0.72	0.72							

RSH: right suprahyoid muscles; LSH: left suprahyoid muscles; RIH: right infrahyoid muscles; AP: anterior-posterior axis; SI: superior-inferior axis; ML: medial-lateral axis.

Table C.8: AUC$_{ROC}$ of the nonlinear dynamics domain features retrieved in Protocol #2 (sEMG+Acc), Experiment #4.

Task	Channel	Sample_ent kurt	LZC min	Hurst_exp skew
Water$_5$	RSH			
	LSH			
	RIH			
	AP			
Water$_{10}$	RSH			
	LSH			
	RIH			
	ML			
Water$_{20}$	RSH			
	LSH			
	RIH			
	AP	0.72		
	ML			
Saliva	LSH			
	RIH			
	AP			
	SI			
	ML			0.71
Yogurt$_5$	RSH			
	LSH			
	RIH			
	SI			
	ML			
Yogurt$_{10}$	RSH			
	LSH			
	RIH			
	ML			
Yogurt$_{20}$	RSH			
	LSH			
	RIH			
	AP		0.72	
	SI			
	ML			

RSH: right suprahyoid muscles; LSH: left suprahyoid muscles; RIH: right infrahyoid muscles; AP: anterior-posterior axis; SI: superior-inferior axis; ML: medial-lateral axis.

Table C.9: AUC_{ROC} of the phonation-related features retrieved in Protocol #3 (speech), Experiment #4.

Feature	Functional	/a/	/e/	/i/	/o/	/u/
APQ	mean	0.72				0.78
	std	0.70				
	skew					0.74
	kurt				0.72	0.75
Jitter	mean	0.80	0.81	0.81	0.83	0.84
	std	0.76	0.80	0.82	0.80	0.81
	skew	0.75	0.72	0.72	0.86	0.81
	kurt	0.75	0.72	0.73	0.85	0.80
Shimmer	mean	0.81	0.70		0.85	0.83
	std	0.79	0.75		0.83	0.82
F_0	std	0.71	0.76	0.75	0.76	0.77
	skew	0.74	0.70	0.74	0.70	0.73
	kurt				0.80	0.72
ΔF_0	std	0.73	0.75	0.77	0.72	0.77
	skew	0.73				0.72
	kurt	0.76	0.72	0.72	0.85	0.79
$\Delta^2 F_0$	std	0.73	0.75	0.76	0.72	0.76
	kurt	0.73	0.70	0.71	0.84	0.77
logE	std	0.70	0.72	0.75	0.74	0.74
	kurt				0.70	

Table C.10: AUC$_{ROC}$ of the articulation-related features retrieved in sustained vowels of Protocol #3 (speech), Experiment #4.

Feature	Functional	/a/	/e/	/i/	/o/	/u/
F_1	std		0.73	0.71	0.76	0.73
F_2	std	0.77			0.74	
ΔF_1	std	0.73	0.77	0.74	0.82	0.80
	kurt			0.73		
ΔF_2	mean	0.75			0.72	0.72
	std	0.81			0.83	
$\Delta^2 F_1$	std	0.75	0.79	0.75	0.79	0.78
$\Delta^2 F_2$	std	0.80			0.80	
MFCC1	std	0.76	0.75	0.78	0.78	0.81
MFCC2	mean			0.76		
MFCC3	mean	0.80				
	std	0.77			0.81	
MFCC4	mean				0.81	0.78
	std		0.70			0.74
	skew				0.70	
	kurt				0.78	0.70
MFCC5	mean		0.83			0.71
	std		0.75	0.77	0.72	0.76
MFCC6	mean					0.72
	skew		0.72	0.71		
MFCC7	mean	0.71		0.75		
	std			0.76		
	skew					0.70
MFCC8	std			0.72		0.71
ΔMFCC1	std	0.85	0.83	0.82	0.83	0.85
ΔMFCC2	mean			0.72		
	skew	0.76				
ΔMFCC3	std	0.70			0.71	
	skew	0.80	0.71		0.79	
	kurt	0.71				
ΔMFCC4	std					0.72
	skew	0.78		0.71		0.71
ΔMFCC5	std		0.71	0.75		
	skew				0.70	
ΔMFCC6	std		0.72			
	kurt	0.73				
ΔMFCC8	std		0.71			
ΔMFCC11	mean	0.71				
Δ^2MFCC1	mean	0.75	0.71	0.70	0.76	0.74
	std	0.85	0.86	0.83	0.82	0.86
Δ^2MFCC2	mean					0.73
	kurt	0.73				
Δ^2MFCC3	mean	0.78			0.76	
Δ^2MFCC9	kurt	0.73				
Δ^2MFCC10	mean			0.70		

Table C.11: AUC$_{\text{ROC}}$ of the articulation-related features retrieved during reading of a text in Protocol #3 (reading of a text), Experiment #4.

Feature	Onset		Offset	
	mean	std	mean	std
BBE_1	0.80	0.81	0.79	0.83
BBE_2	0.73	0.83	0.75	0.85
BBE_3		0.78		0.81
BBE_6		0.71		0.77
BBE_7	0.70			0.74
BBE_8	0.71			
BBE_10	0.71			
BBE_11	0.72			
BBE_12	0.72			
BBE_13	0.76		0.72	
BBE_14	0.77		0.76	
BBE_15	0.72		0.71	
BBE_16	0.74		0.72	
BBE_17	0.71		0.72	
BBE_18			0.71	

Table C.12: AUC$_{\text{ROC}}$ of the DDK and prosody-related features retrieved in Protocol #3, Experiment #4.

Speech dimension	Feature	Functional	AUC$_{\text{ROC}}$
DDK	DDK	mean	0.73
		rate	0.78
	logE	var	0.74
Prosody	logE	kurt	0.74
		skew	0.72
		std	0.83

Appendix D

Unimodal classification performance in Experiment #4

This appendix includes the classification performance retrieved by unimodal scenarios in Experiment #4, with different test set size.

Table D.1: Classification results for unimodal scenarios using classical machine learning algorithms and 10% of the database for test.

Scenario	Feature selection	Classifier	Hyperparameters mode	AUC	F1	Accuracy	Precision	Sensitivity	Specificity
sEMG$_{P1}$	✓	SVM_lin	1	0.78 ± 0.12	0.78 ± 0.11	0.78 ± 0.11	0.81 ± 0.13	0.78 ± 0.16	0.79 ± 0.16
		SVM_RBF	{1, 0.0001}	0.88 ± 0.08	0.85 ± 0.09	0.87 ± 0.08	1.00 ± 0.00	0.75 ± 0.15	1.00 ± 0.00
		SVM_sigmoid	{1, 0.001}	0.89 ± 0.08	0.870 ± 0.1	0.88 ± 0.08	1.00 ± 0.00	0.78 ± 0.16	1.00 ± 0.00
		XGBoost	{2, 1}	0.89 ± 0.12	0.87 ± 0.15	0.89 ± 0.12	0.94 ± 0.09	0.82 ± 0.21	0.95 ± 0.07
		kNN	23	0.85 ± 0.10	0.83 ± 0.12	0.85 ± 0.11	0.92 ± 0.12	0.78 ± 0.16	0.92 ± 0.11
	✗	SVM_lin	1	0.86 ± 0.12	0.85 ± 0.13	0.86 ± 0.12	0.89 ± 0.12	0.82 ± 0.17	0.90 ± 0.10
		SVM_RBF	{10, 0.0001}	0.83 ± 0.11	0.84 ± 0.11	0.83 ± 0.11	0.830 ± 0.1	0.85 ± 0.14	0.82 ± 0.12
		SVM_sigmoid	{1, 0.0001}	0.86 ± 0.12	0.85 ± 0.13	0.86 ± 0.12	0.92 ± 0.12	0.80 ± 0.19	0.92 ± 0.11
		XGBoost	{2, 1}	0.79 ± 0.13	0.74 ± 0.18	0.78 ± 0.13	0.87 ± 0.12	0.68 ± 0.26	0.90 ± 0.11
		kNN	10	0.78 ± 0.12	0.82 ± 0.09	0.78 ± 0.12	0.77 ± 0.17	0.900 ± 0.1	0.670 ± 0.3
sEMG$_{P2}$	✓	SVM_lin	1	0.77 ± 0.13	0.76 ± 0.15	0.77 ± 0.13	0.79 ± 0.14	0.750 ± 0.2	0.79 ± 0.16
		SVM_RBF	{10, 0.0001}	0.82 ± 0.14	0.80 ± 0.16	0.82 ± 0.14	0.87 ± 0.13	0.78 ± 0.22	0.87 ± 0.13
		SVM_sigmoid	{1, 0.01}	0.86 ± 0.14	0.83 ± 0.17	0.86 ± 0.14	0.96 ± 0.09	0.75 ± 0.23	0.98 ± 0.06
		XGBoost	{2, 1}	0.81 ± 0.15	0.78 ± 0.18	0.81 ± 0.15	0.86 ± 0.08	0.72 ± 0.26	0.90 ± 0.06
		kNN	16	0.87 ± 0.09	0.870 ± 0.1	0.87 ± 0.09	0.89 ± 0.06	0.85 ± 0.14	0.90 ± 0.06
	✗	SVM_lin	1	0.82 ± 0.13	0.80 ± 0.14	0.82 ± 0.13	0.87 ± 0.13	0.78 ± 0.22	0.87 ± 0.13
		SVM_RBF	{10, 0.0001}	0.78 ± 0.09	0.76 ± 0.12	0.78 ± 0.09	0.86 ± 0.13	0.72 ± 0.26	0.84 ± 0.14
		SVM_sigmoid	{1, 0.0001}	0.82 ± 0.14	0.80 ± 0.17	0.82 ± 0.14	0.89 ± 0.15	0.75 ± 0.25	0.90 ± 0.14
		XGBoost	{5, 1}	0.84 ± 0.12	0.80 ± 0.16	0.84 ± 0.12	0.94 ± 0.09	0.72 ± 0.22	0.95 ± 0.07
		kNN	7	0.70 ± 0.14	0.760 ± 0.1	0.71 ± 0.14	0.66 ± 0.10	0.90 ± 0.10	0.51 ± 0.19
Acc	✓	SVM_lin	1	0.84 ± 0.07	0.80 ± 0.10	0.84 ± 0.07	0.98 ± 0.06	0.70 ± 0.17	0.98 ± 0.06
		SVM_RBF	{1, 0.001}	0.84 ± 0.08	0.82 ± 0.11	0.84 ± 0.08	0.88 ± 0.07	0.78 ± 0.16	0.90 ± 0.06
		SVM_sigmoid	{1, 0.0001}	0.78 ± 0.17	0.67 ± 0.38	0.77 ± 0.18	0.71 ± 0.42	0.65 ± 0.37	0.90 ± 0.16
		XGBoost	{2, 1}	0.84 ± 0.09	0.82 ± 0.12	0.84 ± 0.09	0.88 ± 0.12	0.80 ± 0.19	0.87 ± 0.15
		kNN	30	0.82 ± 0.05	0.81 ± 0.07	0.82 ± 0.05	0.86 ± 0.02	0.78 ± 0.10	0.87 ± 0.01
	✗	SVM_lin	1	0.77 ± 0.06	0.75 ± 0.06	0.77 ± 0.06	0.85 ± 0.09	0.68 ± 0.07	0.87 ± 0.10
		SVM_RBF	{10, 0.0001}	0.81 ± 0.07	0.80 ± 0.06	0.81 ± 0.07	0.89 ± 0.11	0.72 ± 0.06	0.89 ± 0.12
		SVM_sigmoid	{1, 0.0001}	0.78 ± 0.11	0.800 ± 0.1	0.78 ± 0.11	0.76 ± 0.12	0.850 ± 0.1	0.72 ± 0.16
		XGBoost	{2, 1}	0.81 ± 0.06	0.81 ± 0.06	0.81 ± 0.06	0.81 ± 0.06	0.82 ± 0.07	0.79 ± 0.08
		kNN	29	0.71 ± 0.10	0.65 ± 0.17	0.710 ± 0.1	0.79 ± 0.13	0.57 ± 0.21	0.85 ± 0.10
Speech	✓	SVM_lin	1	0.96 ± 0.03	0.96 ± 0.04	0.96 ± 0.03	0.98 ± 0.05	0.95 ± 0.07	0.98 ± 0.06
		SVM_RBF	{1, 0.001}	0.96 ± 0.03	0.96 ± 0.04	0.96 ± 0.03	0.98 ± 0.05	0.95 ± 0.07	0.98 ± 0.06
		SVM_sigmoid	{1, 0.001}	0.96 ± 0.03	0.96 ± 0.03	0.96 ± 0.03	0.96 ± 0.06	0.98 ± 0.06	0.95 ± 0.07
		XGBoost	{2, 1}	0.96 ± 0.03	0.96 ± 0.04	0.96 ± 0.04	0.98 ± 0.05	0.95 ± 0.07	0.98 ± 0.06
		kNN	9	0.96 ± 0.03	0.96 ± 0.03	0.96 ± 0.03	0.96 ± 0.06	0.98 ± 0.06	0.95 ± 0.07
	✗	SVM_lin	1	0.96 ± 0.03	0.96 ± 0.04	0.96 ± 0.03	0.98 ± 0.05	0.95 ± 0.07	0.98 ± 0.06
		SVM_RBF	{1, 0.001}	0.96 ± 0.03	0.96 ± 0.04	0.96 ± 0.03	0.98 ± 0.05	0.95 ± 0.07	0.98 ± 0.06
		SVM_sigmoid	{1, 0.001}	0.96 ± 0.03	0.96 ± 0.03	0.96 ± 0.03	0.96 ± 0.06	0.98 ± 0.06	0.95 ± 0.07
		XGBoost	{2, 1}	0.96 ± 0.03	0.96 ± 0.04	0.96 ± 0.04	0.98 ± 0.05	0.95 ± 0.07	0.98 ± 0.06
		kNN	9	0.96 ± 0.03	0.96 ± 0.03	0.96 ± 0.03	0.96 ± 0.06	0.98 ± 0.06	0.95 ± 0.07

The hyperparameters are provided in the following order: SVM: $\{C, \gamma\}$ (for linear kernel, γ doesn't apply); XGBoost: {depth, negative/positive balance}; kNN: {neighbors}.

Table D.2: Classification results for unimodal scenarios using classical machine learning algorithms and 15% of the database for test.

Scenario	Feature selection	Classifier	Hyperparameters mode	AUC	F1	Accuracy	Precision	Sensitivity	Specificity
sEMG$_{P1}$	✓	SVM_lin	1	0.84 ± 0.05	0.84 ± 0.04	0.84 ± 0.05	0.87 ± 0.06	0.81 ± 0.07	0.88 ± 0.08
		SVM_RBF	{1, 0.0001}	0.81 ± 0.07	0.81 ± 0.05	0.82 ± 0.07	0.86 ± 0.15	0.79 ± 0.09	0.84 ± 0.20
		SVM_sigmoid	{1, 0.001}	0.84 ± 0.06	0.82 ± 0.09	0.84 ± 0.06	0.93 ± 0.07	0.76 ± 0.16	0.93 ± 0.08
		XGBoost	{2, 1}	0.84 ± 0.03	0.83 ± 0.04	0.84 ± 0.03	0.87 ± 0.05	0.81 ± 0.07	0.88 ± 0.05
		kNN	4	0.85 ± 0.07	0.84 ± 0.06	0.85 ± 0.07	0.91 ± 0.10	0.79 ± 0.04	0.91 ± 0.11
	✗	SVM_lin	1	0.79 ± 0.07	0.79 ± 0.06	0.79 ± 0.08	0.80 ± 0.13	0.79 ± 0.04	0.79 ± 0.15
		SVM_RBF	{10, 0.0001}	0.78 ± 0.03	0.79 ± 0.02	0.78 ± 0.03	0.76 ± 0.04	0.82 ± 0.06	0.74 ± 0.09
		SVM_sigmoid	{1, 0.0001}	0.78 ± 0.06	0.78 ± 0.05	0.78 ± 0.06	0.81 ± 0.10	0.75 ± 0.07	0.81 ± 0.14
		XGBoost	{2, 1}	0.74 ± 0.09	0.75 ± 0.07	0.75 ± 0.08	0.76 ± 0.11	0.75 ± 0.09	0.74 ± 0.18
		kNN	11	0.71 ± 0.10	0.66 ± 0.19	0.71 ± 0.10	0.84 ± 0.19	0.63 ± 0.28	0.78 ± 0.32
sEMG$_{P2}$	✓	SVM_lin	1	0.73 ± 0.11	0.72 ± 0.10	0.73 ± 0.11	0.77 ± 0.13	0.67 ± 0.09	0.79 ± 0.16
		SVM_RBF	{1, 0.001}	0.80 ± 0.11	0.81 ± 0.10	0.80 ± 0.11	0.77 ± 0.10	0.87 ± 0.11	0.74 ± 0.15
		SVM_sigmoid	{1, 0.01}	0.83 ± 0.09	0.84 ± 0.09	0.83 ± 0.09	0.83 ± 0.12	0.86 ± 0.07	0.81 ± 0.14
		XGBoost	{2, 1}	0.82 ± 0.10	0.83 ± 0.10	0.83 ± 0.10	0.80 ± 0.08	0.86 ± 0.14	0.79 ± 0.08
		kNN	13	0.86 ± 0.10	0.87 ± 0.09	0.86 ± 0.10	0.84 ± 0.13	0.92 ± 0.06	0.81 ± 0.17
	✗	SVM_lin	1	0.73 ± 0.15	0.74 ± 0.14	0.73 ± 0.15	0.72 ± 0.15	0.75 ± 0.13	0.70 ± 0.19
		SVM_RBF	{10, 0.0001}	0.69 ± 0.12	0.71 ± 0.11	0.70 ± 0.12	0.69 ± 0.13	0.74 ± 0.14	0.65 ± 0.18
		SVM_sigmoid	{1, 0.0001}	0.74 ± 0.15	0.74 ± 0.14	0.74 ± 0.15	0.75 ± 0.15	0.74 ± 0.14	0.73 ± 0.20
		XGBoost	{2, 1}	0.82 ± 0.11	0.82 ± 0.11	0.82 ± 0.11	0.80 ± 0.11	0.84 ± 0.11	0.79 ± 0.12
		kNN	6	0.75 ± 0.07	0.78 ± 0.04	0.75 ± 0.07	0.71 ± 0.09	0.90 ± 0.11	0.60 ± 0.20
Acc	✓	SVM_lin	1	0.72 ± 0.07	0.74 ± 0.06	0.72 ± 0.06	0.69 ± 0.05	0.80 ± 0.11	0.65 ± 0.06
		SVM_RBF	{1, 0.0001}	0.80 ± 0.06	0.81 ± 0.05	0.80 ± 0.06	0.77 ± 0.10	0.88 ± 0.12	0.72 ± 0.16
		SVM_sigmoid	{1, 0.0001}	0.73 ± 0.16	0.63 ± 0.36	0.72 ± 0.17	0.61 ± 0.35	0.66 ± 0.38	0.80 ± 0.15
		XGBoost	{2, 10}	0.78 ± 0.06	0.79 ± 0.06	0.78 ± 0.06	0.73 ± 0.06	0.88 ± 0.09	0.67 ± 0.10
		kNN	9	0.76 ± 0.07	0.77 ± 0.06	0.76 ± 0.07	0.75 ± 0.11	0.81 ± 0.11	0.71 ± 0.16
	✗	SVM_lin	1	0.76 ± 0.08	0.76 ± 0.07	0.76 ± 0.08	0.75 ± 0.07	0.78 ± 0.11	0.74 ± 0.10
		SVM_RBF	{10, 0.0001}	0.74 ± 0.03	0.74 ± 0.04	0.74 ± 0.04	0.73 ± 0.06	0.78 ± 0.11	0.71 ± 0.10
		SVM_sigmoid	{1, 0.0001}	0.67 ± 0.02	0.71 ± 0.03	0.67 ± 0.01	0.63 ± 0.02	0.81 ± 0.10	0.53 ± 0.07
		XGBoost	{2, 1}	0.71 ± 0.11	0.70 ± 0.14	0.71 ± 0.11	0.68 ± 0.07	0.74 ± 0.22	0.67 ± 0.05
		kNN	8	0.66 ± 0.09	0.69 ± 0.14	0.67 ± 0.08	0.62 ± 0.09	0.78 ± 0.22	0.55 ± 0.14
Speech	✓	SVM_lin	1	0.93 ± 0.02	0.93 ± 0.02	0.93 ± 0.02	0.98 ± 0.04	0.88 ± 0.04	0.98 ± 0.04
		SVM_RBF	{10, 0.001}	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.02	0.97 ± 0.05	0.91 ± 0.06	0.97 ± 0.05
		SVM_sigmoid	{1, 0.001}	0.94 ± 0.04	0.94 ± 0.03	0.94 ± 0.04	0.95 ± 0.07	0.93 ± 0.04	0.95 ± 0.08
		XGBoost	{2, 1}	0.89 ± 0.04	0.90 ± 0.04	0.90 ± 0.05	0.89 ± 0.08	0.91 ± 0.09	0.88 ± 0.10
		kNN	4	0.93 ± 0.05	0.93 ± 0.05	0.93 ± 0.05	0.92 ± 0.06	0.95 ± 0.05	0.92 ± 0.06
	✗	SVM_lin	1	0.93 ± 0.02	0.93 ± 0.02	0.93 ± 0.02	0.98 ± 0.04	0.88 ± 0.04	0.98 ± 0.04
		SVM_RBF	{10, 0.001}	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.02	0.97 ± 0.05	0.91 ± 0.06	0.97 ± 0.05
		SVM_sigmoid	{1, 0.001}	0.94 ± 0.04	0.94 ± 0.03	0.94 ± 0.04	0.95 ± 0.07	0.93 ± 0.04	0.95 ± 0.08
		XGBoost	{2, 1}	0.89 ± 0.04	0.90 ± 0.04	0.90 ± 0.05	0.89 ± 0.08	0.91 ± 0.09	0.88 ± 0.10
		kNN	4	0.93 ± 0.05	0.93 ± 0.05	0.93 ± 0.05	0.92 ± 0.06	0.95 ± 0.05	0.92 ± 0.06

The hyperparameters are provided in the following order: SVM: $\{C, \gamma\}$ (for linear kernel, γ doesn't apply); XGBoost: {depth, negative/positive balance}; kNN: {neighbors}.

Table D.3: Classification results for unimodal scenarios using classical machine learning algorithms and 20% of the database for test.

Scenario	Feature selection	Classifier	Hyperparameters mode	AUC	F1	Accuracy	Precision	Sensitivity	Specificity
$sEMG_{P1}$	✓	SVM_lin	1	0.85 ± 0.06	0.83 ± 0.07	0.85 ± 0.06	0.89 ± 0.09	0.79 ± 0.05	0.91 ± 0.07
		SVM_RBF	{1, 0.0001}	0.82 ± 0.05	0.78 ± 0.07	0.82 ± 0.05	0.94 ± 0.09	0.67 ± 0.08	0.96 ± 0.06
		SVM_sigmoid	{1, 0.0001}	0.75 ± 0.10	0.65 ± 0.18	0.75 ± 0.10	0.94 ± 0.09	0.53 ± 0.22	0.96 ± 0.06
		XGBoost	{2, 1}	0.84 ± 0.03	0.83 ± 0.03	0.83 ± 0.03	0.86 ± 0.11	0.82 ± 0.11	0.85 ± 0.13
		kNN	4	0.85 ± 0.04	0.82 ± 0.05	0.85 ± 0.04	0.95 ± 0.08	0.73 ± 0.06	0.96 ± 0.06
	✗	SVM_lin	1	0.83 ± 0.07	0.82 ± 0.06	0.83 ± 0.07	0.87 ± 0.14	0.78 ± 0.02	0.88 ± 0.14
		SVM_RBF	{10, 0.0001}	0.83 ± 0.06	0.82 ± 0.05	0.83 ± 0.06	0.86 ± 0.12	0.79 ± 0.05	0.87 ± 0.12
		SVM_sigmoid	{1, 0.0001}	0.80 ± 0.06	0.78 ± 0.06	0.80 ± 0.06	0.87 ± 0.14	0.72 ± 0.05	0.88 ± 0.13
		XGBoost	{2, 1}	0.80 ± 0.07	0.78 ± 0.06	0.80 ± 0.07	0.84 ± 0.13	0.74 ± 0.07	0.85 ± 0.13
		kNN	10	0.68 ± 0.12	0.65 ± 0.24	0.67 ± 0.12	0.66 ± 0.13	0.74 ± 0.34	0.61 ± 0.29
$sEMG_{P2}$	✓	SVM_lin	1	0.77 ± 0.08	0.75 ± 0.07	0.76 ± 0.08	0.82 ± 0.15	0.71 ± 0.12	0.83 ± 0.15
		SVM_RBF	{1, 0.001}	0.85 ± 0.07	0.85 ± 0.07	0.85 ± 0.08	0.87 ± 0.16	0.85 ± 0.05	0.85 ± 0.17
		SVM_sigmoid	{1, 0.01}	0.86 ± 0.05	0.85 ± 0.05	0.86 ± 0.06	0.89 ± 0.13	0.82 ± 0.05	0.90 ± 0.13
		XGBoost	{2, 1}	0.84 ± 0.03	0.84 ± 0.02	0.84 ± 0.03	0.87 ± 0.12	0.82 ± 0.07	0.87 ± 0.13
		kNN	13	0.84 ± 0.05	0.84 ± 0.06	0.84 ± 0.05	0.85 ± 0.10	0.83 ± 0.08	0.85 ± 0.10
	✗	SVM_lin	1	0.81 ± 0.03	0.80 ± 0.03	0.81 ± 0.03	0.86 ± 0.10	0.75 ± 0.04	0.88 ± 0.09
		SVM_RBF	{10, 0.0001}	0.80 ± 0.04	0.80 ± 0.03	0.79 ± 0.04	0.79 ± 0.11	0.82 ± 0.07	0.78 ± 0.14
		SVM_sigmoid	{1, 0.0001}	0.84 ± 0.02	0.82 ± 0.03	0.83 ± 0.03	0.88 ± 0.08	0.78 ± 0.08	0.89 ± 0.07
		XGBoost	{2, 1}	0.83 ± 0.05	0.82 ± 0.05	0.83 ± 0.05	0.87 ± 0.11	0.79 ± 0.05	0.88 ± 0.11
		kNN	5	0.77 ± 0.06	0.79 ± 0.06	0.76 ± 0.07	0.70 ± 0.09	0.93 ± 0.10	0.60 ± 0.16
Acc	✓	SVM_lin	1	0.75 ± 0.07	0.74 ± 0.06	0.75 ± 0.07	0.77 ± 0.09	0.72 ± 0.05	0.78 ± 0.13
		SVM_RBF	{1, 0.001}	0.77 ± 0.09	0.78 ± 0.08	0.77 ± 0.09	0.75 ± 0.13	0.82 ± 0.13	0.72 ± 0.20
		SVM_sigmoid	{1, 0.0001}	0.64 ± 0.19	0.59 ± 0.34	0.62 ± 0.20	0.53 ± 0.35	0.73 ± 0.42	0.55 ± 0.50
		XGBoost	{2, 1}	0.87 ± 0.05	0.87 ± 0.05	0.87 ± 0.05	0.86 ± 0.04	0.88 ± 0.08	0.86 ± 0.05
		kNN	29	0.79 ± 0.08	0.79 ± 0.08	0.79 ± 0.08	0.79 ± 0.08	0.79 ± 0.09	0.80 ± 0.08
	✗	SVM_lin	1	0.80 ± 0.08	0.79 ± 0.08	0.80 ± 0.08	0.83 ± 0.11	0.76 ± 0.08	0.83 ± 0.11
		SVM_RBF	{10, 0.0001}	0.81 ± 0.05	0.79 ± 0.06	0.81 ± 0.05	0.82 ± 0.05	0.77 ± 0.08	0.84 ± 0.04
		SVM_sigmoid	{1, 0.0001}	0.78 ± 0.05	0.77 ± 0.05	0.78 ± 0.05	0.78 ± 0.05	0.77 ± 0.08	0.78 ± 0.08
		XGBoost	{2, 1}	0.77 ± 0.06	0.75 ± 0.07	0.77 ± 0.06	0.82 ± 0.08	0.69 ± 0.09	0.85 ± 0.08
		kNN	6	0.73 ± 0.05	0.72 ± 0.08	0.73 ± 0.05	0.73 ± 0.07	0.73 ± 0.18	0.72 ± 0.16
Speech	✓	SVM_lin	1	0.94 ± 0.04	0.94 ± 0.04	0.94 ± 0.03	1.00 ± 0.00	0.89 ± 0.06	1.00 ± 0.00
		SVM_RBF	{10, 0.001}	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.02	0.98 ± 0.03	0.90 ± 0.04	0.99 ± 0.03
		SVM_sigmoid	{1, 0.001}	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.02	0.98 ± 0.03	0.90 ± 0.04	0.99 ± 0.03
		XGBoost	{2, 1}	0.91 ± 0.03	0.90 ± 0.03	0.90 ± 0.03	0.90 ± 0.07	0.90 ± 0.08	0.91 ± 0.07
		kNN	10	0.95 ± 0.02	0.95 ± 0.02	0.95 ± 0.02	1.00 ± 0.00	0.90 ± 0.04	1.00 ± 0.00
	✗	SVM_lin	1	0.94 ± 0.03	0.94 ± 0.04	0.94 ± 0.03	1.00 ± 0.00	0.89 ± 0.06	1.00 ± 0.00
		SVM_RBF	{10, 0.001}	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.02	0.98 ± 0.03	0.90 ± 0.04	0.99 ± 0.03
		SVM_sigmoid	{1, 0.001}	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.02	0.98 ± 0.03	0.90 ± 0.04	0.99 ± 0.03
		XGBoost	{2, 1}	0.91 ± 0.03	0.90 ± 0.03	0.90 ± 0.03	0.90 ± 0.07	0.90 ± 0.08	0.91 ± 0.07
		kNN	10	0.95 ± 0.02	0.95 ± 0.02	0.95 ± 0.02	1.00 ± 0.00	0.90 ± 0.04	1.00 ± 0.00

The hyperparameters are provided in the following order: SVM: $\{C, \gamma\}$ (for linear kernel, γ doesn't apply); XGBoost: {depth, negative/positive balance}; kNN: {neighbors}.

Table D.4: Classification results for unimodal scenarios using classical machine learning algorithms and 25% of the database for test.

Scenario	Feature selection	Classifier	Hyperparameters mode	AUC	F1	Accuracy	Precision	Sensitivity	Specificity
sEMG$_{P1}$	✓	SVM_lin	1	0.84 ± 0.04	0.84 ± 0.04	0.84 ± 0.04	0.85 ± 0.09	0.84 ± 0.08	0.85 ± 0.09
		SVM_RBF	{1, 0.0001}	0.81 ± 0.05	0.79 ± 0.05	0.80 ± 0.05	0.87 ± 0.08	0.73 ± 0.12	0.88 ± 0.08
		SVM_sigmoid	{1, 0.001}	0.79 ± 0.07	0.75 ± 0.11	0.79 ± 0.07	0.90 ± 0.09	0.67 ± 0.16	0.91 ± 0.08
		XGBoost	{2, 1}	0.81 ± 0.07	0.80 ± 0.09	0.81 ± 0.07	0.86 ± 0.06	0.75 ± 0.12	0.88 ± 0.04
		kNN	4	0.83 ± 0.04	0.82 ± 0.05	0.83 ± 0.04	0.88 ± 0.05	0.77 ± 0.08	0.89 ± 0.05
	✗	SVM_lin	1	0.83 ± 0.06	0.81 ± 0.07	0.83 ± 0.06	0.88 ± 0.03	0.75 ± 0.12	0.90 ± 0.03
		SVM_RBF	{10, 0.0001}	0.82 ± 0.03	0.81 ± 0.04	0.81 ± 0.03	0.84 ± 0.05	0.78 ± 0.10	0.85 ± 0.06
		SVM_sigmoid	{1, 0.0001}	0.81 ± 0.05	0.79 ± 0.06	0.81 ± 0.05	0.89 ± 0.03	0.72 ± 0.11	0.91 ± 0.03
		XGBoost	{2, 1}	0.78 ± 0.05	0.76 ± 0.06	0.78 ± 0.05	0.81 ± 0.07	0.73 ± 0.11	0.83 ± 0.08
		kNN	6	0.69 ± 0.08	0.67 ± 0.16	0.68 ± 0.08	0.73 ± 0.18	0.73 ± 0.32	0.64 ± 0.32
sEMG$_{P2}$	✓	SVM_lin	1	0.79 ± 0.02	0.79 ± 0.03	0.79 ± 0.02	0.81 ± 0.06	0.77 ± 0.08	0.82 ± 0.08
		SVM_RBF	{1, 0.01}	0.83 ± 0.07	0.83 ± 0.07	0.83 ± 0.07	0.86 ± 0.12	0.80 ± 0.08	0.86 ± 0.12
		SVM_sigmoid	{1, 0.001}	0.85 ± 0.05	0.84 ± 0.06	0.85 ± 0.06	0.89 ± 0.08	0.80 ± 0.08	0.90 ± 0.07
		XGBoost	{2, 1}	0.83 ± 0.05	0.83 ± 0.05	0.83 ± 0.05	0.85 ± 0.06	0.82 ± 0.08	0.85 ± 0.07
		kNN	12	0.85 ± 0.06	0.84 ± 0.07	0.85 ± 0.06	0.88 ± 0.08	0.80 ± 0.08	0.89 ± 0.07
	✗	SVM_lin	1	0.83 ± 0.05	0.82 ± 0.06	0.82 ± 0.05	0.84 ± 0.07	0.81 ± 0.12	0.85 ± 0.09
		SVM_RBF	{10, 0.0001}	0.80 ± 0.06	0.80 ± 0.06	0.80 ± 0.06	0.80 ± 0.08	0.81 ± 0.10	0.79 ± 0.10
		SVM_sigmoid	{1, 0.0001}	0.82 ± 0.04	0.81 ± 0.05	0.81 ± 0.05	0.84 ± 0.10	0.79 ± 0.10	0.84 ± 0.12
		XGBoost	{2, 1}	0.84 ± 0.07	0.83 ± 0.07	0.84 ± 0.07	0.87 ± 0.09	0.79 ± 0.10	0.88 ± 0.09
		kNN	4	0.78 ± 0.05	0.79 ± 0.06	0.78 ± 0.06	0.75 ± 0.09	0.86 ± 0.11	0.71 ± 0.13
Acc	✓	SVM_lin	1	0.80 ± 0.05	0.79 ± 0.07	0.79 ± 0.05	0.81 ± 0.14	0.78 ± 0.08
		SVM_RBF	{1, 0.0001}	0.73 ± 0.11	0.77 ± 0.08	0.72 ± 0.13	0.68 ± 0.13	0.92 ± 0.08	0.54 ± 0.28
		SVM_sigmoid	{1, 0.0001}	0.75 ± 0.14	0.65 ± 0.37	0.74 ± 0.15	0.61 ± 0.34	0.71 ± 0.40	0.78 ± 0.13
		XGBoost	{2, 1}	0.88 ± 0.05	0.88 ± 0.05	0.88 ± 0.05	0.87 ± 0.06	0.90 ± 0.06	0.86 ± 0.07
		kNN	11	0.85 ± 0.05	0.86 ± 0.04	0.85 ± 0.05	0.82 ± 0.05	0.89 ± 0.03	0.80 ± 0.08
	✗	SVM_lin	1	0.79 ± 0.05	0.78 ± 0.07	0.79 ± 0.05	0.83 ± 0.07	0.75 ± 0.14	0.84 ± 0.08
		SVM_RBF	{10, 0.0001}	0.81 ± 0.04	0.79 ± 0.06	0.80 ± 0.05	0.83 ± 0.09	0.78 ± 0.15	0.83 ± 0.11
		SVM_sigmoid	{1, 0.0001}	0.75 ± 0.07	0.78 ± 0.05	0.75 ± 0.08	0.74 ± 0.15	0.87 ± 0.16	0.64 ± 0.27
		XGBoost	{2, 1}	0.79 ± 0.07	0.79 ± 0.08	0.79 ± 0.07	0.78 ± 0.05	0.81 ± 0.14	0.77 ± 0.06
		kNN	6	0.79 ± 0.06	0.77 ± 0.09	0.79 ± 0.07	0.84 ± 0.09	0.74 ± 0.18	0.84 ± 0.14
Speech	✓	SVM_lin	1	0.95 ± 0.05	0.94 ± 0.06	0.95 ± 0.05	0.99 ± 0.03	0.90 ± 0.11	0.99 ± 0.02
		SVM_RBF	{1, 0.001}	0.96 ± 0.03	0.96 ± 0.04	0.96 ± 0.03	1.00 ± 0.00	0.92 ± 0.06	1.00 ± 0.00
		SVM_sigmoid	{1, 0.001}	0.96 ± 0.03	0.96 ± 0.04	0.96 ± 0.03	1.00 ± 0.00	0.92 ± 0.06	1.00 ± 0.00
		XGBoost	{2, 1}	0.90 ± 0.04	0.89 ± 0.05	0.90 ± 0.04	0.89 ± 0.05	0.90 ± 0.11	0.89 ± 0.07
		kNN	5	0.93 ± 0.05	0.93 ± 0.06	0.93 ± 0.05	0.96 ± 0.05	0.90 ± 0.07	0.97 ± 0.05
	✗	SVM_lin	1	0.95 ± 0.05	0.94 ± 0.06	0.95 ± 0.05	0.99 ± 0.03	0.90 ± 0.11	0.99 ± 0.02
		SVM_RBF	{1, 0.001}	0.96 ± 0.03	0.96 ± 0.04	0.96 ± 0.03	1.00 ± 0.00	0.92 ± 0.06	1.00 ± 0.00
		SVM_sigmoid	{1, 0.001}	0.96 ± 0.03	0.96 ± 0.04	0.96 ± 0.03	1.00 ± 0.00	0.92 ± 0.06	1.00 ± 0.00
		XGBoost	{2, 1}	0.90 ± 0.04	0.89 ± 0.05	0.90 ± 0.04	0.89 ± 0.05	0.90 ± 0.11	0.89 ± 0.07
		kNN	5	0.93 ± 0.05	0.93 ± 0.06	0.93 ± 0.05	0.96 ± 0.05	0.90 ± 0.07	0.97 ± 0.05

The hyperparameters are provided in the following order: SVM: $\{C, \gamma\}$ (for linear kernel, γ doesn't apply); XGBoost: {depth, negative/positive balance}; kNN: {neighbors}.

Table D.5: Classification results for unimodal scenarios using classical machine learning algorithms and 30% of the database for test.

Scenario	Feature selection	Classifier	Hyperparameters mode	AUC	F1	Accuracy	Precision	Sensitivity	Specificity
$sEMG_{P1}$	✓	SVM_lin	1	0.83 ± 0.06	0.83 ± 0.06	0.82 ± 0.07	0.79 ± 0.10	0.88 ± 0.02	0.77 ± 0.12
		SVM_RBF	{1, 0.0001}	0.80 ± 0.04	0.79 ± 0.04	0.80 ± 0.04	0.84 ± 0.11	0.76 ± 0.11	0.85 ± 0.13
		SVM_sigmoid	{1, 0.001}	0.84 ± 0.08	0.80 ± 0.11	0.84 ± 0.07	0.97 ± 0.04	0.69 ± 0.16	0.98 ± 0.03
		XGBoost	{2, 1}	0.86 ± 0.05	0.85 ± 0.05	0.86 ± 0.05	0.88 ± 0.08	0.82 ± 0.04	0.89 ± 0.08
		kNN	9	0.83 ± 0.06	0.81 ± 0.07	0.83 ± 0.06	0.90 ± 0.08	0.73 ± 0.07	0.93 ± 0.06
	✗	SVM_lin	1	0.84 ± 0.06	0.83 ± 0.06	0.84 ± 0.06	0.87 ± 0.08	0.80 ± 0.06	0.89 ± 0.07
		SVM_RBF	{10, 0.0001}	0.81 ± 0.05	0.81 ± 0.05	0.81 ± 0.05	0.81 ± 0.07	0.81 ± 0.04	0.81 ± 0.07
		SVM_sigmoid	{1, 0.0001}	0.84 ± 0.05	0.82 ± 0.06	0.84 ± 0.05	0.90 ± 0.10	0.75 ± 0.06	0.92 ± 0.08
		XGBoost	{2, 1}	0.80 ± 0.03	0.79 ± 0.03	0.80 ± 0.03	0.82 ± 0.06	0.77 ± 0.04	0.84 ± 0.06
		kNN	8	0.67 ± 0.12	0.59 ± 0.28	0.68 ± 0.12	0.80 ± 0.18	0.62 ± 0.36	0.72 ± 0.36
$sEMG_{P2}$	✓	SVM_lin	1	0.79 ± 0.05	0.79 ± 0.04	0.79 ± 0.05	0.78 ± 0.07	0.80 ± 0.05	0.78 ± 0.09
		SVM_RBF	{10, 0.001}	0.81 ± 0.07	0.81 ± 0.08	0.81 ± 0.07	0.79 ± 0.04	0.82 ± 0.11	0.80 ± 0.02
		SVM_sigmoid	{1, 0.001}	0.85 ± 0.05	0.85 ± 0.06	0.85 ± 0.05	0.84 ± 0.03	0.85 ± 0.09	0.85 ± 0.03
		XGBoost	{2, 1}	0.83 ± 0.03	0.83 ± 0.04	0.83 ± 0.03	0.82 ± 0.04	0.84 ± 0.06	0.82 ± 0.03
		kNN	5	0.85 ± 0.06	0.84 ± 0.07	0.85 ± 0.06	0.84 ± 0.04	0.85 ± 0.10	0.85 ± 0.04
	✗	SVM_lin	1	0.82 ± 0.06	0.81 ± 0.05	0.82 ± 0.06	0.84 ± 0.08	0.79 ± 0.04	0.84 ± 0.08
		SVM_RBF	{10, 0.001}	0.78 ± 0.07	0.79 ± 0.06	0.78 ± 0.07	0.75 ± 0.09	0.84 ± 0.05	0.72 ± 0.12
		SVM_sigmoid	{1, 0.0001}	0.85 ± 0.02	0.85 ± 0.03	0.85 ± 0.02	0.85 ± 0.02	0.84 ± 0.05	0.86 ± 0.03
		XGBoost	{2, 1}	0.85 ± 0.01	0.84 ± 0.02	0.85 ± 0.01	0.85 ± 0.02	0.84 ± 0.04	0.86 ± 0.02
		kNN	4	0.77 ± 0.04	0.79 ± 0.03	0.76 ± 0.04	0.70 ± 0.03	0.93 ± 0.06	0.61 ± 0.09
Acc	✓	SVM_lin	1	0.76 ± 0.07	0.74 ± 0.07	0.76 ± 0.07	0.80 ± 0.11	0.70 ± 0.08	0.82 ± 0.13
		SVM_RBF	{1, 0.0001}	0.74 ± 0.10	0.77 ± 0.07	0.74 ± 0.11	0.71 ± 0.11	0.87 ± 0.11	0.61 ± 0.28
		SVM_sigmoid	{1, 0.001}	0.72 ± 0.14	0.63 ± 0.36	0.72 ± 0.15	0.60 ± 0.34	0.69 ± 0.39	0.76 ± 0.21
		XGBoost	{2, 1}	0.79 ± 0.07	0.81 ± 0.04	0.79 ± 0.07	0.77 ± 0.11	0.86 ± 0.09	0.72 ± 0.19
		kNN	30	0.80 ± 0.04	0.80 ± 0.04	0.80 ± 0.04	0.77 ± 0.06	0.85 ± 0.07	0.75 ± 0.08
	✗	SVM_lin	1	0.76 ± 0.06	0.76 ± 0.05	0.76 ± 0.06	0.77 ± 0.10	0.75 ± 0.08	0.76 ± 0.11
		SVM_RBF	{1, 0.0001}	0.76 ± 0.05	0.77 ± 0.04	0.76 ± 0.05	0.75 ± 0.10	0.80 ± 0.09	0.72 ± 0.15
		SVM_sigmoid	{1, 0.0001}	0.70 ± 0.05	0.73 ± 0.03	0.69 ± 0.05	0.65 ± 0.06	0.86 ± 0.06	0.53 ± 0.12
		XGBoost	{2, 1}	0.74 ± 0.04	0.74 ± 0.05	0.74 ± 0.04	0.75 ± 0.10	0.76 ± 0.12	0.73 ± 0.15
		kNN	24	0.68 ± 0.09	0.66 ± 0.10	0.69 ± 0.09	0.72 ± 0.14	0.62 ± 0.13	0.75 ± 0.16
Speech	✓	SVM_lin	1	0.95 ± 0.03	0.95 ± 0.03	0.95 ± 0.03	0.97 ± 0.03	0.94 ± 0.06	0.97 ± 0.03
		SVM_RBF	{10, 0.001}	0.95 ± 0.03	0.95 ± 0.03	0.95 ± 0.03	0.97 ± 0.03	0.94 ± 0.05	0.97 ± 0.03
		SVM_sigmoid	{1, 0.001}	0.96 ± 0.03	0.96 ± 0.03	0.96 ± 0.03	0.97 ± 0.03	0.95 ± 0.06	0.97 ± 0.03
		XGBoost	{2, 1}	0.93 ± 0.06	0.92 ± 0.07	0.92 ± 0.06	0.93 ± 0.03	0.92 ± 0.12	0.93 ± 0.04
		kNN	10	0.96 ± 0.03	0.96 ± 0.03	0.96 ± 0.03	0.98 ± 0.03	0.94 ± 0.04	0.98 ± 0.03
	✗	SVM_lin	1	0.95 ± 0.03	0.95 ± 0.03	0.95 ± 0.03	0.97 ± 0.03	0.94 ± 0.06	0.97 ± 0.03
		SVM_RBF	{10, 0.001}	0.95 ± 0.03	0.95 ± 0.03	0.95 ± 0.03	0.97 ± 0.03	0.94 ± 0.05	0.97 ± 0.03
		SVM_sigmoid	{1, 0.001}	0.96 ± 0.03	0.96 ± 0.03	0.96 ± 0.03	0.97 ± 0.03	0.95 ± 0.06	0.97 ± 0.03
		XGBoost	{2, 1}	0.93 ± 0.06	0.92 ± 0.07	0.92 ± 0.06	0.93 ± 0.03	0.92 ± 0.12	0.93 ± 0.04
		kNN	10	0.96 ± 0.03	0.96 ± 0.03	0.96 ± 0.03	0.98 ± 0.03	0.94 ± 0.04	0.98 ± 0.03

The hyperparameters are provided in the following order: SVM: $\{C, \gamma\}$ (for linear kernel, γ doesn't apply); XGBoost: {depth, negative/positive balance}; kNN: {neighbors}.

Table D.6: Classification results for unimodal scenarios using classical machine learning algorithms and 35% of the database for test.

Scenario	Feature selection	Classifier	Hyperparameters mode	AUC	F1	Accuracy	Precision	Sensitivity	Specificity
sEMG$_{P1}$	✓	SVM_lin	1	0.82 ± 0.05	0.83 ± 0.04	0.83 ± 0.05	0.84 ± 0.07	0.82 ± 0.07	0.82 ± 0.13
		SVM_RBF	{1, 0.0001}	0.83 ± 0.03	0.82 ± 0.04	0.83 ± 0.03	0.87 ± 0.08	0.79 ± 0.11	0.86 ± 0.12
		SVM_sigmoid	{1, 0.001}	0.80 ± 0.14	0.72 ± 0.30	0.79 ± 0.17	0.97 ± 0.04	0.63 ± 0.30	0.97 ± 0.04
		XGBoost	{2, 1}	0.83 ± 0.06	0.83 ± 0.08	0.83 ± 0.07	0.84 ± 0.05	0.82 ± 0.12	0.83 ± 0.06
		kNN	18	0.83 ± 0.04	0.83 ± 0.04	0.84 ± 0.04	0.88 ± 0.04	0.79 ± 0.06	0.88 ± 0.05
	✗	SVM_lin	1	0.86 ± 0.07	0.86 ± 0.06	0.86 ± 0.07	0.89 ± 0.10	0.83 ± 0.07	0.88 ± 0.11
		SVM_RBF	{10, 0.0001}	0.83 ± 0.04	0.83 ± 0.05	0.83 ± 0.04	0.84 ± 0.05	0.82 ± 0.07	0.84 ± 0.06
		SVM_sigmoid	{1, 0.0001}	0.82 ± 0.04	0.81 ± 0.04	0.82 ± 0.05	0.86 ± 0.07	0.77 ± 0.04	0.86 ± 0.08
		XGBoost	{2, 1}	0.77 ± 0.06	0.77 ± 0.06	0.77 ± 0.06	0.80 ± 0.09	0.75 ± 0.07	0.79 ± 0.11
		kNN	6	0.70 ± 0.12	0.65 ± 0.25	0.69 ± 0.13	0.79 ± 0.18	0.68 ± 0.33	0.72 ± 0.32
sEMG$_{P2}$	✓	SVM_lin	1	0.77 ± 0.05	0.77 ± 0.05	0.77 ± 0.05	0.80 ± 0.10	0.74 ± 0.08	0.79 ± 0.12
		SVM_RBF	{10, 0.001}	0.83 ± 0.07	0.84 ± 0.06	0.84 ± 0.07	0.84 ± 0.08	0.84 ± 0.07	0.83 ± 0.11
		SVM_sigmoid	{1, 0.01}	0.87 ± 0.05	0.86 ± 0.05	0.87 ± 0.05	0.89 ± 0.05	0.84 ± 0.07	0.89 ± 0.05
		XGBoost	{2, 1}	0.84 ± 0.05	0.84 ± 0.05	0.84 ± 0.05	0.85 ± 0.07	0.84 ± 0.07	0.84 ± 0.08
		kNN	12	0.86 ± 0.07	0.86 ± 0.08	0.86 ± 0.07	0.88 ± 0.06	0.85 ± 0.11	0.87 ± 0.08
	✗	SVM_lin	1	0.82 ± 0.09	0.82 ± 0.09	0.82 ± 0.09	0.85 ± 0.13	0.79 ± 0.10	0.85 ± 0.14
		SVM_RBF	{10, 0.001}	0.80 ± 0.06	0.81 ± 0.06	0.81 ± 0.06	0.81 ± 0.12	0.84 ± 0.13	0.77 ± 0.17
		SVM_sigmoid	{1, 0.0001}	0.84 ± 0.06	0.84 ± 0.05	0.84 ± 0.06	0.87 ± 0.10	0.81 ± 0.07	0.87 ± 0.12
		XGBoost	{5, 1}	0.85 ± 0.07	0.85 ± 0.08	0.85 ± 0.07	0.87 ± 0.08	0.84 ± 0.12	0.86 ± 0.09
		kNN	5	0.75 ± 0.05	0.76 ± 0.09	0.75 ± 0.05	0.76 ± 0.11	0.80 ± 0.21	0.71 ± 0.19
Acc	✓	SVM_lin	1	0.69 ± 0.08	0.71 ± 0.07	0.69 ± 0.08	0.70 ± 0.06	0.73 ± 0.12	0.66 ± 0.13
		SVM_RBF	{1, 0.0001}	0.65 ± 0.10	0.62 ± 0.22	0.64 ± 0.11	0.73 ± 0.18	0.69 ± 0.35	0.61 ± 0.35
		SVM_sigmoid	{1, 0.001}	0.72 ± 0.13	0.62 ± 0.35	0.71 ± 0.14	0.61 ± 0.35	0.65 ± 0.39	0.79 ± 0.14
		XGBoost	{2, 1}	0.79 ± 0.03	0.80 ± 0.03	0.79 ± 0.03	0.77 ± 0.06	0.85 ± 0.11	0.72 ± 0.10
		kNN	30	0.75 ± 0.06	0.76 ± 0.07	0.75 ± 0.06	0.74 ± 0.06	0.81 ± 0.16	0.69 ± 0.13
	✗	SVM_lin	1	0.76 ± 0.05	0.77 ± 0.04	0.76 ± 0.05	0.76 ± 0.05	0.78 ± 0.08	0.74 ± 0.08
		SVM_RBF	{1, 0.0001}	0.76 ± 0.05	0.76 ± 0.05	0.75 ± 0.05	0.75 ± 0.05	0.79 ± 0.12	0.72 ± 0.08
		SVM_sigmoid	{1, 0.0001}	0.72 ± 0.04	0.75 ± 0.03	0.72 ± 0.04	0.69 ± 0.06	0.85 ± 0.11	0.59 ± 0.11
		XGBoost	{2, 1}	0.75 ± 0.08	0.76 ± 0.07	0.75 ± 0.08	0.76 ± 0.07	0.76 ± 0.11	0.74 ± 0.10
		kNN	8	0.66 ± 0.07	0.65 ± 0.14	0.65 ± 0.08	0.66 ± 0.07	0.69 ± 0.23	0.63 ± 0.16
Speech	✓	SVM_lin	1	0.91 ± 0.05	0.92 ± 0.03	0.92 ± 0.04	0.91 ± 0.06	0.93 ± 0.04	0.89 ± 0.11
		SVM_RBF	{1, 0.001}	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.02	0.95 ± 0.01	0.93 ± 0.04	0.95 ± 0.02
		SVM_sigmoid	{1, 0.001}	0.95 ± 0.03	0.95 ± 0.03	0.95 ± 0.03	0.96 ± 0.05	0.93 ± 0.04	0.96 ± 0.05
		XGBoost	{2, 1}	0.92 ± 0.02	0.92 ± 0.01	0.92 ± 0.02	0.91 ± 0.04	0.93 ± 0.04	0.90 ± 0.05
		kNN	5	0.94 ± 0.03	0.94 ± 0.03	0.94 ± 0.03	0.95 ± 0.03	0.93 ± 0.03	0.95 ± 0.04
	✗	SVM_lin	1	0.91 ± 0.05	0.92 ± 0.03	0.92 ± 0.04	0.91 ± 0.06	0.93 ± 0.04	0.89 ± 0.11
		SVM_RBF	{1, 0.001}	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.02	0.95 ± 0.01	0.93 ± 0.04	0.95 ± 0.02
		SVM_sigmoid	{1, 0.001}	0.95 ± 0.03	0.95 ± 0.03	0.95 ± 0.03	0.96 ± 0.05	0.93 ± 0.04	0.96 ± 0.05
		XGBoost	{2, 1}	0.92 ± 0.02	0.92 ± 0.01	0.92 ± 0.02	0.91 ± 0.04	0.93 ± 0.04	0.90 ± 0.05
		kNN	5	0.94 ± 0.03	0.94 ± 0.03	0.94 ± 0.03	0.95 ± 0.03	0.93 ± 0.03	0.95 ± 0.04

The hyperparameters are provided in the following order: SVM: {C, γ} (for linear kernel, γ doesn't apply); XGBoost: {depth, negative/positive balance}; kNN: {neighbors}.

Table D.7: Classification results for unimodal scenarios using classical machine learning algorithms and 40% of the database for test.

Scenario	Feature selection	Classifier	Hyperparameters mode	AUC	F1	Accuracy	Precision	Sensitivity	Specificity
sEMG$_{P1}$	✓	SVM_lin	1	0.83 ± 0.04	0.84 ± 0.05	0.84 ± 0.04	0.83 ± 0.05	0.85 ± 0.07	0.82 ± 0.06
		SVM_RBF	{1, 0.0001}	0.85 ± 0.07	0.83 ± 0.10	0.84 ± 0.08	0.93 ± 0.06	0.76 ± 0.16	0.93 ± 0.06
		SVM_sigmoid	{1, 0.001}	0.80 ± 0.15	0.72 ± 0.25	0.79 ± 0.15	0.95 ± 0.05	0.63 ± 0.31	0.96 ± 0.04
		XGBoost	{2, 1}	0.84 ± 0.03	0.84 ± 0.04	0.84 ± 0.03	0.86 ± 0.05	0.83 ± 0.09	0.85 ± 0.06
		kNN	14	0.87 ± 0.05	0.87 ± 0.06	0.87 ± 0.05	0.91 ± 0.06	0.83 ± 0.09	0.92 ± 0.06
	✗	SVM_lin	1	0.85 ± 0.05	0.85 ± 0.05	0.85 ± 0.05	0.89 ± 0.07	0.82 ± 0.08	0.89 ± 0.08
		SVM_RBF	{10, 0.0001}	0.85 ± 0.03	0.85 ± 0.03	0.85 ± 0.03	0.86 ± 0.05	0.85 ± 0.07	0.85 ± 0.05
		SVM_sigmoid	{1, 0.0001}	0.87 ± 0.03	0.87 ± 0.04	0.87 ± 0.03	0.93 ± 0.04	0.81 ± 0.07	0.94 ± 0.03
		XGBoost	{2, 1}	0.80 ± 0.03	0.80 ± 0.02	0.80 ± 0.02	0.81 ± 0.08	0.81 ± 0.07	0.79 ± 0.10
		kNN	5	0.68 ± 0.12	0.65 ± 0.17	0.67 ± 0.12	0.79 ± 0.20	0.68 ± 0.33	0.67 ± 0.42
sEMG$_{P2}$	✓	SVM_lin	1	0.77 ± 0.07	0.80 ± 0.05	0.77 ± 0.07	0.76 ± 0.10	0.85 ± 0.04	0.70 ± 0.14
		SVM_RBF	{1, 0.01}	0.83 ± 0.02	0.84 ± 0.02	0.84 ± 0.02	0.82 ± 0.04	0.87 ± 0.02	0.80 ± 0.04
		SVM_sigmoid	{1, 0.001}	0.84 ± 0.04	0.85 ± 0.04	0.84 ± 0.04	0.83 ± 0.06	0.87 ± 0.02	0.81 ± 0.07
		XGBoost	{2, 1}	0.82 ± 0.04	0.83 ± 0.04	0.82 ± 0.04	0.82 ± 0.06	0.85 ± 0.05	0.79 ± 0.09
		kNN	9	0.83 ± 0.06	0.83 ± 0.06	0.83 ± 0.06	0.84 ± 0.08	0.83 ± 0.06	0.82 ± 0.09
	✗	SVM_lin	1	0.83 ± 0.06	0.84 ± 0.05	0.83 ± 0.06	0.85 ± 0.10	0.83 ± 0.04	0.83 ± 0.12
		SVM_RBF	{1, 0.001}	0.80 ± 0.05	0.83 ± 0.04	0.81 ± 0.05	0.77 ± 0.07	0.90 ± 0.05	0.71 ± 0.11
		SVM_sigmoid	{1, 0.0001}	0.80 ± 0.05	0.80 ± 0.05	0.80 ± 0.05	0.81 ± 0.07	0.80 ± 0.07	0.80 ± 0.07
		XGBoost	{2, 1}	0.81 ± 0.08	0.82 ± 0.06	0.81 ± 0.08	0.81 ± 0.12	0.85 ± 0.03	0.77 ± 0.16
		kNN	9	0.74 ± 0.09	0.73 ± 0.15	0.74 ± 0.09	0.75 ± 0.10	0.77 ± 0.25	0.71 ± 0.18
Acc	✓	SVM_lin	1	0.74 ± 0.05	0.76 ± 0.04	0.74 ± 0.05	0.73 ± 0.09	0.80 ± 0.07	0.67 ± 0.17
		SVM_RBF	{1, 0.0001}	0.62 ± 0.12	0.39 ± 0.36	0.61 ± 0.14	0.68 ± 0.41	0.39 ± 0.43	0.86 ± 0.20
		SVM_sigmoid	{1, 0.001}	0.79 ± 0.04	0.81 ± 0.04	0.79 ± 0.04	0.78 ± 0.05	0.84 ± 0.07	0.75 ± 0.06
		XGBoost	{2, 1}	0.79 ± 0.04	0.79 ± 0.03	0.79 ± 0.04	0.79 ± 0.05	0.80 ± 0.05	0.77 ± 0.08
		kNN	30	0.79 ± 0.04	0.80 ± 0.05	0.79 ± 0.04	0.80 ± 0.02	0.80 ± 0.08	0.79 ± 0.03
	✗	SVM_lin	1	0.74 ± 0.06	0.75 ± 0.05	0.74 ± 0.06	0.75 ± 0.07	0.76 ± 0.05	0.72 ± 0.10
		SVM_RBF	{10, 0.0001}	0.78 ± 0.03	0.79 ± 0.04	0.78 ± 0.03	0.77 ± 0.03	0.81 ± 0.09	0.75 ± 0.05
		SVM_sigmoid	{1, 0.0001}	0.73 ± 0.05	0.76 ± 0.04	0.73 ± 0.05	0.72 ± 0.07	0.82 ± 0.09	0.65 ± 0.14
		XGBoost	{2, 1}	0.73 ± 0.03	0.74 ± 0.02	0.73 ± 0.03	0.73 ± 0.05	0.75 ± 0.02	0.71 ± 0.08
		kNN	4	0.64 ± 0.05	0.60 ± 0.16	0.64 ± 0.06	0.69 ± 0.08	0.61 ± 0.31	0.67 ± 0.26
Speech	✓	SVM_lin	1	0.93 ± 0.03	0.92 ± 0.03	0.93 ± 0.03	0.95 ± 0.04	0.91 ± 0.08	0.95 ± 0.04
		SVM_RBF	{1, 0.001}	0.93 ± 0.07	0.93 ± 0.07	0.93 ± 0.07	0.93 ± 0.07	0.93 ± 0.07	0.93 ± 0.08
		SVM_sigmoid	{1, 0.001}	0.95 ± 0.04	0.95 ± 0.04	0.95 ± 0.04	0.97 ± 0.04	0.93 ± 0.07	0.97 ± 0.04
		XGBoost	{2, 1}	0.90 ± 0.04	0.91 ± 0.04	0.90 ± 0.04	0.90 ± 0.05	0.91 ± 0.07	0.89 ± 0.05
		kNN	4	0.92 ± 0.04	0.92 ± 0.04	0.92 ± 0.04	0.95 ± 0.03	0.90 ± 0.06	0.95 ± 0.03
	✗	SVM_lin	1	0.93 ± 0.03	0.92 ± 0.03	0.93 ± 0.03	0.95 ± 0.04	0.91 ± 0.08	0.95 ± 0.04
		SVM_RBF	{1, 0.001}	0.93 ± 0.07	0.93 ± 0.07	0.93 ± 0.07	0.93 ± 0.07	0.93 ± 0.07	0.93 ± 0.08
		SVM_sigmoid	{1, 0.001}	0.95 ± 0.04	0.95 ± 0.04	0.95 ± 0.04	0.97 ± 0.04	0.93 ± 0.07	0.97 ± 0.04
		XGBoost	{2, 1}	0.90 ± 0.04	0.91 ± 0.04	0.90 ± 0.04	0.90 ± 0.05	0.91 ± 0.07	0.89 ± 0.05
		kNN	4	0.92 ± 0.04	0.92 ± 0.04	0.92 ± 0.04	0.95 ± 0.03	0.90 ± 0.06	0.95 ± 0.03

The hyperparameters are provided in the following order: SVM: $\{C, \gamma\}$ (for linear kernel, γ doesn't apply); XGBoost: {depth, negative/positive balance}; kNN: {neighbors}.

Appendix E

Multimodal classification performance in Experiment #4

This appendix includes the classification performance retrieved by multimodal scenarios in Experiment #4, with different test set size.

Table E.1: Classification results for multimodal scenarios using classical machine learning algorithms and 10% of the database for test.

Scenario	Feature selection	Classifier	Hyperparameters mode	AUC	F1	Accuracy	Precision	Sensitivity	Specificity
sEMG$_{P1}$ + Acc + Speech	✓	SVM_lin	1	0.96 ± 0.06	0.96 ± 0.06	0.96 ± 0.06	1.00 ± 0.00	0.92 ± 0.11	1.00 ± 0.00
		SVM_RBF	{1, 0.001}	0.98 ± 0.03	0.97 ± 0.04	0.97 ± 0.04	1.00 ± 0.00	0.95 ± 0.07	1.00 ± 0.00
		SVM_sigmoid	{1, 0.0001}	0.96 ± 0.06	0.96 ± 0.06	0.96 ± 0.06	1.00 ± 0.00	0.92 ± 0.11	1.00 ± 0.00
		XGBoost	{2, 1}	0.94 ± 0.06	0.94 ± 0.06	0.94 ± 0.06	0.94 ± 0.09	0.95 ± 0.07	0.92 ± 0.11
		kNN	30	0.98 ± 0.06	0.97 ± 0.06	0.98 ± 0.06	1.00 ± 0.00	0.95 ± 0.11	1.00 ± 0.00
	✗	SVM_lin	1	0.91 ± 0.07	0.90 ± 0.09	0.91 ± 0.07	0.98 ± 0.05	0.85 ± 0.16	0.98 ± 0.06
		SVM_RBF	{1, 0.0001}	0.89 ± 0.05	0.88 ± 0.06	0.89 ± 0.05	0.91 ± 0.09	0.88 ± 0.12	0.90 ± 0.10
		SVM_sigmoid	{1, 0.0001}	0.91 ± 0.07	0.90 ± 0.09	0.91 ± 0.07	0.98 ± 0.05	0.85 ± 0.16	0.98 ± 0.06
		XGBoost	{2, 1}	0.96 ± 0.06	0.96 ± 0.06	0.96 ± 0.06	0.95 ± 0.06	0.98 ± 0.06	0.95 ± 0.07
		kNN	30	0.86 ± 0.08	0.84 ± 0.10	0.86 ± 0.08	0.97 ± 0.07	0.75 ± 0.12	0.98 ± 0.06
sEMG$_{P2}$ + Acc + Speech	✓	SVM_lin	1	0.98 ± 0.06	0.97 ± 0.06	0.98 ± 0.06	1.00 ± 0.00	0.95 ± 0.11	1.00 ± 0.00
		SVM_RBF	{1, 0.001}	0.96 ± 0.03	0.96 ± 0.04	0.96 ± 0.04	0.98 ± 0.05	0.95 ± 0.07	0.98 ± 0.06
		SVM_sigmoid	{1, 0.001}	0.98 ± 0.03	0.97 ± 0.03	0.98 ± 0.03	0.98 ± 0.05	0.98 ± 0.06	0.98 ± 0.06
		XGBoost	{2, 1}	0.96 ± 0.03	0.96 ± 0.04	0.96 ± 0.04	0.98 ± 0.05	0.95 ± 0.07	0.98 ± 0.06
		kNN	11	0.96 ± 0.03	0.96 ± 0.04	0.96 ± 0.04	0.98 ± 0.05	0.95 ± 0.07	0.98 ± 0.06
	✗	SVM_lin	1	0.96 ± 0.03	0.96 ± 0.04	0.96 ± 0.04	1.00 ± 0.00	0.92 ± 0.07	1.00 ± 0.00
		SVM_RBF	{10, 0.0001}	0.96 ± 0.03	0.96 ± 0.04	0.96 ± 0.04	1.00 ± 0.00	0.92 ± 0.07	1.00 ± 0.00
		SVM_sigmoid	{1, 0.0001}	0.96 ± 0.03	0.96 ± 0.04	0.96 ± 0.04	1.00 ± 0.00	0.92 ± 0.07	1.00 ± 0.00
		XGBoost	{2, 1}	0.92 ± 0.07	0.92 ± 0.07	0.92 ± 0.07	0.95 ± 0.07	0.90 ± 0.10	0.95 ± 0.07
		kNN	21	0.86 ± 0.06	0.86 ± 0.06	0.86 ± 0.06	0.91 ± 0.09	0.82 ± 0.11	0.89 ± 0.12
sEMG$_{P1}$ + Acc	✓	SVM_lin	1	0.90 ± 0.10	0.90 ± 0.10	0.90 ± 0.10	0.93 ± 0.11	0.88 ± 0.12	0.92 ± 0.13
		SVM_RBF	{1, 0.0001}	0.90 ± 0.09	0.88 ± 0.12	0.90 ± 0.10	1.00 ± 0.00	0.80 ± 0.19	1.00 ± 0.00
		SVM_sigmoid	{1, 0.0001}	0.88 ± 0.06	0.85 ± 0.08	0.87 ± 0.07	1.00 ± 0.00	0.75 ± 0.12	1.00 ± 0.00
		XGBoost	{2, 1}	0.86 ± 0.12	0.84 ± 0.15	0.86 ± 0.12	0.94 ± 0.09	0.78 ± 0.22	0.95 ± 0.07
		kNN	18	0.90 ± 0.09	0.88 ± 0.12	0.90 ± 0.10	1.00 ± 0.00	0.80 ± 0.19	1.00 ± 0.00
	✗	SVM_lin	1	0.88 ± 0.10	0.86 ± 0.11	0.87 ± 0.10	0.92 ± 0.08	0.82 ± 0.17	0.92 ± 0.07
		SVM_RBF	{1, 0.0001}	0.86 ± 0.11	0.85 ± 0.12	0.86 ± 0.11	0.89 ± 0.07	0.82 ± 0.17	0.90 ± 0.06
		SVM_sigmoid	{1, 0.0001}	0.88 ± 0.10	0.86 ± 0.11	0.87 ± 0.10	0.92 ± 0.08	0.82 ± 0.17	0.92 ± 0.07
		XGBoost	{2, 1}	0.75 ± 0.15	0.70 ± 0.20	0.74 ± 0.15	0.80 ± 0.12	0.65 ± 0.27	0.85 ± 0.10
		kNN	22	0.82 ± 0.07	0.81 ± 0.09	0.82 ± 0.07	0.86 ± 0.02	0.78 ± 0.14	0.87 ± 0.01
sEMG$_{P2}$ + Acc	✓	SVM_lin	1	0.90 ± 0.06	0.89 ± 0.07	0.90 ± 0.06	0.93 ± 0.06	0.88 ± 0.15	0.92 ± 0.07
		SVM_RBF	{1, 0.001}	0.90 ± 0.08	0.89 ± 0.10	0.90 ± 0.08	0.95 ± 0.07	0.85 ± 0.16	0.95 ± 0.07
		SVM_sigmoid	{1, 0.001}	0.90 ± 0.06	0.89 ± 0.06	0.90 ± 0.06	0.95 ± 0.07	0.85 ± 0.10	0.95 ± 0.07
		XGBoost	{2, 1}	0.90 ± 0.11	0.88 ± 0.14	0.90 ± 0.11	0.97 ± 0.06	0.82 ± 0.21	0.98 ± 0.06
		kNN	12	0.89 ± 0.08	0.88 ± 0.08	0.89 ± 0.08	0.90 ± 0.10	0.88 ± 0.12	0.90 ± 0.10
	✗	SVM_lin	1	0.86 ± 0.07	0.85 ± 0.08	0.86 ± 0.07	0.92 ± 0.07	0.80 ± 0.14	0.92 ± 0.07
		SVM_RBF	{10, 0.0001}	0.85 ± 0.12	0.83 ± 0.13	0.85 ± 0.12	0.89 ± 0.11	0.80 ± 0.19	0.89 ± 0.12
		SVM_sigmoid	{1, 0.0001}	0.87 ± 0.09	0.86 ± 0.10	0.87 ± 0.09	0.92 ± 0.07	0.82 ± 0.17	0.92 ± 0.07
		XGBoost	{2, 1}	0.86 ± 0.17	0.85 ± 0.17	0.86 ± 0.17	0.89 ± 0.17	0.82 ± 0.19	0.90 ± 0.16
		kNN	20	0.84 ± 0.08	0.81 ± 0.12	0.83 ± 0.09	0.92 ± 0.08	0.75 ± 0.20	0.92 ± 0.07
sEMG$_{P1}$ + Speech	✓	SVM_lin	1	0.95 ± 0.05	0.95 ± 0.06	0.95 ± 0.05	0.98 ± 0.05	0.92 ± 0.11	0.98 ± 0.06
		SVM_RBF	{1, 0.001}	0.92 ± 0.08	0.93 ± 0.06	0.92 ± 0.07	0.90 ± 0.11	0.98 ± 0.06	0.86 ± 0.18
		SVM_sigmoid	{1, 0.001}	0.98 ± 0.03	0.97 ± 0.03	0.98 ± 0.03	0.98 ± 0.05	0.98 ± 0.06	0.98 ± 0.06
		XGBoost	{2, 1}	0.96 ± 0.06	0.96 ± 0.06	0.96 ± 0.06	0.98 ± 0.05	0.95 ± 0.11	0.98 ± 0.06
		kNN	15	0.99 ± 0.03	0.99 ± 0.03	0.99 ± 0.03	1.00 ± 0.00	0.98 ± 0.06	1.00 ± 0.00
	✗	SVM_lin	1	0.92 ± 0.07	0.92 ± 0.07	0.92 ± 0.07	0.95 ± 0.07	0.90 ± 0.10	0.95 ± 0.07
		SVM_RBF	{10, 0.0001}	0.86 ± 0.08	0.87 ± 0.08	0.86 ± 0.08	0.84 ± 0.10	0.90 ± 0.10	0.82 ± 0.12
		SVM_sigmoid	{1, 0.0001}	0.91 ± 0.10	0.90 ± 0.12	0.91 ± 0.10	0.97 ± 0.07	0.85 ± 0.16	0.98 ± 0.06
		XGBoost	{2, 1}	0.95 ± 0.05	0.95 ± 0.05	0.95 ± 0.05	0.95 ± 0.06	0.95 ± 0.07	0.95 ± 0.07
		kNN	17	0.85 ± 0.05	0.85 ± 0.05	0.85 ± 0.06	0.87 ± 0.13	0.85 ± 0.10	0.85 ± 0.16
sEMG$_{P2}$ + Speech	✓	SVM_lin	1	0.96 ± 0.06	0.96 ± 0.06	0.96 ± 0.06	0.98 ± 0.05	0.95 ± 0.11	0.98 ± 0.06
		SVM_RBF	{1, 0.001}	0.95 ± 0.05	0.95 ± 0.06	0.95 ± 0.05	0.98 ± 0.05	0.92 ± 0.11	0.98 ± 0.06
		SVM_sigmoid	{1, 0.001}	0.96 ± 0.06	0.96 ± 0.06	0.96 ± 0.06	0.98 ± 0.05	0.95 ± 0.11	0.98 ± 0.06
		XGBoost	{2, 1}	0.95 ± 0.05	0.95 ± 0.05	0.95 ± 0.05	0.96 ± 0.09	0.95 ± 0.07	0.95 ± 0.11
		kNN	14	0.96 ± 0.03	0.96 ± 0.04	0.96 ± 0.04	0.98 ± 0.05	0.95 ± 0.07	0.98 ± 0.06
	✗	SVM_lin	1	0.95 ± 0.05	0.94 ± 0.06	0.95 ± 0.05	1.00 ± 0.00	0.90 ± 0.10	1.00 ± 0.00
		SVM_RBF	{10, 0.0001}	0.94 ± 0.04	0.93 ± 0.05	0.94 ± 0.04	0.98 ± 0.05	0.90 ± 0.10	0.98 ± 0.06
		SVM_sigmoid	{10, 0.0001}	0.95 ± 0.05	0.94 ± 0.06	0.95 ± 0.05	1.00 ± 0.00	0.90 ± 0.10	1.00 ± 0.00
		XGBoost	{2, 1}	0.95 ± 0.05	0.95 ± 0.05	0.95 ± 0.05	0.96 ± 0.09	0.95 ± 0.07	0.95 ± 0.11
		kNN	5	0.90 ± 0.12	0.90 ± 0.11	0.90 ± 0.12	0.88 ± 0.12	0.92 ± 0.11	0.87 ± 0.13
Acc + Speech	✓	SVM_lin	1	0.96 ± 0.06	0.96 ± 0.06	0.96 ± 0.06	1.00 ± 0.00	0.92 ± 0.11	1.00 ± 0.00
		SVM_RBF	{1, 0.001}	0.95 ± 0.05	0.95 ± 0.05	0.95 ± 0.05	0.95 ± 0.06	0.95 ± 0.07	0.95 ± 0.07
		SVM_sigmoid	{1, 0.001}	0.98 ± 0.03	0.97 ± 0.03	0.98 ± 0.03	0.98 ± 0.05	0.98 ± 0.06	0.98 ± 0.06
		XGBoost	{2, 1}	0.96 ± 0.03	0.96 ± 0.04	0.96 ± 0.04	0.98 ± 0.05	0.95 ± 0.07	0.98 ± 0.06
		kNN	15	0.98 ± 0.03	0.97 ± 0.03	0.98 ± 0.03	0.98 ± 0.05	0.98 ± 0.06	0.98 ± 0.06
	✗	SVM_lin	1	0.94 ± 0.04	0.93 ± 0.05	0.94 ± 0.04	0.98 ± 0.05	0.90 ± 0.10	0.98 ± 0.06
		SVM_RBF	{10, 0.0001}	0.95 ± 0.03	0.95 ± 0.03	0.95 ± 0.03	0.98 ± 0.05	0.92 ± 0.07	0.98 ± 0.06
		SVM_sigmoid	{10, 0.0001}	0.94 ± 0.04	0.93 ± 0.05	0.94 ± 0.04	0.98 ± 0.05	0.90 ± 0.10	0.98 ± 0.06
		XGBoost	{2, 1}	0.95 ± 0.05	0.95 ± 0.05	0.95 ± 0.05	0.95 ± 0.06	0.95 ± 0.07	0.95 ± 0.07
		kNN	9	0.85 ± 0.07	0.82 ± 0.10	0.85 ± 0.07	0.95 ± 0.07	0.75 ± 0.18	0.95 ± 0.07

The hyperparameters are provided in the following order: SVM: {C, γ} (for linear kernel, γ doesn't apply); XGBoost: {depth, negative/positive balance}; kNN: {neighbors}.

Table E.2: Classification results for multimodal scenarios using classical machine learning algorithms and 15% of the database for test.

Scenario	Feature selection	Classifier	Hyperparameters mode	AUC	F1	Accuracy	Precision	Sensitivity	Specificity
sEMG$_{P1}$ + Acc + Speech	✓	SVM_lin	1	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.02	0.93 ± 0.04	0.95 ± 0.05	0.93 ± 0.04
		SVM_RBF	{1, 0.001}	0.96 ± 0.00	0.96 ± 0.00	0.96 ± 0.00	0.97 ± 0.04	0.95 ± 0.05	0.96 ± 0.05
		SVM_sigmoid	{1, 0.0001}	0.93 ± 0.04	0.93 ± 0.04	0.93 ± 0.04	0.97 ± 0.05	0.89 ± 0.08	0.96 ± 0.05
		XGBoost	{2, 1}	0.92 ± 0.04	0.92 ± 0.04	0.92 ± 0.04	0.91 ± 0.06	0.94 ± 0.09	0.90 ± 0.07
		kNN	25	0.94 ± 0.03	0.94 ± 0.03	0.94 ± 0.03	0.97 ± 0.05	0.91 ± 0.07	0.96 ± 0.05
	✗	SVM_lin	1	0.87 ± 0.04	0.86 ± 0.05	0.87 ± 0.05	0.89 ± 0.04	0.84 ± 0.07	0.90 ± 0.04
		SVM_RBF	{10, 0.0001}	0.83 ± 0.01	0.83 ± 0.01	0.83 ± 0.01	0.81 ± 0.03	0.84 ± 0.04	0.81 ± 0.05
		SVM_sigmoid	{1, 0.0001}	0.87 ± 0.04	0.86 ± 0.04	0.87 ± 0.04	0.92 ± 0.08	0.81 ± 0.03	0.93 ± 0.07
		XGBoost	{2, 1}	0.90 ± 0.04	0.91 ± 0.04	0.90 ± 0.05	0.89 ± 0.08	0.93 ± 0.07	0.88 ± 0.10
		kNN	30	0.76 ± 0.10	0.77 ± 0.04	0.77 ± 0.09	0.81 ± 0.16	0.79 ± 0.16	0.73 ± 0.32
sEMG$_{P2}$ + Acc + Speech	✓	SVM_lin	1	0.90 ± 0.04	0.90 ± 0.04	0.90 ± 0.04	0.90 ± 0.03	0.91 ± 0.07	0.90 ± 0.04
		SVM_RBF	{1, 0.001}	0.89 ± 0.05	0.90 ± 0.04	0.90 ± 0.05	0.87 ± 0.09	0.95 ± 0.05	0.84 ± 0.11
		SVM_sigmoid	{1, 0.001}	0.91 ± 0.03	0.92 ± 0.03	0.91 ± 0.03	0.89 ± 0.07	0.95 ± 0.05	0.88 ± 0.08
		XGBoost	{2, 1}	0.90 ± 0.05	0.91 ± 0.05	0.91 ± 0.05	0.89 ± 0.08	0.93 ± 0.06	0.88 ± 0.10
		kNN	10	0.94 ± 0.04	0.94 ± 0.03	0.94 ± 0.04	0.94 ± 0.09	0.95 ± 0.05	0.93 ± 0.12
	✗	SVM_lin	1	0.92 ± 0.04	0.92 ± 0.03	0.92 ± 0.04	0.92 ± 0.08	0.93 ± 0.04	0.91 ± 0.09
		SVM_RBF	{10, 0.0001}	0.91 ± 0.03	0.91 ± 0.03	0.91 ± 0.03	0.90 ± 0.06	0.93 ± 0.04	0.89 ± 0.08
		SVM_sigmoid	{1, 0.0001}	0.91 ± 0.03	0.91 ± 0.03	0.91 ± 0.03	0.90 ± 0.06	0.93 ± 0.04	0.89 ± 0.08
		XGBoost	{2, 1}	0.93 ± 0.05	0.93 ± 0.05	0.93 ± 0.05	0.91 ± 0.06	0.97 ± 0.05	0.90 ± 0.07
		kNN	20	0.84 ± 0.09	0.82 ± 0.11	0.84 ± 0.09	0.93 ± 0.11	0.76 ± 0.16	0.93 ± 0.12
sEMG$_{P1}$ + Acc	✓	SVM_lin	1	0.87 ± 0.04	0.87 ± 0.04	0.87 ± 0.04	0.88 ± 0.04	0.86 ± 0.07	0.88 ± 0.05
		SVM_RBF	{1, 0.0001}	0.84 ± 0.07	0.84 ± 0.05	0.84 ± 0.07	0.90 ± 0.14	0.81 ± 0.07	0.87 ± 0.19
		SVM_sigmoid	{1, 0.0001}	0.87 ± 0.05	0.86 ± 0.06	0.87 ± 0.05	0.94 ± 0.06	0.79 ± 0.09	0.95 ± 0.05
		XGBoost	{2, 1}	0.81 ± 0.05	0.82 ± 0.03	0.81 ± 0.05	0.80 ± 0.08	0.84 ± 0.07	0.78 ± 0.13
		kNN	27	0.86 ± 0.02	0.85 ± 0.03	0.86 ± 0.02	0.91 ± 0.06	0.81 ± 0.08	0.91 ± 0.06
	✗	SVM_lin	1	0.78 ± 0.06	0.78 ± 0.06	0.78 ± 0.06	0.78 ± 0.09	0.79 ± 0.04	0.78 ± 0.10
		SVM_RBF	{1, 0.0001}	0.79 ± 0.05	0.80 ± 0.05	0.79 ± 0.05	0.77 ± 0.05	0.83 ± 0.06	0.76 ± 0.07
		SVM_sigmoid	{1, 0.0001}	0.82 ± 0.04	0.81 ± 0.04	0.82 ± 0.04	0.82 ± 0.06	0.81 ± 0.07	0.83 ± 0.06
		XGBoost	{2, 1}	0.75 ± 0.05	0.76 ± 0.05	0.75 ± 0.05	0.73 ± 0.06	0.79 ± 0.06	0.71 ± 0.09
		kNN	9	0.74 ± 0.11	0.77 ± 0.06	0.75 ± 0.11	0.77 ± 0.17	0.80 ± 0.12	0.68 ± 0.32
sEMG$_{P2}$ + Acc	✓	SVM_lin	1	0.83 ± 0.09	0.84 ± 0.08	0.84 ± 0.09	0.84 ± 0.10	0.84 ± 0.11	0.83 ± 0.13
		SVM_RBF	{1, 0.001}	0.88 ± 0.06	0.88 ± 0.05	0.88 ± 0.06	0.88 ± 0.10	0.90 ± 0.09	0.86 ± 0.14
		SVM_sigmoid	{1, 0.001}	0.85 ± 0.08	0.86 ± 0.07	0.85 ± 0.08	0.84 ± 0.11	0.88 ± 0.09	0.82 ± 0.14
		XGBoost	{2, 1}	0.90 ± 0.08	0.90 ± 0.07	0.90 ± 0.08	0.90 ± 0.12	0.92 ± 0.06	0.88 ± 0.15
		kNN	14	0.89 ± 0.09	0.90 ± 0.08	0.89 ± 0.09	0.85 ± 0.11	0.95 ± 0.07	0.82 ± 0.14
	✗	SVM_lin	1	0.82 ± 0.08	0.82 ± 0.07	0.82 ± 0.08	0.83 ± 0.12	0.81 ± 0.07	0.82 ± 0.15
		SVM_RBF	{10, 0.0001}	0.83 ± 0.09	0.84 ± 0.08	0.83 ± 0.09	0.84 ± 0.12	0.85 ± 0.06	0.82 ± 0.15
		SVM_sigmoid	{1, 0.0001}	0.80 ± 0.06	0.81 ± 0.05	0.80 ± 0.06	0.79 ± 0.07	0.83 ± 0.06	0.77 ± 0.11
		XGBoost	{2, 1}	0.86 ± 0.10	0.87 ± 0.09	0.86 ± 0.10	0.86 ± 0.11	0.88 ± 0.09	0.84 ± 0.14
		kNN	20	0.79 ± 0.08	0.78 ± 0.07	0.79 ± 0.08	0.83 ± 0.13	0.75 ± 0.12	0.82 ± 0.15
sEMG$_{P1}$ + Speech	✓	SVM_lin	1	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.06	0.95 ± 0.05	0.93 ± 0.08
		SVM_RBF	{1, 0.001}	0.91 ± 0.07	0.92 ± 0.06	0.91 ± 0.07	0.90 ± 0.12	0.96 ± 0.05	0.86 ± 0.17
		SVM_sigmoid	{1, 0.01}	0.93 ± 0.04	0.93 ± 0.04	0.93 ± 0.04	0.93 ± 0.07	0.93 ± 0.04	0.93 ± 0.07
		XGBoost	{2, 1}	0.92 ± 0.03	0.92 ± 0.03	0.92 ± 0.03	0.91 ± 0.06	0.95 ± 0.05	0.90 ± 0.07
		kNN	5	0.96 ± 0.02	0.96 ± 0.02	0.97 ± 0.02	0.98 ± 0.03	0.95 ± 0.05	0.98 ± 0.04
	✗	SVM_lin	1	0.85 ± 0.05	0.85 ± 0.06	0.85 ± 0.05	0.83 ± 0.07	0.88 ± 0.05	0.83 ± 0.06
		SVM_RBF	{10, 0.0001}	0.82 ± 0.05	0.82 ± 0.04	0.82 ± 0.05	0.79 ± 0.05	0.86 ± 0.05	0.77 ± 0.09
		SVM_sigmoid	{1, 0.0001}	0.88 ± 0.02	0.87 ± 0.02	0.88 ± 0.02	0.93 ± 0.07	0.82 ± 0.06	0.93 ± 0.07
		XGBoost	{2, 1}	0.89 ± 0.05	0.89 ± 0.06	0.89 ± 0.05	0.89 ± 0.08	0.89 ± 0.11	0.88 ± 0.10
		kNN	9	0.77 ± 0.12	0.73 ± 0.23	0.78 ± 0.10	0.86 ± 0.15	0.74 ± 0.31	0.81 ± 0.21
sEMG$_{P2}$ + Speech	✓	SVM_lin	1	0.91 ± 0.04	0.91 ± 0.04	0.91 ± 0.03	0.93 ± 0.04	0.89 ± 0.05	0.93 ± 0.04
		SVM_RBF	{1, 0.001}	0.93 ± 0.02	0.93 ± 0.03	0.93 ± 0.03	0.92 ± 0.06	0.95 ± 0.05	0.91 ± 0.06
		SVM_sigmoid	{1, 0.001}	0.92 ± 0.02	0.92 ± 0.02	0.92 ± 0.02	0.90 ± 0.06	0.95 ± 0.05	0.89 ± 0.08
		XGBoost	{2, 1}	0.90 ± 0.05	0.90 ± 0.05	0.90 ± 0.06	0.89 ± 0.08	0.91 ± 0.06	0.88 ± 0.10
		kNN	4	0.92 ± 0.08	0.92 ± 0.07	0.92 ± 0.08	0.93 ± 0.12	0.93 ± 0.04	0.91 ± 0.16
	✗	SVM_lin	1	0.90 ± 0.06	0.91 ± 0.05	0.90 ± 0.06	0.90 ± 0.10	0.91 ± 0.01	0.89 ± 0.12
		SVM_RBF	{10, 0.0001}	0.88 ± 0.07	0.89 ± 0.05	0.89 ± 0.07	0.88 ± 0.10	0.91 ± 0.01	0.86 ± 0.14
		SVM_sigmoid	{10, 0.0001}	0.90 ± 0.06	0.91 ± 0.05	0.90 ± 0.06	0.90 ± 0.10	0.91 ± 0.01	0.89 ± 0.12
		XGBoost	{2, 1}	0.90 ± 0.05	0.90 ± 0.04	0.90 ± 0.05	0.87 ± 0.06	0.93 ± 0.04	0.86 ± 0.08
		kNN	10	0.81 ± 0.06	0.80 ± 0.08	0.81 ± 0.06	0.85 ± 0.15	0.81 ± 0.20	0.81 ± 0.20
Acc + Speech	✓	SVM_lin	1	0.90 ± 0.04	0.90 ± 0.03	0.90 ± 0.03	0.92 ± 0.05	0.89 ± 0.08	0.91 ± 0.06
		SVM_RBF	{1, 0.001}	0.93 ± 0.04	0.93 ± 0.03	0.93 ± 0.04	0.95 ± 0.07	0.91 ± 0.06	0.95 ± 0.08
		SVM_sigmoid	{1, 0.0001}	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.02	0.97 ± 0.07	0.91 ± 0.06	0.96 ± 0.08
		XGBoost	{2, 1}	0.90 ± 0.05	0.91 ± 0.05	0.91 ± 0.05	0.89 ± 0.08	0.93 ± 0.07	0.88 ± 0.10
		kNN	13	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.02	0.97 ± 0.05	0.91 ± 0.06	0.97 ± 0.05
	✗	SVM_lin	1	0.91 ± 0.04	0.91 ± 0.04	0.91 ± 0.04	0.91 ± 0.06	0.91 ± 0.06	0.92 ± 0.06
		SVM_RBF	{10, 0.0001}	0.92 ± 0.03	0.92 ± 0.03	0.92 ± 0.03	0.92 ± 0.06	0.93 ± 0.04	0.92 ± 0.06
		SVM_sigmoid	{10, 0.0001}	0.91 ± 0.04	0.91 ± 0.04	0.91 ± 0.04	0.91 ± 0.06	0.91 ± 0.06	0.92 ± 0.06
		XGBoost	{2, 1}	0.92 ± 0.05	0.93 ± 0.05	0.92 ± 0.05	0.90 ± 0.06	0.95 ± 0.07	0.90 ± 0.07
		kNN	23	0.81 ± 0.09	0.79 ± 0.11	0.81 ± 0.09	0.87 ± 0.13	0.76 ± 0.18	0.86 ± 0.14

The hyperparameters are provided in the following order: SVM: {C, γ} (for linear kernel, γ doesn't apply); XGBoost: {depth, negative/positive balance}; kNN: {neighbors}.

Table E.3: Classification results for multimodal scenarios using classical machine learning algorithms and 20% of the database for test.

Scenario	Feature selection	Classifier	Hyperparameters mode	AUC	F1	Accuracy	Precision	Sensitivity	Specificity
sEMG$_{P1}$ + Acc + Speech	✓	SVM_lin	1	0.94 ± 0.03	0.94 ± 0.04	0.94 ± 0.03	1.00 ± 0.00	0.89 ± 0.06	1.00 ± 0.00
		SVM_RBF	{1, 0.001}	0.94 ± 0.03	0.93 ± 0.03	0.94 ± 0.03	0.98 ± 0.03	0.89 ± 0.06	0.99 ± 0.03
		SVM_sigmoid	{1, 0.001}	0.94 ± 0.03	0.94 ± 0.04	0.95 ± 0.03	1.00 ± 0.00	0.89 ± 0.06	1.00 ± 0.00
		XGBoost	{2, 1}	0.91 ± 0.04	0.91 ± 0.04	0.91 ± 0.04	0.93 ± 0.07	0.89 ± 0.04	0.93 ± 0.06
		kNN	14	0.95 ± 0.03	0.95 ± 0.04	0.95 ± 0.03	1.00 ± 0.00	0.90 ± 0.06	1.00 ± 0.00
	✗	SVM_lin	1	0.89 ± 0.04	0.89 ± 0.04	0.90 ± 0.04	0.95 ± 0.04	0.83 ± 0.04	0.96 ± 0.04
		SVM_RBF	{1, 0.0001}	0.86 ± 0.07	0.85 ± 0.06	0.86 ± 0.06	0.88 ± 0.10	0.83 ± 0.04	0.89 ± 0.10
		SVM_sigmoid	{1, 0.0001}	0.89 ± 0.03	0.89 ± 0.04	0.90 ± 0.04	0.97 ± 0.05	0.82 ± 0.03	0.97 ± 0.04
		XGBoost	{2, 1}	0.92 ± 0.06	0.92 ± 0.06	0.92 ± 0.06	0.95 ± 0.07	0.89 ± 0.06	0.96 ± 0.06
		kNN	14	0.84 ± 0.08	0.83 ± 0.08	0.84 ± 0.08	0.90 ± 0.12	0.77 ± 0.06	0.90 ± 0.11
sEMG$_{P2}$ + Acc + Speech	✓	SVM_lin	1	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.02	0.97 ± 0.04	0.91 ± 0.03	0.97 ± 0.04
		SVM_RBF	{1, 0.001}	0.93 ± 0.04	0.93 ± 0.04	0.93 ± 0.03	0.96 ± 0.06	0.90 ± 0.06	0.96 ± 0.06
		SVM_sigmoid	{1, 0.001}	0.94 ± 0.03	0.93 ± 0.03	0.94 ± 0.03	0.96 ± 0.06	0.91 ± 0.03	0.96 ± 0.06
		XGBoost	{2, 1}	0.93 ± 0.04	0.93 ± 0.04	0.93 ± 0.04	0.93 ± 0.08	0.93 ± 0.01	0.93 ± 0.08
		kNN	8	0.94 ± 0.04	0.93 ± 0.04	0.94 ± 0.04	0.97 ± 0.06	0.90 ± 0.06	0.98 ± 0.06
	✗	SVM_lin	1	0.92 ± 0.02	0.91 ± 0.02	0.92 ± 0.02	0.95 ± 0.04	0.87 ± 0.06	0.96 ± 0.04
		SVM_RBF	{10, 0.0001}	0.90 ± 0.03	0.89 ± 0.04	0.90 ± 0.03	0.91 ± 0.07	0.87 ± 0.06	0.92 ± 0.07
		SVM_sigmoid	{1, 0.0001}	0.92 ± 0.02	0.91 ± 0.02	0.92 ± 0.02	0.95 ± 0.04	0.87 ± 0.06	0.96 ± 0.04
		XGBoost	{2, 1}	0.93 ± 0.04	0.93 ± 0.04	0.93 ± 0.03	0.94 ± 0.06	0.91 ± 0.06	0.95 ± 0.05
		kNN	6	0.87 ± 0.04	0.86 ± 0.06	0.88 ± 0.06	0.93 ± 0.07	0.80 ± 0.08	0.95 ± 0.06
sEMG$_{P1}$ + Acc	✓	SVM_lin	1	0.90 ± 0.05	0.89 ± 0.06	0.90 ± 0.05	0.95 ± 0.07	0.84 ± 0.07	0.96 ± 0.06
		SVM_RBF	{1, 0.0001}	0.87 ± 0.02	0.86 ± 0.03	0.88 ± 0.03	1.00 ± 0.00	0.75 ± 0.04	1.00 ± 0.00
		SVM_sigmoid	{1, 0.0001}	0.85 ± 0.04	0.82 ± 0.06	0.85 ± 0.05	1.00 ± 0.00	0.69 ± 0.08	1.00 ± 0.00
		XGBoost	{2, 1}	0.85 ± 0.06	0.84 ± 0.06	0.86 ± 0.06	0.90 ± 0.11	0.80 ± 0.07	0.91 ± 0.11
		kNN	10	0.87 ± 0.02	0.86 ± 0.03	0.88 ± 0.02	0.98 ± 0.04	0.76 ± 0.04	0.99 ± 0.03
	✗	SVM_lin	1	0.83 ± 0.07	0.81 ± 0.07	0.83 ± 0.07	0.88 ± 0.09	0.75 ± 0.08	0.91 ± 0.08
		SVM_RBF	{10, 0.0001}	0.83 ± 0.06	0.81 ± 0.06	0.83 ± 0.06	0.88 ± 0.10	0.76 ± 0.07	0.89 ± 0.10
		SVM_sigmoid	{1, 0.0001}	0.83 ± 0.07	0.82 ± 0.07	0.83 ± 0.07	0.89 ± 0.09	0.76 ± 0.06	0.90 ± 0.08
		XGBoost	{2, 1}	0.85 ± 0.06	0.84 ± 0.05	0.85 ± 0.06	0.90 ± 0.14	0.80 ± 0.08	0.89 ± 0.17
		kNN	6	0.73 ± 0.08	0.78 ± 0.03	0.74 ± 0.08	0.70 ± 0.13	0.91 ± 0.14	0.56 ± 0.28
sEMG$_{P2}$ + Acc	✓	SVM_lin	1	0.85 ± 0.06	0.84 ± 0.08	0.86 ± 0.06	0.87 ± 0.06	0.83 ± 0.11	0.88 ± 0.05
		SVM_RBF	{10, 0.001}	0.86 ± 0.06	0.85 ± 0.06	0.86 ± 0.06	0.90 ± 0.10	0.82 ± 0.06	0.91 ± 0.09
		SVM_sigmoid	{1, 0.001}	0.88 ± 0.04	0.86 ± 0.05	0.88 ± 0.04	0.92 ± 0.07	0.82 ± 0.06	0.93 ± 0.07
		XGBoost	{2, 1}	0.90 ± 0.07	0.89 ± 0.07	0.90 ± 0.06	0.93 ± 0.11	0.86 ± 0.08	0.93 ± 0.09
		kNN	4	0.88 ± 0.04	0.87 ± 0.05	0.88 ± 0.04	0.91 ± 0.07	0.83 ± 0.06	0.92 ± 0.05
	✗	SVM_lin	1	0.84 ± 0.04	0.82 ± 0.06	0.84 ± 0.04	0.89 ± 0.04	0.77 ± 0.08	0.91 ± 0.04
		SVM_RBF	{10, 0.0001}	0.83 ± 0.06	0.82 ± 0.06	0.83 ± 0.06	0.84 ± 0.09	0.80 ± 0.06	0.85 ± 0.08
		SVM_sigmoid	{1, 0.0001}	0.83 ± 0.06	0.81 ± 0.07	0.83 ± 0.06	0.85 ± 0.06	0.79 ± 0.09	0.86 ± 0.05
		XGBoost	{2, 1}	0.87 ± 0.07	0.86 ± 0.08	0.87 ± 0.07	0.89 ± 0.09	0.83 ± 0.09	0.91 ± 0.07
		kNN	18	0.79 ± 0.05	0.79 ± 0.05	0.79 ± 0.05	0.79 ± 0.05	0.79 ± 0.07	0.79 ± 0.09
sEMG$_{P1}$ + Speech	✓	SVM_lin	1	0.94 ± 0.03	0.93 ± 0.03	0.94 ± 0.03	0.97 ± 0.06	0.90 ± 0.04	0.98 ± 0.06
		SVM_RBF	{1, 0.0001}	0.96 ± 0.02	0.96 ± 0.02	0.96 ± 0.02	0.97 ± 0.04	0.94 ± 0.03	0.97 ± 0.04
		SVM_sigmoid	{1, 0.0001}	0.95 ± 0.03	0.95 ± 0.03	0.95 ± 0.03	1.00 ± 0.00	0.90 ± 0.06	1.00 ± 0.00
		XGBoost	{2, 1}	0.94 ± 0.05	0.93 ± 0.05	0.94 ± 0.04	0.96 ± 0.06	0.91 ± 0.03	0.96 ± 0.06
		kNN	3	0.94 ± 0.02	0.93 ± 0.02	0.94 ± 0.01	0.98 ± 0.03	0.89 ± 0.04	0.99 ± 0.03
	✗	SVM_lin	1	0.87 ± 0.05	0.86 ± 0.05	0.87 ± 0.04	0.91 ± 0.09	0.82 ± 0.06	0.92 ± 0.09
		SVM_RBF	{10, 0.0001}	0.85 ± 0.06	0.85 ± 0.06	0.85 ± 0.06	0.86 ± 0.12	0.85 ± 0.06	0.86 ± 0.14
		SVM_sigmoid	{1, 0.0001}	0.90 ± 0.04	0.89 ± 0.04	0.90 ± 0.04	0.95 ± 0.07	0.83 ± 0.06	0.96 ± 0.06
		XGBoost	{2, 1}	0.94 ± 0.04	0.94 ± 0.04	0.94 ± 0.04	0.95 ± 0.08	0.93 ± 0.01	0.95 ± 0.09
		kNN	7	0.85 ± 0.14	0.80 ± 0.22	0.85 ± 0.13	0.90 ± 0.07	0.76 ± 0.28	0.93 ± 0.04
sEMG$_{P2}$ + Speech	✓	SVM_lin	1	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.02	0.97 ± 0.04	0.91 ± 0.03	0.97 ± 0.04
		SVM_RBF	{1, 0.001}	0.94 ± 0.03	0.93 ± 0.03	0.94 ± 0.03	0.96 ± 0.06	0.91 ± 0.03	0.96 ± 0.06
		SVM_sigmoid	{1, 0.001}	0.94 ± 0.03	0.93 ± 0.03	0.94 ± 0.03	0.96 ± 0.06	0.91 ± 0.03	0.96 ± 0.06
		XGBoost	{2, 1}	0.92 ± 0.06	0.92 ± 0.06	0.92 ± 0.06	0.91 ± 0.09	0.94 ± 0.03	0.91 ± 0.09
		kNN	6	0.94 ± 0.03	0.93 ± 0.03	0.94 ± 0.03	0.97 ± 0.06	0.90 ± 0.03	0.98 ± 0.06
	✗	SVM_lin	1	0.92 ± 0.02	0.91 ± 0.02	0.92 ± 0.02	0.95 ± 0.04	0.87 ± 0.06	0.96 ± 0.04
		SVM_RBF	{10, 0.0001}	0.92 ± 0.02	0.91 ± 0.02	0.92 ± 0.02	0.95 ± 0.04	0.87 ± 0.06	0.96 ± 0.04
		SVM_sigmoid	{10, 0.0001}	0.92 ± 0.02	0.91 ± 0.02	0.92 ± 0.02	0.95 ± 0.04	0.87 ± 0.06	0.96 ± 0.04
		XGBoost	{2, 1}	0.94 ± 0.04	0.94 ± 0.05	0.94 ± 0.04	0.98 ± 0.03	0.90 ± 0.08	0.99 ± 0.03
		kNN	5	0.86 ± 0.06	0.86 ± 0.05	0.86 ± 0.06	0.89 ± 0.12	0.84 ± 0.11	0.87 ± 0.16
Acc + Speech	✓	SVM_lin	1	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.02	0.99 ± 0.03	0.90 ± 0.04	0.99 ± 0.03
		SVM_RBF	{1, 0.001}	0.95 ± 0.02	0.95 ± 0.02	0.95 ± 0.02	0.98 ± 0.03	0.91 ± 0.03	0.99 ± 0.03
		SVM_sigmoid	{1, 0.001}	0.95 ± 0.02	0.95 ± 0.02	0.95 ± 0.02	0.98 ± 0.03	0.91 ± 0.03	0.99 ± 0.03
		XGBoost	{2, 1}	0.93 ± 0.04	0.93 ± 0.04	0.93 ± 0.04	0.93 ± 0.05	0.93 ± 0.05	0.93 ± 0.05
		kNN	11	0.94 ± 0.04	0.94 ± 0.04	0.94 ± 0.03	1.00 ± 0.00	0.88 ± 0.07	1.00 ± 0.00
	✗	SVM_lin	1	0.94 ± 0.03	0.93 ± 0.03	0.94 ± 0.03	1.00 ± 0.00	0.87 ± 0.06	1.00 ± 0.00
		SVM_RBF	{10, 0.0001}	0.92 ± 0.02	0.92 ± 0.02	0.92 ± 0.02	0.97 ± 0.04	0.87 ± 0.06	0.97 ± 0.04
		SVM_sigmoid	{10, 0.0001}	0.92 ± 0.04	0.91 ± 0.05	0.92 ± 0.04	0.98 ± 0.04	0.85 ± 0.06	0.99 ± 0.03
		XGBoost	{5, 1}	0.90 ± 0.03	0.89 ± 0.04	0.90 ± 0.04	0.90 ± 0.08	0.89 ± 0.04	0.91 ± 0.07
		kNN	5	0.87 ± 0.05	0.86 ± 0.06	0.88 ± 0.05	0.97 ± 0.05	0.78 ± 0.07	0.97 ± 0.04

The hyperparameters are provided in the following order: SVM: {C, γ} (for linear kernel, γ doesn't apply); XGBoost: {depth, negative/positive balance}; kNN: {neighbors}.

Table E.4: Classification results for multimodal scenarios using classical machine learning algorithms and 25% of the database for test.

Scenario	Feature selection	Classifier	Hyperparameters mode	AUC	F1	Accuracy	Precision	Sensitivity	Specificity
sEMG$_{P1}$ + Acc + Speech	✓	SVM_lin	1	0.92 ± 0.02	0.91 ± 0.02	0.92 ± 0.02	0.95 ± 0.03	0.88 ± 0.04	0.96 ± 0.02
		SVM_RBF	{1, 0.001}	0.93 ± 0.03	0.93 ± 0.03	0.93 ± 0.03	0.94 ± 0.04	0.92 ± 0.03	0.95 ± 0.04
		SVM_sigmoid	{1, 0.0001}	0.93 ± 0.03	0.93 ± 0.03	0.93 ± 0.03	0.94 ± 0.00	0.92 ± 0.06	0.95 ± 0.00
		XGBoost	{2, 1}	0.88 ± 0.04	0.87 ± 0.06	0.88 ± 0.04	0.91 ± 0.05	0.85 ± 0.11	0.91 ± 0.05
		kNN	7	0.94 ± 0.04	0.94 ± 0.04	0.94 ± 0.04	0.95 ± 0.03	0.92 ± 0.06	0.96 ± 0.02
	✗	SVM_lin	1	0.88 ± 0.01	0.87 ± 0.02	0.87 ± 0.01	0.91 ± 0.05	0.84 ± 0.06	0.91 ± 0.05
		SVM_RBF	{10, 0.0001}	0.83 ± 0.02	0.82 ± 0.02	0.83 ± 0.02	0.85 ± 0.05	0.80 ± 0.09	0.86 ± 0.06
		SVM_sigmoid	{1, 0.0001}	0.89 ± 0.05	0.88 ± 0.05	0.89 ± 0.05	0.94 ± 0.04	0.83 ± 0.07	0.95 ± 0.04
		XGBoost	{2, 1}	0.90 ± 0.04	0.89 ± 0.05	0.90 ± 0.04	0.92 ± 0.02	0.87 ± 0.11	0.92 ± 0.03
		kNN	8	0.82 ± 0.03	0.80 ± 0.04	0.81 ± 0.04	0.86 ± 0.11	0.77 ± 0.09	0.86 ± 0.12
sEMG$_{P2}$ + Acc + Speech	✓	SVM_lin	1	0.95 ± 0.01	0.95 ± 0.01	0.95 ± 0.01	0.97 ± 0.05	0.94 ± 0.04	0.97 ± 0.05
		SVM_RBF	{1, 0.001}	0.93 ± 0.03	0.93 ± 0.03	0.93 ± 0.03	0.95 ± 0.07	0.92 ± 0.03	0.95 ± 0.06
		SVM_sigmoid	{1, 0.001}	0.94 ± 0.01	0.93 ± 0.02	0.93 ± 0.02	0.94 ± 0.06	0.94 ± 0.05	0.94 ± 0.06
		XGBoost	{2, 1}	0.89 ± 0.08	0.88 ± 0.08	0.89 ± 0.08	0.91 ± 0.05	0.86 ± 0.12	0.91 ± 0.05
		kNN	10	0.92 ± 0.02	0.91 ± 0.02	0.92 ± 0.02	0.94 ± 0.04	0.89 ± 0.04	0.95 ± 0.04
	✗	SVM_lin	1	0.91 ± 0.02	0.91 ± 0.03	0.91 ± 0.02	0.93 ± 0.06	0.89 ± 0.04	0.94 ± 0.06
		SVM_RBF	{10, 0.0001}	0.91 ± 0.02	0.90 ± 0.02	0.91 ± 0.02	0.92 ± 0.05	0.89 ± 0.04	0.92 ± 0.05
		SVM_sigmoid	{1, 0.0001}	0.92 ± 0.02	0.91 ± 0.02	0.92 ± 0.02	0.95 ± 0.07	0.89 ± 0.04	0.95 ± 0.06
		XGBoost	{2, 1}	0.91 ± 0.06	0.90 ± 0.08	0.91 ± 0.06	0.95 ± 0.03	0.86 ± 0.11	0.96 ± 0.02
		kNN	14	0.84 ± 0.03	0.83 ± 0.04	0.84 ± 0.03	0.90 ± 0.08	0.78 ± 0.10	0.90 ± 0.09
sEMG$_{P1}$ + Acc	✓	SVM_lin	1	0.87 ± 0.05	0.87 ± 0.05	0.87 ± 0.05	0.90 ± 0.06	0.84 ± 0.07	0.90 ± 0.06
		SVM_RBF	{1, 0.0001}	0.84 ± 0.04	0.83 ± 0.05	0.84 ± 0.05	0.89 ± 0.09	0.80 ± 0.10	0.89 ± 0.10
		SVM_sigmoid	{1, 0.0001}	0.86 ± 0.05	0.85 ± 0.05	0.86 ± 0.05	0.92 ± 0.05	0.78 ± 0.07	0.94 ± 0.05
		XGBoost	{2, 1}	0.84 ± 0.04	0.83 ± 0.04	0.84 ± 0.04	0.89 ± 0.04	0.78 ± 0.06	0.90 ± 0.04
		kNN	7	0.84 ± 0.04	0.83 ± 0.05	0.84 ± 0.04	0.89 ± 0.04	0.78 ± 0.07	0.90 ± 0.04
	✗	SVM_lin	1	0.81 ± 0.02	0.79 ± 0.04	0.80 ± 0.02	0.84 ± 0.08	0.76 ± 0.12	0.85 ± 0.10
		SVM_RBF	{1, 0.0001}	0.81 ± 0.04	0.80 ± 0.04	0.80 ± 0.04	0.83 ± 0.07	0.77 ± 0.10	0.84 ± 0.08
		SVM_sigmoid	{1, 0.0001}	0.81 ± 0.01	0.80 ± 0.02	0.81 ± 0.01	0.85 ± 0.08	0.76 ± 0.10	0.86 ± 0.10
		XGBoost	{2, 1}	0.78 ± 0.05	0.77 ± 0.05	0.78 ± 0.05	0.78 ± 0.07	0.77 ± 0.10	0.78 ± 0.10
		kNN	16	0.76 ± 0.08	0.75 ± 0.08	0.76 ± 0.08	0.79 ± 0.08	0.71 ± 0.13	0.82 ± 0.07
sEMG$_{P2}$ + Acc	✓	SVM_lin	1	0.89 ± 0.03	0.89 ± 0.03	0.89 ± 0.03	0.92 ± 0.09	0.86 ± 0.04	0.92 ± 0.09
		SVM_RBF	{10, 0.001}	0.88 ± 0.02	0.88 ± 0.02	0.88 ± 0.02	0.90 ± 0.06	0.86 ± 0.06	0.90 ± 0.07
		SVM_sigmoid	{1, 0.001}	0.90 ± 0.03	0.89 ± 0.04	0.90 ± 0.03	0.92 ± 0.03	0.87 ± 0.08	0.92 ± 0.03
		XGBoost	{2, 1}	0.91 ± 0.02	0.91 ± 0.03	0.91 ± 0.02	0.94 ± 0.06	0.88 ± 0.05	0.95 ± 0.05
		kNN	7	0.89 ± 0.05	0.88 ± 0.05	0.89 ± 0.05	0.92 ± 0.03	0.85 ± 0.09	0.92 ± 0.03
	✗	SVM_lin	1	0.85 ± 0.03	0.85 ± 0.03	0.85 ± 0.03	0.88 ± 0.08	0.83 ± 0.09	0.88 ± 0.09
		SVM_RBF	{10, 0.0001}	0.86 ± 0.02	0.85 ± 0.03	0.86 ± 0.03	0.88 ± 0.08	0.84 ± 0.09	0.88 ± 0.09
		SVM_sigmoid	{1, 0.0001}	0.86 ± 0.02	0.85 ± 0.03	0.86 ± 0.02	0.87 ± 0.08	0.85 ± 0.07	0.87 ± 0.09
		XGBoost	{2, 1}	0.85 ± 0.07	0.84 ± 0.08	0.85 ± 0.07	0.88 ± 0.05	0.81 ± 0.11	0.89 ± 0.04
		kNN	16	0.80 ± 0.02	0.78 ± 0.03	0.80 ± 0.02	0.84 ± 0.09	0.75 ± 0.11	0.85 ± 0.11
sEMG$_{P1}$ + Speech	✓	SVM_lin	1	0.95 ± 0.03	0.95 ± 0.03	0.95 ± 0.03	0.98 ± 0.03	0.93 ± 0.05	0.98 ± 0.03
		SVM_RBF	{1, 0.001}	0.93 ± 0.05	0.93 ± 0.04	0.92 ± 0.05	0.91 ± 0.09	0.96 ± 0.05	0.89 ± 0.12
		SVM_sigmoid	{1, 0.0001}	0.91 ± 0.07	0.90 ± 0.08	0.91 ± 0.07	0.97 ± 0.06	0.84 ± 0.10	0.98 ± 0.03
		XGBoost	{2, 1}	0.90 ± 0.04	0.89 ± 0.04	0.90 ± 0.04	0.91 ± 0.05	0.88 ± 0.07	0.91 ± 0.05
		kNN	5	0.94 ± 0.03	0.94 ± 0.03	0.94 ± 0.03	0.97 ± 0.03	0.91 ± 0.06	0.97 ± 0.03
	✗	SVM_lin	1	0.88 ± 0.07	0.86 ± 0.08	0.87 ± 0.07	0.92 ± 0.06	0.82 ± 0.11	0.94 ± 0.05
		SVM_RBF	{10, 0.0001}	0.85 ± 0.05	0.84 ± 0.06	0.85 ± 0.05	0.87 ± 0.03	0.82 ± 0.11	0.88 ± 0.02
		SVM_sigmoid	{1, 0.0001}	0.87 ± 0.07	0.86 ± 0.09	0.87 ± 0.07	0.96 ± 0.04	0.78 ± 0.12	0.97 ± 0.03
		XGBoost	{2, 1}	0.92 ± 0.03	0.92 ± 0.03	0.92 ± 0.03	0.92 ± 0.03	0.91 ± 0.06	0.92 ± 0.03
		kNN	6	0.77 ± 0.09	0.70 ± 0.15	0.77 ± 0.09	0.97 ± 0.04	0.57 ± 0.20	0.98 ± 0.03
sEMG$_{P2}$ + Speech	✓	SVM_lin	1	0.94 ± 0.02	0.94 ± 0.03	0.94 ± 0.02	0.96 ± 0.07	0.92 ± 0.06	0.96 ± 0.07
		SVM_RBF	{1, 0.001}	0.95 ± 0.03	0.94 ± 0.04	0.94 ± 0.04	0.97 ± 0.07	0.92 ± 0.03	0.97 ± 0.07
		SVM_sigmoid	{1, 0.001}	0.95 ± 0.03	0.94 ± 0.03	0.95 ± 0.03	0.96 ± 0.07	0.94 ± 0.05	0.96 ± 0.07
		XGBoost	{2, 1}	0.92 ± 0.06	0.91 ± 0.07	0.92 ± 0.06	0.94 ± 0.04	0.89 ± 0.11	0.95 ± 0.04
		kNN	12	0.92 ± 0.01	0.92 ± 0.02	0.92 ± 0.01	0.94 ± 0.04	0.90 ± 0.05	0.95 ± 0.04
	✗	SVM_lin	1	0.93 ± 0.03	0.93 ± 0.03	0.93 ± 0.03	0.97 ± 0.07	0.89 ± 0.03	0.97 ± 0.07
		SVM_RBF	{10, 0.0001}	0.91 ± 0.02	0.91 ± 0.02	0.91 ± 0.02	0.93 ± 0.06	0.89 ± 0.03	0.94 ± 0.06
		SVM_sigmoid	{10, 0.0001}	0.92 ± 0.05	0.91 ± 0.05	0.92 ± 0.04	0.95 ± 0.06	0.88 ± 0.05	0.96 ± 0.06
		XGBoost	{2, 1}	0.92 ± 0.03	0.92 ± 0.03	0.92 ± 0.03	0.92 ± 0.05	0.91 ± 0.03	0.92 ± 0.05
		kNN	8	0.89 ± 0.04	0.87 ± 0.04	0.88 ± 0.04	0.96 ± 0.09	0.81 ± 0.10	0.96 ± 0.09
Acc + Speech	✓	SVM_lin	1	0.94 ± 0.04	0.93 ± 0.05	0.93 ± 0.04	0.97 ± 0.05	0.90 ± 0.11	0.97 ± 0.05
		SVM_RBF	{1, 0.001}	0.96 ± 0.03	0.96 ± 0.03	0.96 ± 0.03	0.98 ± 0.03	0.94 ± 0.06	0.98 ± 0.03
		SVM_sigmoid	{1, 0.001}	0.95 ± 0.02	0.94 ± 0.02	0.95 ± 0.02	0.96 ± 0.02	0.94 ± 0.06	0.96 ± 0.02
		XGBoost	{2, 1}	0.91 ± 0.06	0.91 ± 0.06	0.91 ± 0.06	0.93 ± 0.03	0.89 ± 0.13	0.94 ± 0.05
		kNN	25	0.93 ± 0.03	0.93 ± 0.03	0.93 ± 0.03	0.94 ± 0.01	0.91 ± 0.05	0.95 ± 0.00
	✗	SVM_lin	1	0.91 ± 0.03	0.90 ± 0.03	0.91 ± 0.03	0.94 ± 0.04	0.87 ± 0.08	0.95 ± 0.04
		SVM_RBF	{10, 0.0001}	0.92 ± 0.03	0.92 ± 0.03	0.92 ± 0.03	0.96 ± 0.04	0.89 ± 0.05	0.96 ± 0.05
		SVM_sigmoid	{10, 0.0001}	0.91 ± 0.03	0.90 ± 0.03	0.91 ± 0.03	0.94 ± 0.04	0.87 ± 0.07	0.95 ± 0.04
		XGBoost	{2, 1}	0.90 ± 0.06	0.89 ± 0.07	0.90 ± 0.06	0.93 ± 0.03	0.87 ± 0.12	0.94 ± 0.02
		kNN	4	0.85 ± 0.08	0.84 ± 0.09	0.85 ± 0.08	0.94 ± 0.10	0.76 ± 0.11	0.94 ± 0.10

The hyperparameters are provided in the following order: SVM: {C, γ} (for linear kernel, γ doesn't apply); XGBoost: {depth, negative/positive balance}; kNN: {neighbors}.

Table E.5: Classification results for multimodal scenarios using classical machine learning algorithms and 30% of the database for test.

Scenario	Feature selection	Classifier	Hyperparameters mode	AUC	F1	Accuracy	Precision	Sensitivity	Specificity
sEMG$_{P1}$ + Acc + Speech	✓	SVM_lin	1	0.94 ± 0.03	0.94 ± 0.04	0.94 ± 0.03	0.97 ± 0.03	0.92 ± 0.05	0.97 ± 0.03
		SVM_RBF	{1, 0.001}	0.91 ± 0.08	0.92 ± 0.07	0.91 ± 0.09	0.88 ± 0.15	0.97 ± 0.03	0.84 ± 0.20
		SVM_sigmoid	{1, 0.0001}	0.95 ± 0.01	0.95 ± 0.02	0.95 ± 0.02	0.97 ± 0.04	0.93 ± 0.03	0.97 ± 0.04
		XGBoost	{2, 1}	0.93 ± 0.06	0.92 ± 0.07	0.92 ± 0.06	0.94 ± 0.02	0.91 ± 0.12	0.94 ± 0.02
		kNN	20	0.94 ± 0.03	0.94 ± 0.03	0.94 ± 0.03	0.98 ± 0.03	0.90 ± 0.04	0.98 ± 0.03
	✗	SVM_lin	1	0.88 ± 0.05	0.87 ± 0.05	0.88 ± 0.05	0.93 ± 0.08	0.82 ± 0.08	0.93 ± 0.07
		SVM_RBF	{10, 0.0001}	0.82 ± 0.05	0.82 ± 0.05	0.82 ± 0.04	0.81 ± 0.06	0.85 ± 0.08	0.80 ± 0.08
		SVM_sigmoid	{1, 0.0001}	0.88 ± 0.04	0.88 ± 0.05	0.89 ± 0.04	0.94 ± 0.05	0.83 ± 0.08	0.94 ± 0.05
		XGBoost	{2, 10}	0.92 ± 0.04	0.93 ± 0.03	0.92 ± 0.04	0.89 ± 0.05	0.97 ± 0.03	0.88 ± 0.07
		kNN	18	0.80 ± 0.08	0.75 ± 0.12	0.81 ± 0.08	0.98 ± 0.04	0.61 ± 0.16	0.99 ± 0.02
sEMG$_{P2}$ + Acc + Speech	✓	SVM_lin	1	0.97 ± 0.01	0.96 ± 0.02	0.97 ± 0.01	0.97 ± 0.02	0.96 ± 0.05	0.97 ± 0.02
		SVM_RBF	{1, 0.001}	0.95 ± 0.02	0.95 ± 0.01	0.95 ± 0.02	0.95 ± 0.05	0.96 ± 0.02	0.94 ± 0.05
		SVM_sigmoid	{1, 0.001}	0.94 ± 0.01	0.94 ± 0.01	0.94 ± 0.01	0.94 ± 0.05	0.95 ± 0.04	0.93 ± 0.05
		XGBoost	{2, 1}	0.92 ± 0.04	0.92 ± 0.04	0.92 ± 0.04	0.91 ± 0.04	0.93 ± 0.08	0.91 ± 0.04
		kNN	9	0.93 ± 0.02	0.93 ± 0.02	0.93 ± 0.02	0.92 ± 0.03	0.95 ± 0.01	0.91 ± 0.04
	✗	SVM_lin	1	0.94 ± 0.04	0.94 ± 0.04	0.94 ± 0.04	0.95 ± 0.06	0.93 ± 0.03	0.95 ± 0.06
		SVM_RBF	{10, 0.0001}	0.92 ± 0.05	0.92 ± 0.05	0.92 ± 0.05	0.91 ± 0.07	0.92 ± 0.04	0.91 ± 0.08
		SVM_sigmoid	{1, 0.0001}	0.93 ± 0.05	0.93 ± 0.05	0.93 ± 0.05	0.95 ± 0.06	0.91 ± 0.06	0.95 ± 0.06
		XGBoost	{2, 1}	0.92 ± 0.03	0.91 ± 0.04	0.92 ± 0.03	0.92 ± 0.02	0.91 ± 0.09	0.92 ± 0.02
		kNN	11	0.85 ± 0.07	0.83 ± 0.09	0.85 ± 0.07	0.93 ± 0.04	0.76 ± 0.13	0.94 ± 0.04
sEMG$_{P1}$ + Acc	✓	SVM_lin	1	0.88 ± 0.04	0.87 ± 0.04	0.88 ± 0.04	0.90 ± 0.08	0.85 ± 0.05	0.90 ± 0.09
		SVM_RBF	{1, 0.0001}	0.87 ± 0.03	0.86 ± 0.03	0.87 ± 0.03	0.92 ± 0.07	0.82 ± 0.07	0.92 ± 0.09
		SVM_sigmoid	{1, 0.0001}	0.83 ± 0.05	0.80 ± 0.06	0.83 ± 0.05	0.96 ± 0.04	0.68 ± 0.08	0.97 ± 0.03
		XGBoost	{2, 1}	0.83 ± 0.04	0.83 ± 0.05	0.83 ± 0.05	0.80 ± 0.06	0.87 ± 0.08	0.78 ± 0.08
		kNN	11	0.86 ± 0.06	0.84 ± 0.07	0.86 ± 0.06	0.95 ± 0.03	0.76 ± 0.09	0.96 ± 0.02
	✗	SVM_lin	1	0.82 ± 0.03	0.82 ± 0.03	0.82 ± 0.03	0.84 ± 0.05	0.80 ± 0.05	0.85 ± 0.05
		SVM_RBF	{1, 0.0001}	0.79 ± 0.04	0.78 ± 0.04	0.79 ± 0.04	0.78 ± 0.07	0.79 ± 0.05	0.78 ± 0.07
		SVM_sigmoid	{1, 0.0001}	0.82 ± 0.03	0.81 ± 0.04	0.82 ± 0.03	0.85 ± 0.06	0.79 ± 0.06	0.86 ± 0.05
		XGBoost	{2, 1}	0.80 ± 0.04	0.80 ± 0.03	0.80 ± 0.04	0.81 ± 0.08	0.78 ± 0.05	0.82 ± 0.09
		kNN	6	0.75 ± 0.08	0.70 ± 0.14	0.76 ± 0.08	0.86 ± 0.11	0.63 ± 0.21	0.87 ± 0.16
sEMG$_{P2}$ + Acc	✓	SVM_lin	1	0.82 ± 0.02	0.81 ± 0.04	0.82 ± 0.02	0.84 ± 0.03	0.79 ± 0.07	0.86 ± 0.04
		SVM_RBF	{10, 0.001}	0.89 ± 0.02	0.89 ± 0.03	0.89 ± 0.02	0.89 ± 0.02	0.89 ± 0.05	0.89 ± 0.02
		SVM_sigmoid	{1, 0.001}	0.89 ± 0.04	0.89 ± 0.04	0.89 ± 0.04	0.89 ± 0.04	0.89 ± 0.05	0.89 ± 0.04
		XGBoost	{2, 1}	0.90 ± 0.05	0.90 ± 0.05	0.90 ± 0.04	0.88 ± 0.03	0.92 ± 0.08	0.88 ± 0.02
		kNN	14	0.87 ± 0.05	0.87 ± 0.06	0.87 ± 0.05	0.87 ± 0.02	0.87 ± 0.10	0.88 ± 0.02
	✗	SVM_lin	1	0.83 ± 0.04	0.83 ± 0.04	0.83 ± 0.04	0.83 ± 0.04	0.83 ± 0.07	0.83 ± 0.05
		SVM_RBF	{10, 0.0001}	0.81 ± 0.04	0.81 ± 0.04	0.81 ± 0.04	0.80 ± 0.06	0.82 ± 0.07	0.80 ± 0.07
		SVM_sigmoid	{1, 0.0001}	0.83 ± 0.03	0.83 ± 0.03	0.83 ± 0.03	0.84 ± 0.05	0.83 ± 0.05	0.84 ± 0.06
		XGBoost	{2, 1}	0.86 ± 0.06	0.85 ± 0.07	0.86 ± 0.06	0.84 ± 0.04	0.87 ± 0.11	0.84 ± 0.05
		kNN	18	0.80 ± 0.07	0.77 ± 0.08	0.80 ± 0.08	0.87 ± 0.12	0.71 ± 0.11	0.88 ± 0.13
sEMG$_{P1}$ + Speech	✓	SVM_lin	1	0.96 ± 0.02	0.95 ± 0.03	0.96 ± 0.02	0.98 ± 0.02	0.93 ± 0.05	0.98 ± 0.02
		SVM_RBF	{1, 0.01}	0.87 ± 0.06	0.88 ± 0.05	0.87 ± 0.06	0.81 ± 0.11	0.98 ± 0.04	0.76 ± 0.15
		SVM_sigmoid	{1, 0.0001}	0.95 ± 0.03	0.95 ± 0.03	0.95 ± 0.03	1.00 ± 0.00	0.90 ± 0.05	1.00 ± 0.00
		XGBoost	{2, 10}	0.92 ± 0.04	0.92 ± 0.04	0.92 ± 0.04	0.91 ± 0.05	0.94 ± 0.06	0.91 ± 0.06
		kNN	10	0.92 ± 0.03	0.92 ± 0.03	0.92 ± 0.03	0.98 ± 0.03	0.87 ± 0.06	0.98 ± 0.03
	✗	SVM_lin	1	0.90 ± 0.04	0.89 ± 0.04	0.90 ± 0.03	0.96 ± 0.04	0.84 ± 0.07	0.96 ± 0.04
		SVM_RBF	{10, 0.0001}	0.85 ± 0.04	0.85 ± 0.04	0.85 ± 0.04	0.84 ± 0.08	0.86 ± 0.06	0.84 ± 0.09
		SVM_sigmoid	{1, 0.0001}	0.89 ± 0.03	0.88 ± 0.04	0.89 ± 0.03	0.96 ± 0.04	0.82 ± 0.06	0.96 ± 0.04
		XGBoost	{2, 10}	0.91 ± 0.05	0.90 ± 0.06	0.91 ± 0.05	0.91 ± 0.05	0.91 ± 0.12	0.91 ± 0.06
		kNN	8	0.73 ± 0.13	0.60 ± 0.27	0.73 ± 0.12	0.96 ± 0.06	0.48 ± 0.28	0.97 ± 0.04
sEMG$_{P2}$ + Speech	✓	SVM_lin	1	0.97 ± 0.02	0.97 ± 0.02	0.97 ± 0.02	0.97 ± 0.03	0.97 ± 0.03	0.97 ± 0.03
		SVM_RBF	{10, 0.0001}	0.96 ± 0.01	0.96 ± 0.01	0.96 ± 0.01	0.96 ± 0.02	0.96 ± 0.02	0.96 ± 0.02
		SVM_sigmoid	{1, 0.001}	0.96 ± 0.01	0.96 ± 0.01	0.96 ± 0.01	0.95 ± 0.03	0.96 ± 0.02	0.95 ± 0.03
		XGBoost	{2, 1}	0.92 ± 0.05	0.92 ± 0.05	0.92 ± 0.05	0.93 ± 0.02	0.91 ± 0.09	0.93 ± 0.02
		kNN	10	0.94 ± 0.01	0.94 ± 0.01	0.94 ± 0.01	0.93 ± 0.03	0.95 ± 0.01	0.93 ± 0.03
	✗	SVM_lin	1	0.95 ± 0.02	0.95 ± 0.02	0.95 ± 0.02	0.97 ± 0.03	0.93 ± 0.03	0.97 ± 0.02
		SVM_RBF	{10, 0.0001}	0.93 ± 0.02	0.93 ± 0.02	0.93 ± 0.02	0.93 ± 0.02	0.93 ± 0.03	0.93 ± 0.02
		SVM_sigmoid	{10, 0.0001}	0.96 ± 0.02	0.95 ± 0.02	0.96 ± 0.02	0.98 ± 0.03	0.93 ± 0.03	0.98 ± 0.03
		XGBoost	{2, 1}	0.93 ± 0.03	0.93 ± 0.03	0.93 ± 0.03	0.90 ± 0.03	0.96 ± 0.04	0.89 ± 0.05
		kNN	3	0.80 ± 0.12	0.73 ± 0.22	0.80 ± 0.12	0.95 ± 0.05	0.65 ± 0.29	0.95 ± 0.05
Acc + Speech	✓	SVM_lin	1	0.96 ± 0.03	0.96 ± 0.03	0.96 ± 0.03	0.98 ± 0.03	0.94 ± 0.06	0.98 ± 0.02
		SVM_RBF	{1, 0.001}	0.98 ± 0.02	0.98 ± 0.02	0.98 ± 0.02	0.98 ± 0.02	0.97 ± 0.03	0.98 ± 0.03
		SVM_sigmoid	{1, 0.001}	0.97 ± 0.01	0.97 ± 0.01	0.97 ± 0.01	0.97 ± 0.02	0.97 ± 0.03	0.97 ± 0.02
		XGBoost	{2, 1}	0.93 ± 0.04	0.93 ± 0.04	0.93 ± 0.04	0.93 ± 0.04	0.93 ± 0.05	0.93 ± 0.04
		kNN	11	0.97 ± 0.02	0.96 ± 0.02	0.97 ± 0.02	0.97 ± 0.03	0.96 ± 0.02	0.97 ± 0.03
	✗	SVM_lin	1	0.94 ± 0.04	0.94 ± 0.04	0.94 ± 0.04	0.95 ± 0.05	0.93 ± 0.04	0.95 ± 0.05
		SVM_RBF	{10, 0.0001}	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.02	0.95 ± 0.05	0.92 ± 0.03	0.95 ± 0.05
		SVM_sigmoid	{10, 0.0001}	0.93 ± 0.04	0.92 ± 0.04	0.93 ± 0.04	0.95 ± 0.05	0.90 ± 0.08	0.95 ± 0.05
		XGBoost	{2, 1}	0.92 ± 0.03	0.92 ± 0.03	0.92 ± 0.03	0.93 ± 0.03	0.91 ± 0.07	0.93 ± 0.03
		kNN	6	0.80 ± 0.07	0.75 ± 0.11	0.80 ± 0.07	0.95 ± 0.09	0.64 ± 0.14	0.96 ± 0.07

The hyperparameters are provided in the following order: SVM: {C, γ} (for linear kernel, γ doesn't apply); XGBoost: {depth, negative/positive balance}; kNN: {neighbors}.

Table E.6: Classification results for multimodal scenarios using classical machine learning algorithms and 35% of the database for test.

Scenario	Feature selection	Classifier	Hyperparameters mode	AUC	F1	Accuracy	Precision	Sensitivity	Specificity
sEMG$_{P1}$ + Acc + Speech	✓	SVM_lin	1	0.94 ± 0.02	0.95 ± 0.01	0.95 ± 0.02	0.95 ± 0.04	0.95 ± 0.02	0.94 ± 0.05
		SVM_RBF	{1, 0.001}	0.92 ± 0.07	0.93 ± 0.05	0.92 ± 0.07	0.91 ± 0.12	0.97 ± 0.04	0.87 ± 0.17
		SVM_sigmoid	{1, 0.001}	0.93 ± 0.07	0.94 ± 0.04	0.94 ± 0.06	0.95 ± 0.09	0.94 ± 0.02	0.92 ± 0.15
		XGBoost	{5, 1}	0.93 ± 0.05	0.94 ± 0.04	0.93 ± 0.05	0.92 ± 0.06	0.95 ± 0.05	0.91 ± 0.07
		kNN	21	0.96 ± 0.01	0.96 ± 0.01	0.96 ± 0.01	0.98 ± 0.02	0.94 ± 0.02	0.98 ± 0.02
	✗	SVM_lin	1	0.89 ± 0.04	0.90 ± 0.03	0.89 ± 0.04	0.89 ± 0.05	0.90 ± 0.05	0.88 ± 0.07
		SVM_RBF	{10, 0.0001}	0.86 ± 0.03	0.86 ± 0.02	0.86 ± 0.03	0.85 ± 0.06	0.88 ± 0.03	0.84 ± 0.08
		SVM_sigmoid	{1, 0.0001}	0.90 ± 0.04	0.90 ± 0.03	0.89 ± 0.03	0.91 ± 0.07	0.89 ± 0.03	0.91 ± 0.07
		XGBoost	{2, 1}	0.91 ± 0.02	0.92 ± 0.02	0.91 ± 0.02	0.91 ± 0.04	0.93 ± 0.03	0.90 ± 0.05
		kNN	28	0.80 ± 0.05	0.79 ± 0.06	0.80 ± 0.04	0.86 ± 0.07	0.73 ± 0.06	0.88 ± 0.05
sEMG$_{P2}$ + Acc + Speech	✓	SVM_lin	1	0.93 ± 0.03	0.93 ± 0.03	0.93 ± 0.03	0.92 ± 0.06	0.94 ± 0.02	0.91 ± 0.07
		SVM_RBF	{1, 0.001}	0.93 ± 0.03	0.94 ± 0.02	0.93 ± 0.03	0.92 ± 0.05	0.95 ± 0.02	0.91 ± 0.06
		SVM_sigmoid	{1, 0.001}	0.94 ± 0.03	0.94 ± 0.03	0.94 ± 0.03	0.93 ± 0.06	0.95 ± 0.02	0.92 ± 0.07
		XGBoost	{2, 1}	0.89 ± 0.06	0.90 ± 0.04	0.90 ± 0.05	0.88 ± 0.07	0.93 ± 0.06	0.85 ± 0.14
		kNN	9	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.03	0.95 ± 0.02	0.93 ± 0.04
	✗	SVM_lin	1	0.91 ± 0.03	0.92 ± 0.03	0.91 ± 0.03	0.92 ± 0.07	0.92 ± 0.05	0.91 ± 0.08
		SVM_RBF	{10, 0.0001}	0.88 ± 0.07	0.89 ± 0.05	0.88 ± 0.07	0.89 ± 0.10	0.90 ± 0.07	0.87 ± 0.13
		SVM_sigmoid	{1, 0.0001}	0.91 ± 0.03	0.91 ± 0.03	0.91 ± 0.03	0.93 ± 0.07	0.90 ± 0.06	0.92 ± 0.08
		XGBoost	{2, 1}	0.91 ± 0.03	0.91 ± 0.03	0.91 ± 0.03	0.91 ± 0.05	0.92 ± 0.05	0.91 ± 0.07
		kNN	4	0.84 ± 0.13	0.78 ± 0.24	0.83 ± 0.16	0.92 ± 0.06	0.75 ± 0.31	0.92 ± 0.05
sEMG$_{P1}$ + Acc	✓	SVM_lin	1	0.88 ± 0.06	0.88 ± 0.06	0.88 ± 0.06	0.91 ± 0.07	0.85 ± 0.07	0.91 ± 0.07
		SVM_RBF	{1, 0.0001}	0.86 ± 0.05	0.85 ± 0.05	0.86 ± 0.05	0.91 ± 0.09	0.81 ± 0.07	0.90 ± 0.12
		SVM_sigmoid	{1, 0.001}	0.88 ± 0.04	0.88 ± 0.05	0.88 ± 0.04	0.95 ± 0.06	0.82 ± 0.07	0.95 ± 0.05
		XGBoost	{2, 1}	0.81 ± 0.05	0.82 ± 0.05	0.81 ± 0.05	0.83 ± 0.05	0.80 ± 0.06	0.82 ± 0.08
		kNN	21	0.87 ± 0.04	0.86 ± 0.05	0.87 ± 0.04	0.93 ± 0.06	0.79 ± 0.05	0.94 ± 0.05
	✗	SVM_lin	1	0.85 ± 0.02	0.85 ± 0.02	0.85 ± 0.02	0.87 ± 0.05	0.83 ± 0.04	0.87 ± 0.06
		SVM_RBF	{1, 0.0001}	0.82 ± 0.05	0.82 ± 0.04	0.81 ± 0.05	0.83 ± 0.06	0.81 ± 0.04	0.82 ± 0.08
		SVM_sigmoid	{1, 0.0001}	0.83 ± 0.03	0.83 ± 0.03	0.83 ± 0.04	0.86 ± 0.07	0.81 ± 0.02	0.85 ± 0.08
		XGBoost	{2, 1}	0.76 ± 0.04	0.77 ± 0.05	0.76 ± 0.04	0.75 ± 0.04	0.80 ± 0.07	0.72 ± 0.03
		kNN	19	0.75 ± 0.04	0.74 ± 0.03	0.74 ± 0.04	0.79 ± 0.09	0.70 ± 0.05	0.79 ± 0.11
sEMG$_{P2}$ + Acc	✓	SVM_lin	1	0.85 ± 0.05	0.85 ± 0.04	0.85 ± 0.05	0.86 ± 0.07	0.85 ± 0.03	0.84 ± 0.09
		SVM_RBF	{1, 0.001}	0.87 ± 0.06	0.88 ± 0.06	0.87 ± 0.06	0.89 ± 0.08	0.87 ± 0.08	0.88 ± 0.10
		SVM_sigmoid	{1, 0.001}	0.88 ± 0.06	0.89 ± 0.06	0.88 ± 0.06	0.89 ± 0.08	0.89 ± 0.05	0.88 ± 0.10
		XGBoost	{2, 1}	0.90 ± 0.06	0.91 ± 0.06	0.90 ± 0.06	0.90 ± 0.05	0.92 ± 0.08	0.89 ± 0.05
		kNN	6	0.90 ± 0.05	0.90 ± 0.06	0.90 ± 0.05	0.92 ± 0.05	0.88 ± 0.08	0.91 ± 0.07
	✗	SVM_lin	1	0.83 ± 0.04	0.83 ± 0.03	0.83 ± 0.04	0.86 ± 0.06	0.81 ± 0.02	0.85 ± 0.07
		SVM_RBF	{10, 0.0001}	0.82 ± 0.04	0.82 ± 0.03	0.82 ± 0.04	0.84 ± 0.07	0.80 ± 0.02	0.83 ± 0.10
		SVM_sigmoid	{1, 0.0001}	0.81 ± 0.05	0.82 ± 0.04	0.81 ± 0.05	0.82 ± 0.05	0.81 ± 0.07	0.81 ± 0.07
		XGBoost	{2, 1}	0.84 ± 0.06	0.84 ± 0.06	0.84 ± 0.06	0.84 ± 0.05	0.85 ± 0.10	0.83 ± 0.05
		kNN	7	0.74 ± 0.07	0.69 ± 0.15	0.73 ± 0.09	0.84 ± 0.13	0.65 ± 0.26	0.82 ± 0.20
sEMG$_{P1}$ + Speech	✓	SVM_lin	1	0.95 ± 0.03	0.95 ± 0.03	0.95 ± 0.03	0.96 ± 0.06	0.95 ± 0.03	0.95 ± 0.08
		SVM_RBF	{1, 0.0001}	0.95 ± 0.02	0.96 ± 0.01	0.95 ± 0.02	0.95 ± 0.05	0.97 ± 0.03	0.94 ± 0.07
		SVM_sigmoid	{1, 0.01}	0.95 ± 0.04	0.95 ± 0.03	0.95 ± 0.03	0.96 ± 0.07	0.95 ± 0.05	0.94 ± 0.10
		XGBoost	{2, 1}	0.92 ± 0.03	0.92 ± 0.03	0.92 ± 0.03	0.91 ± 0.04	0.93 ± 0.04	0.90 ± 0.05
		kNN	3	0.95 ± 0.02	0.95 ± 0.02	0.95 ± 0.02	0.97 ± 0.03	0.93 ± 0.04	0.96 ± 0.04
	✗	SVM_lin	1	0.89 ± 0.04	0.89 ± 0.04	0.89 ± 0.04	0.91 ± 0.08	0.88 ± 0.06	0.90 ± 0.10
		SVM_RBF	{10, 0.0001}	0.86 ± 0.05	0.87 ± 0.05	0.86 ± 0.05	0.85 ± 0.06	0.88 ± 0.06	0.84 ± 0.06
		SVM_sigmoid	{1, 0.0001}	0.90 ± 0.03	0.89 ± 0.04	0.89 ± 0.03	0.92 ± 0.06	0.87 ± 0.06	0.92 ± 0.06
		XGBoost	{2, 1}	0.93 ± 0.03	0.93 ± 0.03	0.93 ± 0.03	0.91 ± 0.04	0.96 ± 0.05	0.90 ± 0.04
		kNN	5	0.79 ± 0.04	0.79 ± 0.04	0.79 ± 0.04	0.83 ± 0.10	0.78 ± 0.12	0.81 ± 0.14
sEMG$_{P2}$ + Speech	✓	SVM_lin	1	0.92 ± 0.03	0.93 ± 0.02	0.92 ± 0.02	0.92 ± 0.03	0.93 ± 0.04	0.91 ± 0.05
		SVM_RBF	{10, 0.001}	0.94 ± 0.03	0.94 ± 0.02	0.94 ± 0.03	0.92 ± 0.05	0.96 ± 0.03	0.91 ± 0.05
		SVM_sigmoid	{1, 0.001}	0.94 ± 0.03	0.94 ± 0.03	0.94 ± 0.03	0.94 ± 0.05	0.95 ± 0.02	0.93 ± 0.06
		XGBoost	{2, 1}	0.93 ± 0.03	0.93 ± 0.03	0.93 ± 0.03	0.93 ± 0.03	0.94 ± 0.07	0.92 ± 0.03
		kNN	7	0.94 ± 0.01	0.94 ± 0.01	0.94 ± 0.01	0.94 ± 0.02	0.95 ± 0.02	0.93 ± 0.02
	✗	SVM_lin	1	0.94 ± 0.03	0.94 ± 0.03	0.94 ± 0.03	0.94 ± 0.06	0.94 ± 0.04	0.93 ± 0.07
		SVM_RBF	{10, 0.0001}	0.93 ± 0.04	0.93 ± 0.04	0.93 ± 0.04	0.93 ± 0.06	0.93 ± 0.04	0.92 ± 0.06
		SVM_sigmoid	{10, 0.0001}	0.94 ± 0.03	0.94 ± 0.03	0.94 ± 0.03	0.95 ± 0.06	0.94 ± 0.04	0.94 ± 0.07
		XGBoost	{2, 1}	0.92 ± 0.03	0.93 ± 0.03	0.92 ± 0.03	0.91 ± 0.02	0.94 ± 0.07	0.91 ± 0.02
		kNN	14	0.82 ± 0.09	0.79 ± 0.12	0.82 ± 0.09	0.92 ± 0.05	0.71 ± 0.19	0.93 ± 0.05
Acc + Speech	✓	SVM_lin	1	0.92 ± 0.04	0.92 ± 0.03	0.92 ± 0.04	0.92 ± 0.05	0.93 ± 0.04	0.90 ± 0.09
		SVM_RBF	{1, 0.001}	0.95 ± 0.02	0.95 ± 0.02	0.95 ± 0.02	0.96 ± 0.04	0.94 ± 0.02	0.96 ± 0.04
		SVM_sigmoid	{1, 0.001}	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.02	0.93 ± 0.02	0.94 ± 0.03
		XGBoost	{2, 1}	0.91 ± 0.04	0.92 ± 0.03	0.91 ± 0.04	0.90 ± 0.05	0.93 ± 0.05	0.88 ± 0.07
		kNN	11	0.95 ± 0.02	0.95 ± 0.02	0.95 ± 0.02	0.97 ± 0.03	0.94 ± 0.02	0.97 ± 0.04
	✗	SVM_lin	1	0.93 ± 0.03	0.94 ± 0.02	0.93 ± 0.03	0.94 ± 0.03	0.93 ± 0.02	0.93 ± 0.04
		SVM_RBF	{10, 0.0001}	0.91 ± 0.05	0.91 ± 0.04	0.91 ± 0.05	0.91 ± 0.05	0.92 ± 0.04	0.89 ± 0.08
		SVM_sigmoid	{10, 0.0001}	0.91 ± 0.05	0.91 ± 0.05	0.90 ± 0.05	0.92 ± 0.03	0.89 ± 0.06	0.92 ± 0.04
		XGBoost	{2, 1}	0.93 ± 0.04	0.93 ± 0.03	0.93 ± 0.04	0.93 ± 0.06	0.94 ± 0.06	0.91 ± 0.07
		kNN	6	0.82 ± 0.08	0.78 ± 0.15	0.81 ± 0.10	0.93 ± 0.07	0.71 ± 0.20	0.93 ± 0.06

The hyperparameters are provided in the following order: SVM: {C, γ} (for linear kernel, γ doesn't apply); XGBoost: {depth, negative/positive balance}; kNN: {neighbors}.

Table E.7: Classification results for multimodal scenarios using classical machine learning algorithms and 40% of the database for test.

Scenario	Feature selection	Classifier	Hyperparameters mode	AUC	F1	Accuracy	Sensitivity	Specificity	Precision
sEMG$_{P1}$ + Acc + Speech	✓	SVM_lin	1	0.96 ± 0.03	0.96 ± 0.03	0.96 ± 0.03	0.93 ± 0.05	0.98 ± 0.03	0.98 ± 0.03
		SVM_RBF	{1, 0.001}	0.95 ± 0.04	0.95 ± 0.03	0.95 ± 0.04	0.96 ± 0.05	0.94 ± 0.10	0.95 ± 0.08
		SVM_sigmoid	{1, 0.0001}	0.96 ± 0.03	0.96 ± 0.03	0.96 ± 0.03	0.93 ± 0.06	1.00 ± 0.00	1.00 ± 0.00
		XGBoost	{2, 1}	0.91 ± 0.06	0.91 ± 0.06	0.91 ± 0.06	0.93 ± 0.07	0.88 ± 0.06	0.89 ± 0.05
		kNN	10	0.95 ± 0.03	0.95 ± 0.03	0.95 ± 0.03	0.93 ± 0.06	0.98 ± 0.02	0.98 ± 0.02
	✗	SVM_lin	1	0.90 ± 0.03	0.90 ± 0.03	0.90 ± 0.03	0.86 ± 0.06	0.94 ± 0.05	0.94 ± 0.05
		SVM_RBF	{1, 0.0001}	0.88 ± 0.03	0.88 ± 0.04	0.88 ± 0.03	0.88 ± 0.08	0.88 ± 0.05	0.89 ± 0.04
		SVM_sigmoid	{1, 0.0001}	0.90 ± 0.03	0.90 ± 0.03	0.90 ± 0.03	0.87 ± 0.05	0.94 ± 0.04	0.94 ± 0.04
		XGBoost	{2, 1}	0.90 ± 0.03	0.91 ± 0.03	0.90 ± 0.03	0.94 ± 0.03	0.86 ± 0.08	0.87 ± 0.06
		kNN	18	0.82 ± 0.06	0.79 ± 0.09	0.82 ± 0.07	0.70 ± 0.15	0.95 ± 0.04	0.94 ± 0.04
sEMG$_{P2}$ + Acc + Speech	✓	SVM_lin	1	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.02	0.95 ± 0.05	0.93 ± 0.03	0.94 ± 0.03
		SVM_RBF	{1, 0.001}	0.93 ± 0.03	0.93 ± 0.03	0.93 ± 0.03	0.94 ± 0.04	0.91 ± 0.04	0.91 ± 0.04
		SVM_sigmoid	{1, 0.001}	0.92 ± 0.03	0.93 ± 0.03	0.93 ± 0.03	0.94 ± 0.04	0.91 ± 0.02	0.91 ± 0.02
		XGBoost	{2, 1}	0.89 ± 0.03	0.89 ± 0.03	0.89 ± 0.03	0.91 ± 0.07	0.86 ± 0.04	0.87 ± 0.03
		kNN	9	0.94 ± 0.03	0.94 ± 0.03	0.94 ± 0.03	0.93 ± 0.05	0.94 ± 0.03	0.94 ± 0.03
	✗	SVM_lin	1	0.91 ± 0.03	0.92 ± 0.02	0.91 ± 0.03	0.92 ± 0.03	0.91 ± 0.05	0.92 ± 0.04
		SVM_RBF	{10, 0.0001}	0.91 ± 0.02	0.91 ± 0.02	0.91 ± 0.02	0.92 ± 0.04	0.89 ± 0.04	0.90 ± 0.03
		SVM_sigmoid	{1, 0.0001}	0.92 ± 0.03	0.92 ± 0.03	0.92 ± 0.03	0.93 ± 0.04	0.91 ± 0.05	0.92 ± 0.04
		XGBoost	{2, 1}	0.89 ± 0.03	0.90 ± 0.03	0.89 ± 0.03	0.91 ± 0.07	0.87 ± 0.04	0.88 ± 0.03
		kNN	4	0.81 ± 0.04	0.80 ± 0.08	0.81 ± 0.04	0.77 ± 0.18	0.85 ± 0.13	0.87 ± 0.09
sEMG$_{P1}$ + Acc	✓	SVM_lin	1	0.89 ± 0.04	0.89 ± 0.03	0.89 ± 0.04	0.90 ± 0.05	0.88 ± 0.08	0.90 ± 0.07
		SVM_RBF	{1, 0.0001}	0.90 ± 0.04	0.90 ± 0.03	0.90 ± 0.04	0.86 ± 0.06	0.94 ± 0.08	0.94 ± 0.06
		SVM_sigmoid	{1, 0.001}	0.84 ± 0.11	0.80 ± 0.16	0.83 ± 0.11	0.72 ± 0.24	0.95 ± 0.06	0.96 ± 0.06
		XGBoost	{2, 1}	0.79 ± 0.05	0.81 ± 0.04	0.79 ± 0.05	0.84 ± 0.09	0.74 ± 0.16	0.79 ± 0.11
		kNN	8	0.90 ± 0.04	0.90 ± 0.04	0.90 ± 0.04	0.86 ± 0.07	0.94 ± 0.03	0.94 ± 0.03
	✗	SVM_lin	1	0.84 ± 0.03	0.84 ± 0.03	0.84 ± 0.03	0.82 ± 0.07	0.85 ± 0.08	0.86 ± 0.06
		SVM_RBF	{1, 0.0001}	0.86 ± 0.03	0.86 ± 0.04	0.86 ± 0.03	0.87 ± 0.08	0.85 ± 0.05	0.86 ± 0.03
		SVM_sigmoid	{1, 0.0001}	0.85 ± 0.04	0.85 ± 0.04	0.85 ± 0.04	0.84 ± 0.07	0.86 ± 0.07	0.87 ± 0.06
		XGBoost	{2, 1}	0.78 ± 0.04	0.79 ± 0.03	0.78 ± 0.03	0.80 ± 0.08	0.76 ± 0.13	0.79 ± 0.10
		kNN	29	0.78 ± 0.06	0.730 ± 0.1	0.77 ± 0.07	0.62 ± 0.13	0.93 ± 0.09	0.92 ± 0.09
sEMG$_{P2}$ + Acc	✓	SVM_lin	1	0.86 ± 0.03	0.86 ± 0.03	0.86 ± 0.03	0.88 ± 0.04	0.84 ± 0.06	0.85 ± 0.04
		SVM_RBF	{10, 0.0001}	0.88 ± 0.01	0.89 ± 0.02	0.88 ± 0.01	0.88 ± 0.04	0.89 ± 0.04	0.89 ± 0.04
		SVM_sigmoid	{1, 0.001}	0.89 ± 0.03	0.89 ± 0.03	0.89 ± 0.03	0.87 ± 0.08	0.91 ± 0.03	0.91 ± 0.03
		XGBoost	{2, 1}	0.87 ± 0.03	0.88 ± 0.03	0.87 ± 0.03	0.91 ± 0.07	0.82 ± 0.09	0.85 ± 0.06
		kNN	3	0.87 ± 0.04	0.87 ± 0.04	0.87 ± 0.04	0.87 ± 0.07	0.86 ± 0.06	0.87 ± 0.06
	✗	SVM_lin	1	0.82 ± 0.04	0.83 ± 0.04	0.82 ± 0.04	0.83 ± 0.06	0.82 ± 0.05	0.83 ± 0.04
		SVM_RBF	{10, 0.0001}	0.83 ± 0.05	0.84 ± 0.05	0.83 ± 0.05	0.83 ± 0.06	0.83 ± 0.05	0.84 ± 0.04
		SVM_sigmoid	{1, 0.0001}	0.82 ± 0.04	0.82 ± 0.04	0.82 ± 0.04	0.84 ± 0.06	0.79 ± 0.04	0.81 ± 0.03
		XGBoost	{2, 1}	0.83 ± 0.04	0.84 ± 0.04	0.83 ± 0.04	0.82 ± 0.03	0.85 ± 0.08	0.85 ± 0.06
		kNN	5	0.70 ± 0.10	0.71 ± 0.04	0.70 ± 0.09	0.73 ± 0.18	0.66 ± 0.36	0.76 ± 0.18
sEMG$_{P1}$ + Speech	✓	SVM_lin	1	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.01	0.93 ± 0.05	0.95 ± 0.06	0.95 ± 0.05
		SVM_RBF	{1, 0.0001}	0.94 ± 0.03	0.94 ± 0.03	0.94 ± 0.03	0.95 ± 0.06	0.92 ± 0.10	0.94 ± 0.09
		SVM_sigmoid	{1, 0.0001}	0.96 ± 0.03	0.95 ± 0.04	0.95 ± 0.03	0.93 ± 0.07	0.98 ± 0.02	0.99 ± 0.02
		XGBoost	{5, 1}	0.91 ± 0.04	0.91 ± 0.04	0.91 ± 0.04	0.93 ± 0.05	0.89 ± 0.06	0.90 ± 0.05
		kNN	8	0.96 ± 0.02	0.96 ± 0.02	0.96 ± 0.02	0.93 ± 0.05	0.98 ± 0.02	0.99 ± 0.02
	✗	SVM_lin	1	0.92 ± 0.03	0.91 ± 0.03	0.91 ± 0.03	0.87 ± 0.07	0.96 ± 0.03	0.96 ± 0.03
		SVM_RBF	{10, 0.0001}	0.90 ± 0.03	0.90 ± 0.03	0.90 ± 0.03	0.90 ± 0.05	0.91 ± 0.04	0.91 ± 0.04
		SVM_sigmoid	{1, 0.0001}	0.90 ± 0.03	0.90 ± 0.03	0.90 ± 0.03	0.85 ± 0.06	0.95 ± 0.03	0.95 ± 0.03
		XGBoost	{2, 1}	0.90 ± 0.04	0.90 ± 0.04	0.90 ± 0.04	0.92 ± 0.07	0.88 ± 0.05	0.89 ± 0.05
		kNN	4	0.72 ± 0.15	0.60 ± 0.34	0.71 ± 0.16	0.58 ± 0.41	0.85 ± 0.25	0.89 ± 0.15
sEMG$_{P2}$ + Speech	✓	SVM_lin	1	0.93 ± 0.02	0.93 ± 0.02	0.93 ± 0.02	0.93 ± 0.05	0.92 ± 0.03	0.93 ± 0.02
		SVM_RBF	{1, 0.001}	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.04	0.93 ± 0.02	0.93 ± 0.02
		SVM_sigmoid	{1, 0.001}	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.06	0.93 ± 0.03	0.94 ± 0.03
		XGBoost	{2, 1}	0.89 ± 0.04	0.89 ± 0.04	0.89 ± 0.04	0.91 ± 0.07	0.88 ± 0.04	0.89 ± 0.03
		kNN	3	0.92 ± 0.01	0.92 ± 0.02	0.92 ± 0.01	0.90 ± 0.04	0.95 ± 0.02	0.95 ± 0.02
	✗	SVM_lin	1	0.94 ± 0.01	0.94 ± 0.02	0.94 ± 0.02	0.91 ± 0.04	0.96 ± 0.03	0.96 ± 0.03
		SVM_RBF	{10, 0.0001}	0.91 ± 0.02	0.91 ± 0.02	0.91 ± 0.02	0.91 ± 0.04	0.90 ± 0.04	0.91 ± 0.04
		SVM_sigmoid	{10, 0.0001}	0.93 ± 0.02	0.93 ± 0.02	0.93 ± 0.02	0.90 ± 0.03	0.95 ± 0.04	0.96 ± 0.04
		XGBoost	{2, 1}	0.90 ± 0.05	0.90 ± 0.05	0.90 ± 0.05	0.93 ± 0.05	0.87 ± 0.06	0.88 ± 0.05
		kNN	4	0.810 ± 0.1	0.78 ± 0.16	0.81 ± 0.09	0.71 ± 0.21	0.91 ± 0.12	0.91 ± 0.10
Acc + Speech	✓	SVM_lin	1	0.93 ± 0.03	0.93 ± 0.04	0.93 ± 0.03	0.91 ± 0.07	0.95 ± 0.06	0.95 ± 0.05
		SVM_RBF	{1, 0.001}	0.92 ± 0.05	0.93 ± 0.05	0.92 ± 0.05	0.92 ± 0.07	0.93 ± 0.05	0.93 ± 0.05
		SVM_sigmoid	{1, 0.001}	0.93 ± 0.05	0.93 ± 0.06	0.93 ± 0.05	0.90 ± 0.13	0.97 ± 0.03	0.97 ± 0.03
		XGBoost	{2, 1}	0.88 ± 0.04	0.89 ± 0.04	0.88 ± 0.04	0.91 ± 0.09	0.86 ± 0.05	0.87 ± 0.04
		kNN	4	0.94 ± 0.05	0.94 ± 0.06	0.94 ± 0.05	0.90 ± 0.09	0.99 ± 0.03	0.98 ± 0.04
	✗	SVM_lin	1	0.93 ± 0.03	0.93 ± 0.03	0.93 ± 0.03	0.93 ± 0.06	0.92 ± 0.05	0.93 ± 0.04
		SVM_RBF	{10, 0.0001}	0.93 ± 0.04	0.93 ± 0.04	0.93 ± 0.04	0.92 ± 0.07	0.93 ± 0.03	0.93 ± 0.03
		SVM_sigmoid	{10, 0.0001}	0.93 ± 0.03	0.93 ± 0.03	0.93 ± 0.03	0.93 ± 0.06	0.92 ± 0.05	0.93 ± 0.04
		XGBoost	{2, 1}	0.89 ± 0.06	0.90 ± 0.05	0.89 ± 0.06	0.93 ± 0.04	0.84 ± 0.10	0.86 ± 0.08
		kNN	17	0.77 ± 0.07	0.71 ± 0.13	0.76 ± 0.07	0.61 ± 0.18	0.93 ± 0.06	0.91 ± 0.06

The hyperparameters are provided in the following order: SVM: {C, γ} (for linear kernel, γ doesn't apply); XGBoost: {depth, negative/positive balance}; kNN: {neighbors}.

Bibliography

[Afka 07]   S. Afkari *et al.* "Measuring frequency of spontaneous swallowing". *Australas Phys Eng Sci Med*, Vol. 30, No. 4, pp. 313–317, 2007.

[Alfo 13]   E. Alfonsi, R. Bergamaschi, G. Cosentino, M. Ponzio, C. Montomoli, D. Restivo, F. Brighina, S. Ravaglia, P. Prunetti, G. Bertino, *et al.* "Electrophysiological patterns of oropharyngeal swallowing in multiple sclerosis". *Clinical Neurophysiology*, Vol. 124, No. 8, pp. 1638–1645, 2013.

[Alos 21]   N. Aloshban, A. Esposito, and A. Vinciarelli. "Language or Paralanguage, This is the Problem: Comparing Depressed and Non-Depressed Speakers Through the Analysis of Gated Multimodal Units.". In: *Interspeech*, pp. 2496–2500, 2021.

[Amrh 19]   V. Amrhein, S. Greenland, and B. McShane. "Scientists rise up against statistical significance". 2019.

[Anto 10]   N. Antonios, G. Carnaby-Mann, M. Crary, L. Miller, H. Hubbard, K. Hood, R. Sambandam, A. Xavier, and S. Silliman. "Analysis of a physician tool for evaluating dysphagia on an inpatient stroke unit: the modified Mann Assessment of Swallowing Ability". *Journal of Stroke and Cerebrovascular Diseases*, Vol. 19, No. 1, pp. 49–57, 2010.

[Arch 13]   S. K. Archer, R. Garrod, N. Hart, and S. Miller. "Dysphagia in Duchenne muscular dystrophy assessed objectively by surface electromyography". *Dysphagia*, Vol. 28, No. 2, pp. 188–198, 2013.

[Arev 17]   J. Arevalo, T. Solorio, M. Montes-y Gómez, and F. A. González. "Gated multimodal units for information fusion". *arXiv preprint arXiv:1702.01992*, 2017.

[Arev 20]   J. Arevalo, T. Solorio, M. Montes-y Gomez, and F. A. González. "Gated multimodal networks". *Neural Computing and Applications*, Vol. 32, No. 14, pp. 10209–10228, 2020.

[Aria 10]   J. D. Arias-Londoño, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz, and G. Castellanos-Domínguez. "Automatic detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients". *IEEE Transactions on biomedical engineering*, Vol. 58, No. 2, pp. 370–379, 2010.

[Aria 21]   T. Arias-Vergara, P. Klumpp, J. C. Vasquez-Correa, E. Nöth, J. R. Orozco-Arroyave, and M. Schuster. "Multi-channel spectrograms for speech processing applications using deep learning methods". *Pattern Analysis and Applications*, Vol. 24, No. 2, pp. 423–431, 2021.

[Attr 18]   S. Attrill, S. White, J. Murray, S. Hammond, and S. Doeltgen. "Impact of oropharyngeal dysphagia on healthcare cost and length of stay in hospital: a systematic review". *BMC health services research*, Vol. 18, No. 1, p. 594, 2018.

[Aviv 00]   J. E. Aviv. "Prospective, randomized outcome study of endoscopy versus modified barium swallow in patients with dysphagia". *The Laryngoscope*, Vol. 110, No. 4, pp. 563–574, 2000.

[Aydo 15]   I. Aydogdu, N. Kiylioglu, S. Tarlaci, Z. Tanriverdi, S. Alpaydin, A. Acarer, L. Baysal, E. Arpaci, N. Yuceyar, Y. Secil, *et al.* "Diagnostic value of "dysphagia limit" for neurogenic dysphagia: 17 years of experience in 1278 adults". *Clinical Neurophysiology*, Vol. 126, No. 3, pp. 634–643, 2015.

[Bahi 16]   M. M. Bahia, L. F. Mourao, and R. Y. S. Chun. "Dysarthria as a predictor of dysphagia following stroke". *NeuroRehabilitation*, Vol. 38, No. 2, pp. 155–162, 2016.

[Baij 09]   L. W. Baijens and R. Speyer. "Effects of therapy for dysphagia in Parkinson's disease: systematic review". *Dysphagia*, Vol. 24, No. 1, pp. 91–102, 2009.

[Barr 14]   K. E. Barrett. "Chapter 7. Esophageal Motility". In: *Gastrointestinal Physiology, 2e*, The McGraw-Hill Companies, New York, NY, 2014.

[Barr 20]   R. R. Barreira and L. L. Ling. "Kullback–Leibler divergence and sample skewness for pathological voice quality assessment". *Biomedical Signal Processing and Control*, Vol. 57, p. 101697, 2020.

[Bass 14]   D. Bassi, A. M. Furkim, C. A. Silva, M. S. P. H. Coelho, M. R. P. Rolim, M. L. A. d. Alencar, and M. J. Machado. "Identification of risk groups for oropharyngeal dysphagia in hospitalized patients in a university hospital". In: *Codas*, pp. 17–27, SciELO Brasil, 2014.

[Berg 14]   L. Bergström, P. Svensson, and L. Hartelius. "Cervical auscultation as an adjunct to the clinical swallow examination: A comparison with fibre-optic endoscopic evaluation of swallowing". *International journal of speech-language pathology*, Vol. 16, No. 5, pp. 517–528, 2014.

[Berr 19]   D. Berrar. "Cross-Validation.". 2019.

[Bhat 17]   C. Bhat, B. Vachhani, and S. K. Kopparapu. "Automatic assessment of dysarthria severity level using audio descriptors". In: *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*, pp. 5070–5074, IEEE, 2017.

[Bish 06]   C. M. Bishop and N. M. Nasrabadi. *Pattern recognition and machine learning*. Vol. 4, Springer, 2006.

[Blau 13]   Blausen.com staff. "Upper Respiratory System — Wikipedia, The Free Encyclopedia". 2013. [Online; accessed 24-July-2018].

[Boer 01]   P. Boersma. "Praat, a system for doing phonetics by computer". *Glot. Int.*, Vol. 5, No. 9, pp. 341–345, 2001.

[Boer 93]   P. Boersma. "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound". In: *Proceedings of the institute of phonetic sciences*, pp. 97–110, Institute of Phonetic Sciences,University of Amsterdam, 1993.

[Bolo 13]   V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos. "A review of feature selection methods on synthetic data". *Knowledge and information systems*, Vol. 34, No. 3, pp. 483–519, 2013.

[Bols 13]   D. C. Bolser, C. Gestreau, K. F. Morris, P. W. Davenport, and T. E. Pitts. "Central neural circuits for coordination of swallowing, breathing, and coughing: predictions from computational modeling and simulation". *Otolaryngologic Clinics of North America*, Vol. 46, No. 6, 2013.

[Bolz 13]   G. d. P. Bolzan, M. K. Christmann, L. C. Berwig, C. C. Costa, and R. M. Rocha. "Contribution of the cervical auscultation in clinical assessment of the oropharyngeal dysphagia". *Revista CEFAC*, Vol. 15, No. 2, pp. 455–465, 2013.

[Brit 18]   D. Britton, A. Roeske, S. K. Ennis, J. O. Benditt, C. Quinn, and D. Graville. "Utility of pulse oximetry to detect aspiration: an evidence-based systematic review". *Dysphagia*, Vol. 33, No. 3, pp. 282–292, 2018.

[Brou 00]   D. L. Broussard and S. M. Altschuler. "Central integration of swallow and airway-protective reflexes". *The American journal of medicine*, Vol. 108, No. 4, pp. 62–67, 2000.

[Brui 13]   M. J. de Bruijn, R. N. Rinkel, I. C. Cnossen, B. I. Witte, J. A. Langendijk, C. R. Leemans, and I. M. Verdonck-de Leeuw. "Associations between voice quality and swallowing function in patients treated for oral or oropharyngeal cancer". *Supportive Care in Cancer*, Vol. 21, No. 7, pp. 2025–2032, 2013.

[Cao 97]   L. Cao. "Practical method for determining the minimum embedding dimension of a scalar time series". *Physica D: Nonlinear Phenomena*, Vol. 110, No. 1-2, pp. 43–50, 1997.

[Card 10]   M. Cardoso and D. H. Gomes. "Ausculta cervical em adultos sem queixas de alteração na deglutição". *Arq Int Otorrinolaringol*, Vol. 14, No. 4, pp. 404–9, 2010.

[Carn 08]   G. Carnaby-Mann and K. Lenius. "The bedside examination in dysphagia". *Physical medicine and rehabilitation clinics of North America*, Vol. 19, No. 4, pp. 747–768, 2008.

[Cart 15]   S. Carter and G. Gutierrez. "The concurrent validity of three computerized methods of muscle activity onset detection". *Journal of electromyography and kinesiology*, Vol. 25, No. 5, pp. 731–741, 2015.

[Cavi 10]   I. R. Caviedes, P. M. Lavados, A. J. Hoppe, and M. A. López. "Nasolaryngoscopic validation of a set of clinical predictors of aspiration in a critical care setting". *Journal of bronchology & interventional pulmonology*, Vol. 17, No. 1, pp. 33–38, 2010.

[Chan 12]   H. Y. Chang, P. C. Torng, T. G. Wang, and Y. C. Chang. "Acoustic voice analysis does not identify presence of penetration/aspiration as confirmed by videofluoroscopic swallowing study". *Archives of physical medicine and rehabilitation*, Vol. 93, No. 11, pp. 1991–1994, 2012.

[Chen 16a]   D. F. Chen. "Dysphagia in the Hospitalized Patient". *Volume 6, Issue 1, An Issue of Hospital Medicine Clinics, E-Book*, Vol. 6, No. 1, p. 38, 2016.

[Chen 16b]   T. Chen and C. Guestrin. "Xgboost: A scalable tree boosting system". In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.

[Chow 13]   R. H. Chowdhury, M. B. Reaz, M. A. Ali, A. A. Bakar, K. Chellappan, and T. G. Chang. "Surface electromyography signal processing and classification techniques". *Sensors*, Vol. 13, No. 9, pp. 12431–12466, 2013.

[Cich 12]   J. A. Cichero and K. W. Altman. "Definition, prevalence and burden of oropharyngeal dysphagia: a serious problem among older adults worldwide and the impact on prognosis and hospital resources". In: *Stepping stones to living well with dysphagia*, pp. 1–11, Karger Publishers, 2012.

[Clav 06]   P. Clavé, M. De Kraa, V. Arreola, M. Girvent, R. Farre, E. Palomera, and M. Serra-Prat. "The effect of bolus viscosity on swallowing function in neurogenic dysphagia". *Alimentary pharmacology & therapeutics*, Vol. 24, No. 9, pp. 1385–1394, 2006.

[Clav 15]   P. Clavé and R. Shaker. "Dysphagia: current reality and scope of the problem". *Nature Reviews Gastroenterology and Hepatology*, Vol. 12, No. 5, p. 259, 2015.

[Cohe 16]   J. F. Cohen, D. A. Korevaar, D. G. Altman, D. E. Bruns, C. A. Gatsonis, L. Hooft, L. Irwig, D. Levine, J. B. Reitsma, H. C. De Vet, *et al.* "STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration". *BMJ open*, Vol. 6, No. 11, p. e012799, 2016.

[Cons 17]   G. Constantinescu, W. Hodgetts, D. Scott, K. Kuffel, B. King, C. Brodt, and J. Rieger. "Electromyography and mechanomyography signals during swallowing in healthy adults and head and neck cancer survivors". *Dysphagia*, Vol. 32, No. 1, pp. 90–103, 2017.

[Cons 18]   G. Constantinescu, K. Kuffel, D. Aalto, W. Hodgetts, and J. Rieger. "Evaluation of an automated swallow-detection algorithm using visual biofeedback in healthy adults and head and neck cancer survivors". *Dysphagia*, Vol. 33, No. 3, pp. 345–357, 2018.

[Cook 99]   I. J. Cook and P. J. Kahrilas. "AGA technical review on management of oropharyngeal dysphagia". *Gastroenterology*, Vol. 116, No. 2, pp. 455–478, 1999.

[Cort 95]   C. Cortes and V. Vapnik. "Support-vector networks". *Machine learning*, Vol. 20, No. 3, pp. 273–297, 1995.

[Cost 18]   M. M. B. Costa. "Neural control of swallowing". *Arquivos de gastroenterologia*, Vol. 55, pp. 61–75, 2018.

[Crar 05]   M. A. Crary, G. D. C. Mann, and M. E. Groher. "Initial psychometric assessment of a functional oral intake scale for dysphagia in stroke patients". *Archives of physical medicine and rehabilitation*, Vol. 86, No. 8, pp. 1516–1520, 2005.

[Crar 07]   M. A. Crary, G. D. Carnaby, and M. E. Groher. "Identification of swallowing events from sEMG signals obtained from healthy adults". *Dysphagia*, Vol. 22, No. 2, pp. 94–99, 2007.

[Crar 13a]   M. A. Crary, G. D. Carnaby, I. Sia, A. Khanna, and M. F. Waters. "Spontaneous swallowing frequency has potential to identify dysphagia in acute stroke". *Stroke*, Vol. 44, No. 12, pp. 3452–3457, 2013.

[Crar 13b]   M. A. Crary, L. Sura, and G. Carnaby. "Validation and demonstration of an isolated acoustic recording technique to estimate spontaneous swallow frequency". *Dysphagia*, Vol. 28, No. 1, pp. 86–94, 2013.

[Crar 97]   M. A. Crary and B. O. Baldwin. "Surface electromyographic characteristics of swallowing in dysphagia secondary to brainstem stroke". *Dysphagia*, Vol. 12, No. 4, pp. 180–187, 1997.

[Dani 15]   S. K. Daniels, S. Pathak, C. B. Stach, T. M. Mohr, R. O. Morgan, and J. A. Anderson. "Speech pathology reliability for stroke swallowing screening items". *Dysphagia*, Vol. 30, No. 5, pp. 565–570, 2015.

[Dani 97]   S. K. Daniels, C. P. McAdam, K. Brailey, and A. L. Foundas. "Clinical assessment of swallowing and prediction of dysphagia severity". *American Journal of Speech-Language Pathology*, Vol. 6, No. 4, pp. 17–24, 1997.

[Davi 80]   S. Davis and P. Mermelstein. "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences". *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 28, No. 4, pp. 357–366, 1980.

[Daza 09]   G. Daza-Santacoloma, J. D. Arias-Londono, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz, and G. Castellanos-Dominguez. "Dynamic feature extraction: an application to voice pathology detection". *Intelligent Automation & Soft Computing*, Vol. 15, No. 4, pp. 667–682, 2009.

[Delg 19]   A. Delgado-Bonal and A. Marshak. "Approximate entropy and sample entropy: A comprehensive tutorial". *Entropy*, Vol. 21, No. 6, p. 541, 2019.

[Dell 18]   C. Dellavia, R. Rosati, F. Musto, G. Pellegrini, G. Begnoni, and V. Ferrario. "Preliminary approach for the surface electromyographical evaluation of the oral phase of swallowing". *Journal of oral rehabilitation*, Vol. 45, No. 7, pp. 518–525, 2018.

[DePi 92]   K. L. DePippo, M. A. Holas, and M. J. Reding. "Validation of the 3-oz water swallow test for aspiration following stroke". *Archives of neurology*, Vol. 49, No. 12, pp. 1259–1261, 1992.

[Ding 02]   R. Ding, C. R. Larson, J. A. Logemann, and A. W. Rademaker. "Surface electromyographic and electroglottographic studies in normal subjects under two swallow conditions: normal and during the Mendelsohn manuever". *Dysphagia*, Vol. 17, No. 1, pp. 1–12, 2002.

[Ding 05]   C. Ding and H. Peng. "Minimum redundancy feature selection from microarray gene expression data". *Journal of bioinformatics and computational biology*, Vol. 3, No. 02, pp. 185–205, 2005.

[Dono 20]   C. Donohue, Y. Khalifa, S. Perera, E. Sejdić, and J. L. Coyle. "How closely do machine ratings of duration of UES opening during videofluoroscopy approximate clinician ratings using temporal kinematic analyses and the mbsimp?". *Dysphagia*, pp. 1–12, 2020.

[Dono 21a]   C. Donohue, Y. Khalifa, S. Mao, S. Perera, E. Sejdić, and J. L. Coyle. "Characterizing Swallows From People With Neurodegenerative Diseases Using High-Resolution Cervical Auscultation Signals and Temporal and Spatial Swallow Kinematic Measurements". *Journal of Speech, Language, and Hearing Research*, pp. 1–16, 2021.

[Dono 21b]   C. Donohue, Y. Khalifa, S. Perera, E. Sejdić, and J. L. Coyle. "A preliminary investigation of whether HRCA signals can differentiate between swallows from healthy people and swallows from people with neurodegenerative diseases". *Dysphagia*, Vol. 36, No. 4, pp. 635–643, 2021.

[Dono 21c] C. Donohue, S. Mao, E. Sejdić, and J. L. Coyle. "Tracking hyoid bone displacement during swallowing without videofluoroscopy using machine learning of vibratory signals". *Dysphagia*, Vol. 36, pp. 259–269, 2021.

[Dono 22a] 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*, pp. 1–12, 2022.

[Dono 22b] 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, No. 5, pp. 1103–1111, 2022.

[Dos 21] K. W. Dos Santos, E. da Cunha Rodrigues, R. S. Rech, E. M. da Ros Wendland, M. Neves, F. N. Hugo, and J. B. Hilgert. "Using Voice Change as an Indicator of Dysphagia: A Systematic Review". *Dysphagia*, pp. 1–13, 2021.

[Duan 16] F. Duan, L. Dai, W. Chang, Z. Chen, C. Zhu, and W. Li. "sEMG-based identification of hand motion commands using wavelet neural network combined with discrete wavelet transform". *IEEE Transactions on Industrial Electronics*, Vol. 63, No. 3, pp. 1923–1934, 2016.

[Dudi 15a] J. M. Dudik, J. L. Coyle, and E. Sejdić. "Dysphagia screening: contributions of cervical auscultation signals and modern signal-processing techniques". *IEEE transactions on human-machine systems*, Vol. 45, No. 4, pp. 465–477, 2015.

[Dudi 15b] J. M. Dudik, I. Jestrović, B. Luan, J. L. Coyle, and E. Sejdić. "Characteristics of dry chin-tuck swallowing vibrations and sounds". *IEEE Transactions on Biomedical Engineering*, Vol. 62, No. 10, pp. 2456–2464, 2015.

[Dudi 15c] J. M. Dudik, I. Jestrović, B. Luan, J. L. Coyle, and E. Sejdić. "A comparative analysis of swallowing accelerometry and sounds during saliva swallows". *Biomedical engineering online*, Vol. 14, No. 1, p. 3, 2015.

[Dudi 15d] J. M. Dudik, A. Kurosu, J. L. Coyle, and E. Sejdić. "A comparative analysis of DBSCAN, K-means, and quadratic variation algorithms for automatic identification of swallows from swallowing accelerometry signals". *Computers in biology and medicine*, Vol. 59, pp. 10–18, 2015.

[Dudi 16] J. M. Dudik, A. Kurosu, J. L. Coyle, and E. Sejdić. "A statistical analysis of cervical auscultation signals from adults with unsafe airway protection". *Journal of neuroengineering and rehabilitation*, Vol. 13, No. 1, p. 7, 2016.

[Dudi 18a] J. M. Dudik, J. L. Coyle, A. El-Jaroudi, Z.-H. Mao, M. Sun, and E. Sejdić. "Deep learning for classification of normal swallows in adults". *Neurocomputing*, Vol. 285, pp. 1–9, 2018.

[Dudi 18b] J. M. Dudik, A. Kurosu, J. L. Coyle, and E. Sejdić. "Dysphagia and its effects on swallowing sounds and vibrations in adults". *Biomedical engineering online*, Vol. 17, No. 1, p. 69, 2018.

[Earl 19] 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, 2019.

[Ekbe 02]   O. Ekberg, S. Hamdy, V. Woisard, A. Wuttge-Hannig, and P. Ortega. "Social and psychological burden of dysphagia: its impact on diagnosis and treatment". *Dysphagia*, Vol. 17, No. 2, pp. 139–146, 2002.

[Ende 08]   P. M. Enderby and R. Palmer. *FDA-2: Frenchay dysarthria assessment: examiner's manual*. Pro-ed, 2008.

[Enge 12]   L. van den Engel-Hoek, I. J. de Groot, E. Esser, B. Gorissen, J. C. Hendriks, B. J. de Swart, and A. C. Geurts. "Biomechanical events of swallowing are determined more by bolus consistency than by age or gender". *Physiology & behavior*, Vol. 106, No. 2, pp. 285–290, 2012.

[Enge 13]   L. van den Engel-Hoek, C. E. Erasmus, J. C. Hendriks, A. C. Geurts, W. M. Klein, S. Pillen, L. T. Sie, B. J. de Swart, and I. J. de Groot. "Oral muscles are progressively affected in Duchenne muscular dystrophy: implications for dysphagia treatment". *Journal of neurology*, Vol. 260, No. 5, pp. 1295–1303, 2013.

[Engl 01]   K. Englehart, B. Hudgin, and P. A. Parker. "A wavelet-based continuous classification scheme for multifunction myoelectric control". *IEEE Transactions on Biomedical Engineering*, Vol. 48, No. 3, pp. 302–311, 2001.

[Erco 13]   B. Ercolin, F. C. Sassi, L. D. Mangilli, L. I. Z. Mendonça, S. C. O. Limongi, and C. R. F. de Andrade. "Oral motor movements and swallowing in patients with myotonic dystrophy type 1". *Dysphagia*, Vol. 28, No. 3, pp. 446–454, 2013.

[Erma 09]   A. B. Erman, A. E. Kejner, N. D. Hogikyan, and E. L. Feldman. "Disorders of cranial nerves IX and X". In: *Seminars in neurology*, p. 85, NIH Public Access, 2009.

[Erte 03]   C. Ertekin and I. Aydogdu. "Neurophysiology of swallowing". *Clinical Neurophysiology*, Vol. 114, No. 12, pp. 2226–2244, 2003.

[Erte 14]   C. Ertekin. "Electrophysiological evaluation of oropharyngeal Dysphagia in Parkinson's disease". *Journal of movement disorders*, Vol. 7, No. 2, p. 31, 2014.

[Erte 98]   C. Ertekin, I. Aydogdu, N. Yüceyar, S. Tarlaci, N. Kiylioglu, M. Pehlivan, and G. Çelebi. "Electrodiagnostic methods for neurogenic dysphagia1". *Electroencephalography and Clinical Neurophysiology/Electromyography and Motor Control*, Vol. 109, No. 4, pp. 331–340, 1998.

[ETSI 00]   ETSI. "Digital cellular telecommunications system (Phase 2+) (GSM); Full rate speech; Transcoding (GSM 06.10 version 8.1.1 Release 1999)". Tech. Rep., European Telecommunications Standards Institute, 2000.

[Fals 09]   P. Falsetti, C. Acciai, R. Palilla, M. Bosi, F. Carpinteri, A. Zingarelli, C. Pedace, and L. Lenzi. "Oropharyngeal dysphagia after stroke: incidence, diagnosis, and clinical predictors in patients admitted to a neurorehabilitation unit". *Journal of Stroke and Cerebrovascular Diseases*, Vol. 18, No. 5, pp. 329–335, 2009.

[Farn 14]   D. Farneti, B. Fattori, A. Nacci, V. Mancini, M. Simonelli, G. Ruoppolo, and E. Genovese. "The Pooling-score (P-score): inter-and intra-rater reliability in endoscopic assessment of the severity of dysphagia". *ACTA otorhinolaryngologica italica*, Vol. 34, No. 2, p. 105, 2014.

[Farn 17]   D. Farneti. "Voice and dysphagia". In: *Dysphagia*, pp. 257–274, Springer, 2017.

[Farr 07]   A. Farri, A. Accornero, and C. Burdese. "Social importance of dysphagia: its impact on diagnosis and therapy". *Acta Otorhinolaryngologica Italica*, Vol. 27, No. 2, p. 83, 2007.

[Fest 16]   E. Festic, J. S. Soto, L. A. Pitre, M. Leveton, D. M. Ramsey, W. D. Freeman, M. G. Heckman, and A. S. Lee. "Novel bedside Phonetic evaluation to identify dysphagia and aspiration risk". *Chest*, Vol. 149, No. 3, pp. 649–659, 2016.

[Fine 03]   H. M. Finestone and L. S. Greene-Finestone. "Rehabilitation medicine: 2. Diagnosis of dysphagia and its nutritional management for stroke patients". *Cmaj*, Vol. 169, No. 10, pp. 1041–1044, 2003.

[Flow 13]   H. L. Flowers, F. L. Silver, J. Fang, E. Rochon, and R. Martino. "The incidence, co-occurrence, and predictors of dysphagia, dysarthria, and aphasia after first-ever acute ischemic stroke". *Journal of communication disorders*, Vol. 46, No. 3, pp. 238–248, 2013.

[Frai 09]   R. Fraile, N. Saenz-Lechon, J. Godino-Llorente, V. Osma-Ruiz, and C. Fredouille. "Automatic detection of laryngeal pathologies in records of sustained vowels by means of mel-frequency cepstral coefficient parameters and differentiation of patients by sex". *Folia phoniatrica et logopaedica*, Vol. 61, No. 3, pp. 146–152, 2009.

[Frit 12]   C. O. Fritz, P. E. Morris, and J. J. Richler. "Effect size estimates: current use, calculations, and interpretation.". *Journal of Experimental Psychology*, Vol. 141, No. 1, pp. 2–18, 2012.

[Gall 98]   L. Gallo, P. Guerra, and S. Palla. "Automatic on-line one-channel recognition of masseter activity". *Journal of dental research*, Vol. 77, No. 7, pp. 1539–1546, 1998.

[Gari 19]   M. R. Garita Sánchez, M. I. González Lutz, and N. Solís Pérez. "English vowel sounds: Pronunciation issues and student and faculty perceptions". *Actualidades Investigativas en Educación*, Vol. 19, No. 3, pp. 33–67, 2019.

[Gero 22]   A. Géron. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. " O'Reilly Media, Inc.", 2022.

[Gill 17]   S. Gillespie, Y.-Y. Logan, E. Moore, J. Laures-Gore, S. Russell, and R. Patel. "Cross-database models for the classification of dysarthria presence". In: *Annual Conference of the International Speech Communication Association (INTERSPEECH 2017)*, 2017.

[Gira 16]   L. F. Giraldo-Cadavid, A. M. Gutiérrez-Achury, K. Ruales-Suárez, M. L. Rengifo-Varona, C. Barros, A. Posada, C. Romero, and A. M. Galvis. "Validation of the Spanish Version of the Eating Assessment Tool-10 (EAT-10 spa) in Colombia. A Blinded Prospective Cohort Study". *Dysphagia*, Vol. 31, No. 3, pp. 398–406, 2016.

[Gola 14]   M. Golabbakhsh, A. Rajaei, M. Derakhshan, S. Sadri, M. Taheri, and P. Adibi. "Automated acoustic analysis in detection of spontaneous swallows in Parkinson's disease". *Dysphagia*, Vol. 29, No. 5, pp. 572–577, 2014.

[Good 16]   I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. MIT press, 2016.

[Gord 12]   G. Gordon and R. Tibshirani. "Karush-kuhn-tucker conditions". *Optimization*, Vol. 10, No. 725/36, p. 725, 2012.

[Grac 12]   W. Maria das Graças, L. R. Belo, D. Carneiro, A. G. Asano, P. J. A. Oliveira, D. M. da Silva, and O. G. Lins. "Swallowing in patients with Parkinson's disease: a surface electromyography study". *Dysphagia*, Vol. 27, No. 4, pp. 550–555, 2012.

[Gras 04]   P. Grassberger and I. Procaccia. "Measuring the strangeness of strange attractors". In: *The Theory of Chaotic Attractors*, pp. 170–189, Springer, 2004.

[Groh 06]   M. E. Groher, M. A. Crary, G. Carnaby, Z. Vickers, and C. Aguilar. "The impact of rheologically controlled materials on the identification of airway compromise on the clinical and videofluoroscopic swallowing examinations". *Dysphagia*, Vol. 21, No. 4, pp. 218–225, 2006.

[Grov 07]   K. J. Groves-Wright. *Acoustics and perception of wet vocal quality in identifying penetration/aspiration during swallowing*. PhD thesis, University of Cincinnati, 2007.

[Gupt 14]   H. Gupta and A. Banerjee. "Recovery of dysphagia in lateral medullary stroke". *Case reports in neurological medicine*, Vol. 2014, 2014.

[Gurg 13]   N. Gürgör, Ş. Arıcı, T. K. Incesu, Y. Seçil, F. Tokuçoğlu, and C. Ertekin. "An electrophysiological study of the sequential water swallowing". *Journal of Electromyography and Kinesiology*, Vol. 23, No. 3, pp. 619–626, 2013.

[Gust 91]   B. Gustafsson and L. Tibbling. "Dysphagia, an unrecognized handicap". *Dysphagia*, Vol. 6, No. 4, pp. 193–199, 1991.

[Guti 15]   A. Gutiérrez-Achury, K. Ruales Suárez, L. F. Giraldo Cadavid, and M. L. Rengifo Varona. "Escalas de calidad de vida y valoración de los síntomas en disfagia". *Revista Med*, Vol. 23, No. 1, pp. 50–55, 2015.

[Guyo 03]   I. Guyon and A. Elisseeff. "An introduction to variable and feature selection". *Journal of machine learning research*, Vol. 3, No. Mar, pp. 1157–1182, 2003.

[Guyo 08]   I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh. *Feature extraction: foundations and applications*. Vol. 207, Springer, 2008.

[Hack 90]   J. K. Hackett and M. Shah. "Multi-sensor fusion: a perspective". In: *Proceedings., IEEE International Conference on Robotics and Automation*, pp. 1324–1330, IEEE, 1990.

[Hadj 02]   S. Hadjitodorov and P. Mitev. "A computer system for acoustic analysis of pathological voices and laryngeal diseases screening". *Med. Eng. Phys.*, Vol. 24, No. 6, pp. 419–429, 2002.

[Haml 94]   S. Hamlet, D. G. Penney, and J. Formolo. "Stethoscope acoustics and cervical auscultation of swallowing". *Dysphagia*, Vol. 9, No. 1, pp. 63–68, 1994.

[Hamm 14]   K. Hammoudi, M. Boiron, N. Hernandez, C. Bobillier, and S. Morinière. "Acoustic study of pharyngeal swallowing as a function of the volume and consistency of the bolus". *Dysphagia*, Vol. 29, No. 4, pp. 468–474, 2014.

[Han 19]   Q. Han, C. Chen, R. Fu, L. Tan, and L. Xia. "Portable fibrobronchoscopic treatment for non-severe ischemic stroke-associated pneumonia patients with dysphagia: a pilot study". *Neurological research*, Vol. 41, No. 3, pp. 216–222, 2019.

[Hann 10]   F. Hanna, S. M. Molfenter, R. E. Cliffe, T. Chau, and C. M. Steele. "Anthropometric and demographic correlates of dual-axis swallowing accelerometry signal characteristics: a canonical correlation analysis". *Dysphagia*, Vol. 25, No. 2, pp. 94–103, 2010.

[Hart 00]   P. E. Hart, D. G. Stork, and R. O. Duda. *Pattern classification*. Wiley Hoboken, 2000.

[Hass 14]   H. E. Hassan and A. I. Aboloyoun. "The value of bedside tests in dysphagia evaluation". *Egyptian Journal of Ear, Nose, Throat and Allied Sciences*, Vol. 15, No. 3, pp. 197–203, 2014.

[Hast 09]   T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2, Springer, 2009.

[Hey 15]   C. Hey, P. Pluschinski, R. Pajunk, A. Almahameed, L. Girth, R. Sader, T. Stöver, and Y. Zaretsky. "Penetration–Aspiration: Is Their Detection in FEES® Reliable Without Video Recording?". *Dysphagia*, Vol. 30, No. 4, pp. 418–422, 2015.

[Hill 89]   A. D. Hillel, R. M. Miller, K. Yorkston, E. McDonald, F. H. Norris, and N. Konikow. "Amyotrophic lateral sclerosis severity scale". *Neuroepidemiology*, Vol. 8, No. 3, pp. 142–150, 1989.

[Hira 18]   I. Hirano and P. J. Kahrilas. "Dysphagia". In: J. L. Jameson, A. S. Fauci, D. L. Kasper, S. L. Hauser, D. L. Longo, and J. Loscalzo, Eds., *Harrison's Principles of Internal Medicine, 20e*, McGraw-Hill Education, New York, NY, 2018.

[Hsu 13]   C.-C. Hsu, W.-H. Chen, and H.-C. Chiu. "Using swallow sound and surface electromyography to determine the severity of dysphagia in patients with myasthenia gravis". *Biomedical Signal Processing and Control*, Vol. 8, No. 3, pp. 237–243, 2013.

[Huds 00]   D. L. Hudson and M. E. Cohen. *Neural networks and artificial intelligence for biomedical engineering*. Wiley Online Library, 2000.

[Hurs 57]   H. E. Hurst. "A suggested statistical model of some time series which occur in nature". *Nature*, Vol. 180, No. 4584, pp. 494–494, 1957.

[ICF]   "ICF Browser b5105 Swallowing". http://apps.who.int/classifications/icfbrowser/. Accessed: 2018-06-03.

[Ijit 17]   T. Ijitona, J. Soraghan, A. Lowit, G. Di-Caterina, and H. Yue. "Automatic detection of speech disorder in dysarthria using extended speech feature extraction and neural networks classification". In: *IET 3rd International Conference on Intelligent Signal Processing (ISP 2017)*, pp. 1–6, IET, 2017.

[Inou 18]   K. Inoue, M. Yoshioka, N. Yagi, S. Nagami, and Y. Oku. "Using Machine Learning and a Combination of Respiratory Flow, Laryngeal Motion, and Swallowing Sounds to Classify Safe and Unsafe Swallowing". *IEEE Transactions on Biomedical Engineering*, Vol. 9294, pp. 1–12, 2018.

[Ipin 18]   K. López-de Ipiña, P. Calvo, M. Faundez-Zanuy, P. Clavé, W. Nascimento, U. Martinez de Lizarduy, D. Alvarez, V. Arreola, O. Ortega, J. Mekyska, *et al.* "Automatic voice analysis for dysphagia detection". *Speech, Language and Hearing*, Vol. 21, No. 2, pp. 86–89, 2018.

[Jest 13] I. Jestrović, J. M. Dudik, B. Luan, J. L. Coyle, and E. Sejdić. "Baseline characteristics of cervical auscultation signals during various head maneuvers". *Computers in biology and medicine*, Vol. 43, No. 12, pp. 2014–2020, 2013.

[Jest 14] I. Jestrović, J. Coyle, and E. Sejdić. "The effects of increased fluid viscosity on stationary characteristics of EEG signal in healthy adults". *Brain research*, Vol. 1589, pp. 45–53, 2014.

[Jest 16] I. Jestrović, J. L. Coyle, S. Perera, and E. Sejdić. "Functional connectivity patterns of normal human swallowing: difference among various viscosity swallows in normal and chin-tuck head positions". *Brain research*, Vol. 1652, pp. 158–169, 2016.

[Jone 17a] O. Jones. "The Infrahyoid Muscles". http://teachmeanatomy.info/neck/muscles/infrahyoid/, 2017. [Online; accessed 18-September-2018].

[Jone 17b] O. Jones. "The Suprahyoid Muscles". http://teachmeanatomy.info/neck/muscles/suprahyoid-muscles/, 2017. [Online; accessed 18-September-2018].

[Kala 15] H. Kalantarian, N. Alshurafa, T. Le, and M. Sarrafzadeh. "Monitoring eating habits using a piezoelectric sensor-based necklace". *Computers in biology and medicine*, Vol. 58, pp. 46–55, 2015.

[Kang 17] Y. A. Kang, J. Kim, S. J. Jee, C. W. Jo, and B. S. Koo. "Detection of voice changes due to aspiration via acoustic voice analysis". *Auris Nasus Larynx*, 2017.

[Keag 17] M. J. Keage, M. B. Delatycki, I. Gupta, L. A. Corben, and A. P. Vogel. "Dysphagia in Friedreich Ataxia". *Dysphagia*, Vol. 32, No. 5, pp. 626–635, 2017.

[Khal 20a] Y. Khalifa, J. L. Coyle, and E. Sejdić. "Non-invasive identification of swallows via deep learning in high resolution cervical auscultation recordings". *Scientific Reports*, Vol. 10, No. 1, pp. 1–13, 2020.

[Khal 20b] 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 journal of biomedical and health informatics*, Vol. 25, No. 2, pp. 493–503, 2020.

[King 14] D. P. Kingma and J. Ba. "Adam: A method for stochastic optimization". *arXiv preprint arXiv:1412.6980*, 2014.

[Klah 99] M. S. Klahn and A. L. Perlman. "Temporal and durational patterns associating respiration and swallowing". *Dysphagia*, Vol. 14, No. 3, pp. 131–138, 1999.

[Ko 18] E. J. Ko, M. Chae, and S.-R. Cho. "Relationship between swallowing function and maximum phonation time in patients with Parkinsonism". *Annals of Rehabilitation Medicine*, Vol. 42, No. 3, p. 425, 2018.

[Kuma 14] S. Kumar, C. Doughty, G. Doros, M. Selim, S. Lahoti, S. Gokhale, and G. Schlaug. "Recovery of swallowing after dysphagic stroke: an analysis of prognostic factors". *Journal of Stroke and Cerebrovascular Diseases*, Vol. 23, No. 1, pp. 56–62, 2014.

[Kuro 19]   A. Kurosu, J. L. Coyle, J. M. Dudik, and E. Sejdic. "Detection of swallow kinematic events from acoustic high-resolution cervical auscultation signals in patients with stroke". *Archives of physical medicine and rehabilitation*, Vol. 100, No. 3, pp. 501–508, 2019.

[Laga 16]   M. L. Lagarde, D. M. Kamalski, and L. Van Den Engel-Hoek. "The reliability and validity of cervical auscultation in the diagnosis of dysphagia: a systematic review". *Clinical rehabilitation*, Vol. 30, No. 2, pp. 199–207, 2016.

[Lang 03]   S. E. Langmore. "Evaluation of oropharyngeal dysphagia: which diagnostic tool is superior?". *Current opinion in otolaryngology & head and neck surgery*, Vol. 11, No. 6, pp. 485–489, 2003.

[Lapa 17]   S. Lapa, S. Luger, W. Pfeilschifter, C. Henke, M. Wagner, and C. Foerch. "Predictors of dysphagia in acute pontine infarction". *Stroke*, Vol. 48, No. 5, pp. 1397–1399, 2017.

[Laza 04]   L. J. Lazareck and Z. M. Moussavi. "Classification of normal and dysphagic swallows by acoustical means". *IEEE Transactions on Biomedical Engineering*, Vol. 51, No. 12, pp. 2103–2112, 2004.

[LeCu 15]   Y. LeCun, Y. Bengio, and G. Hinton. "Deep learning". *nature*, Vol. 521, No. 7553, pp. 436–444, 2015.

[Lede 02]   S. B. Leder and J. F. Espinosa. "Aspiration risk after acute stroke: comparison of clinical examination and fiberoptic endoscopic evaluation of swallowing". *Dysphagia*, Vol. 17, No. 3, pp. 214–218, 2002.

[Lee 06]   J. Lee, S. Blain, M. Casas, D. Kenny, G. Berall, and T. Chau. "A radial basis classifier for the automatic detection of aspiration in children with dysphagia". *Journal of NeuroEngineering and Rehabilitation*, Vol. 3, No. 1, p. 14, 2006.

[Lee 08]   J. Lee, C. M. Steele, and T. Chau. "Time and time–frequency characterization of dual-axis swallowing accelerometry signals". *Physiological measurement*, Vol. 29, No. 9, p. 1105, 2008.

[Lee 09a]   J. Lee, C. M. Steele, and T. Chau. "Swallow segmentation with artificial neural networks and multi-sensor fusion". *Medical engineering & physics*, Vol. 31, No. 9, pp. 1049–1055, 2009.

[Lee 09b]   J. Lee. *Investigation of accelerometry, mechanomyography, and nasal airflow signals for abnormal swallow detection*. PhD thesis, University of Toronto, 2009.

[Lee 10]   J. Lee, E. Sejdić, C. M. Steele, and T. Chau. "Effects of liquid stimuli on dual-axis swallowing accelerometry signals in a healthy population". *BioMedical Engineering OnLine*, Vol. 9, No. 1, p. 7, 2010.

[Lee 11]   J. Lee, C. M. Steele, and T. Chau. "Classification of healthy and abnormal swallows based on accelerometry and nasal airflow signals". *Artificial Intelligence in Medicine*, Vol. 52, No. 1, pp. 17–25, 2011.

[Lesl 04]   P. Leslie, M. J. Drinnan, P. Finn, G. A. Ford, and J. A. Wilson. "Reliability and validity of cervical auscultation: a controlled comparison using videofluoroscopy". *Dysphagia*, Vol. 19, No. 4, pp. 231–240, 2004.

[Lesl 07]   P. Leslie, M. J. Drinnan, I. Zammit-Maempel, J. L. Coyle, G. A. Ford, and J. A. Wilson. "Cervical auscultation synchronized with images from endoscopy swallow evaluations". *Dysphagia*, Vol. 22, No. 4, pp. 290–298, 2007.

[Li 17]   Q. Li, Y. Minagi, T. Ono, Y. Chen, K. Hori, S. Fujiwara, and Y. Maeda. "The biomechanical coordination during oropharyngeal swallowing: an evaluation with a non-invasive sensing system". *Scientific Reports*, Vol. 7, No. 1, pp. 1–8, 2017.

[List 90]   M. A. List, C. Ritter-Sterr, and S. B. Lansky. "A performance status scale for head and neck cancer patients". *Cancer*, Vol. 66, No. 3, pp. 564–569, 1990.

[Liu 14]   J. Liu, X. Li, G. Li, and P. Zhou. "EMG feature assessment for myoelectric pattern recognition and channel selection: A study with incomplete spinal cord injury". *Medical engineering & physics*, Vol. 36, No. 7, pp. 975–980, 2014.

[Lope 14]   R. Lopez-Liria, M. Fernandez-Alonso, F. A. Vega-Ramirez, M. Salido-Campos, and D. Padilla-Góngora. "Treatment and rehabilitation of dysphagia following cerebrovascular disease". *Revista de neurologia*, Vol. 58, No. 6, pp. 259–267, 2014.

[Lope 17]   L. M. López-Soto, O. P. López-Soto, A. Osorio-Forero, F. Restrepo, and L. Tamayo-Orrego. "Muscle Activity and Muscle Strength in Atypical Swallowing". *Revista Salud Uninorte*, Vol. 33, No. 3, pp. 273–284, 2017.

[Mall 89]   S. Mallat. "A theory for multiresolution signal decomposition: the wavelet representation". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 7, pp. 674–693, 1989.

[Mand 10]   J. N. Mandrekar. "Receiver operating characteristic curve in diagnostic test assessment". *Journal of Thoracic Oncology*, Vol. 5, No. 9, pp. 1315–1316, 2010.

[Mao 19]   S. Mao, Z. Zhang, Y. Khalifa, C. Donohue, J. L. Coyle, and E. Sejdic. "Neck sensor-supported hyoid bone movement tracking during swallowing". *Royal Society open science*, Vol. 6, No. 7, p. 181982, 2019.

[Mao 21]   S. Mao, A. Sabry, Y. Khalifa, J. L. Coyle, and E. Sejdic. "Estimation of laryngeal closure duration during swallowing without invasive X-rays". *Future Generation Computer Systems*, Vol. 115, pp. 610–618, 2021.

[Mari 17]   T. Marian, J. Schröder, P. Muhle, I. Claus, S. Oelenberg, C. Hamacher, T. Warnecke, S. Suntrup-Krüger, and R. Dziewas. "Measurement of oxygen desaturation is not useful for the detection of aspiration in dysphagic stroke patients". *Cerebrovascular diseases extra*, Vol. 7, No. 1, pp. 44–50, 2017.

[Mart 08]   B. Martin-Harris and B. Jones. "The videofluorographic swallowing study". *Physical medicine and rehabilitation clinics of North America*, Vol. 19, No. 4, pp. 769–785, 2008.

[Maru 10]   M. Marusteri and V. Bacarea. "Comparing groups for statistical differences: how to choose the right statistical test?". *Biochemia medica*, Vol. 20, No. 1, pp. 15–32, 2010.

[Mats 08a]   K. Matsuo, K. M. Hiiemae, M. Gonzalez-Fernandez, and J. B. Palmer. "Respiration during feeding on solid food: alterations in breathing during mastication, pharyngeal bolus aggregation, and swallowing". *Journal of Applied Physiology*, Vol. 104, No. 3, pp. 674–681, 2008.

[Mats 08b] K. Matsuo and J. B. Palmer. "Anatomy and physiology of feeding and swallowing: normal and abnormal". *Physical medicine and rehabilitation clinics of North America*, Vol. 19, No. 4, pp. 691–707, 2008.

[Mats 09] K. Matsuo and J. B. Palmer. "Coordination of mastication, swallowing and breathing". *Japanese Dental Science Review*, Vol. 45, No. 1, pp. 31–40, 2009.

[McCu 01a] G. H. McCullough, R. T. Wertz, J. C. Rosenbek, R. H. Mills, W. G. Webb, and K. B. Ross. "Inter-and intrajudge reliability for videofluoroscopic swallowing evaluation measures". *Dysphagia*, Vol. 16, No. 2, pp. 110–118, 2001.

[McCu 01b] G. McCullough, R. Wertz, and J. Rosenbek. "Sensitivity and specificity of clinical/bedside examination signs for detecting aspiration in adults subsequent to stroke". *Journal of communication disorders*, Vol. 34, No. 1-2, pp. 55–72, 2001.

[McHo 00a] C. A. McHorney, D. E. Bricker, A. E. Kramer, J. C. Rosenbek, J. Robbins, K. A. Chignell, J. A. Logemann, and C. Clarke. "The SWAL-QOL outcomes tool for oropharyngeal dysphagia in adults: I. Conceptual foundation and item development". *Dysphagia*, Vol. 15, No. 3, pp. 115–121, 2000.

[McHo 00b] C. A. McHorney, D. E. Bricker, A. E. Kramer, J. C. Rosenbek, J. Robbins, K. A. Chignell, J. A. Logemann, and C. Clarke. "The SWAL-QOL outcomes tool for oropharyngeal dysphagia in adults: II. item reduction and preliminary scaling". *Dysphagia*, Vol. 15, No. 3, pp. 115–121, 2000.

[McKe 02] M. J. McKeown, D. C. Torpey, and W. C. Gehm. "Non-invasive monitoring of functionally distinct muscle activations during swallowing". *Clinical Neurophysiology*, Vol. 113, No. 3, pp. 354–366, 2002.

[McNu 21] J. McNulty, K. de Jager, H. T. Lancashire, J. Graveston, M. Birchall, and A. Vanhoestenberghe. "Prediction of larynx function using multichannel surface EMG classification". *IEEE Transactions on Medical Robotics and Bionics*, Vol. 3, No. 4, pp. 1032–1039, 2021.

[Meky 15] J. Mekyska, E. Janousova, P. Gomez-Vilda, Z. Smekal, I. Rektorova, I. Eliasova, M. Kostalova, M. Mrackova, J. B. Alonso-Hernandez, M. Faundez-Zanuy, *et al.* "Robust and complex approach of pathological speech signal analysis". *Neurocomputing*, Vol. 167, pp. 94–111, 2015.

[Merl 99] R. Merletti and P. Di Torino. "Standards for reporting EMG data". *J Electromyogr Kinesiol*, Vol. 9, No. 1, pp. 3–4, 1999.

[Mesa 14] I. Mesa, A. Rubio, I. Tubia, J. De No, and J. Diaz. "Channel and feature selection for a surface electromyographic pattern recognition task". *Expert Systems with Applications*, Vol. 41, No. 11, pp. 5190–5200, 2014.

[Miot 18] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley. "Deep learning for healthcare: review, opportunities and challenges". *Briefings in bioinformatics*, Vol. 19, No. 6, pp. 1236–1246, 2018.

[Mitc 97] T. M. Mitchell and T. M. Mitchell. *Machine learning*. Vol. 1, McGraw-hill New York, 1997.

[Miya 20] S. Miyagi, S. Sugiyama, K. Kozawa, S. Moritani, S.-i. Sakamoto, and O. Sakai. "Classifying dysphagic swallowing sounds with support vector machines". In: *Healthcare*, p. 103, MDPI, 2020.

[Moha 20]   M. A. Mohammed, K. H. Abdulkareem, S. A. Mostafa, M. K. A. Ghani, M. S. Maashi, B. Garcia-Zapirain, I. Oleagordia, H. Alhakami, and F. T. Al-Dhief. "Voice pathology detection and classification using convolutional neural network model". *Applied Sciences*, Vol. 10, No. 11, p. 3723, 2020.

[Moli 17]   M. Molina Arias. "¿ Qué significa realmente el valor de p?". *Pediatría Atención Primaria*, Vol. 19, No. 76, pp. 377–381, 2017.

[Mona 08]   A. Monaco, R. Cattaneo, A. Spadaro, and M. Giannoni. "Surface electromyography pattern of human swallowing". *BMC oral health*, Vol. 8, No. 1, p. 6, 2008.

[Mont 18]   D. Montaña, Y. Campos-Roca, and C. J. Pérez. "A Diadochokinesis-based expert system considering articulatory features of plosive consonants for early detection of Parkinson's disease". *Computer Methods and Programs in Biomedicine*, Vol. 154, pp. 89–97, 2018.

[Morg 10]   A. Morgan, S.-D. Mageandran, and C. Mei. "Incidence and clinical presentation of dysarthria and dysphagia in the acute setting following paediatric traumatic brain injury". *Child: care, health and development*, Vol. 36, No. 1, pp. 44–53, 2010.

[Mova 17a] F. Movahedi, A. Kurosu, J. L. Coyle, S. Perera, and E. Sejdić. "Anatomical directional dissimilarities in tri-axial swallowing accelerometry signals". *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 25, No. 5, pp. 447–458, 2017.

[Mova 17b] F. Movahedi, A. Kurosu, J. L. Coyle, S. Perera, and E. Sejdić. "A comparison between swallowing sounds and vibrations in patients with dysphagia". *Computer methods and programs in biomedicine*, Vol. 144, pp. 179–187, 2017.

[Mozz 17]   F. Mozzanica, L. Scarponi, S. Pedrali, N. Pizzorni, C. Pinotti, F. Foieni, G. Zuccotti, and A. Schindler. "Dysphagia screening in subacute care settings using the Italian version of the Royal Brisbane and Women's Hospital (I-RBWH) dysphagia screening tool". *Acta Otorhinolaryngologica Italica*, Vol. 37, No. 1, p. 25, 2017.

[Much 17]   J. Mucha, Z. Galaz, J. Mekyska, T. Kiska, V. Zvoncak, Z. Smekal, I. Eliasova, M. Mrackova, M. Kostalova, I. Rektorova, *et al.* "Identification of hypokinetic dysarthria using acoustic analysis of poem recitation". In: *Telecommunications and Signal Processing (TSP), 2017 40th International Conference on*, pp. 739–742, IEEE, 2017.

[Mukh 21]   R. I. Mukhamediev, A. Symagulov, Y. Kuchin, K. Yakunin, and M. Yelis. "From classical machine learning to deep neural networks: A simplified scientometric review". *Applied Sciences*, Vol. 11, No. 12, p. 5541, 2021.

[Muru 10]   S. Murugappan, S. Boyce, S. Khosla, L. Kelchner, and E. Gutmark. "Acoustic characteristics of phonation in "wet voice" conditions". *The Journal of the Acoustical Society of America*, Vol. 127, No. 4, pp. 2578–2589, 2010.

[Must 17]   F. Musto, R. Rosati, C. Sforza, M. Toma, and C. Dellavia. "Standardised surface electromyography allows effective submental muscles assessment". *Journal of Electromyography and Kinesiology*, Vol. 34, pp. 1–5, 2017.

[Nacc 08]   A. Nacci, F. Ursino, R. La Vela, F. Matteucci, V. Mallardi, and B. Fattori. "Fiberoptic endoscopic evaluation of swallowing (FEES): proposal for informed consent". *Acta Otorhinolaryngologica Italica*, Vol. 28, No. 4, p. 206, 2008.

[Nade 19]   E. Naderifar, A. Ghorbani, N. Moradi, and H. Ansari. "Use of formant cent-ralization ratio for vowel impairment detection in normal hearing and different degrees of hearing impairment". *Logopedics Phoniatrics Vocology*, Vol. 44, No. 4, pp. 159–165, 2019.

[Nazm 16]   N. Nazmi, M. A. Abdul Rahman, S.-I. Yamamoto, S. A. Ahmad, H. Zamzuri, and S. A. Mazlan. "A review of classification techniques of EMG signals during isotonic and isometric contractions". *Sensors*, Vol. 16, No. 8, p. 1304, 2016.

[Nikj 11]   M. S. Nikjoo, C. M. Steele, E. Sejdić, and T. Chau. "Automatic discrimination between safe and unsafe swallowing using a reputation-based classifier". *Biomedical engineering online*, Vol. 10, No. 1, p. 100, 2011.

[Nozu 17]   S. Nozue, Y. Ihara, K. Takahashi, Y. Harada, Y. Takei, K. Yuasa, and K. Yokoyama. "Accuracy of cervical auscultation in detecting the presence of material in the airway". *Clinical and experimental dental research*, Vol. 3, No. 6, pp. 209–214, 2017.

[Okub 12]   P. Okubo, S. Fábio, D. Domenis, and O. Takayanagui. "Using the National Institute of Health Stroke Scale to predict dysphagia in acute ischemic stroke". *Cerebrovascular diseases*, Vol. 33, No. 6, pp. 501–507, 2012.

[ONei 99]   K. H. O'Neil, M. Purdy, J. Falk, and L. Gallo. "The dysphagia outcome and severity scale". *Dysphagia*, Vol. 14, No. 3, pp. 139–145, 1999.

[Oroz 14]   J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. V. Bonilla, M. C. Gonzalez-Rátiva, and E. Nöth. "New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease.". In: *LREC*, pp. 342–347, 2014.

[Oroz 15]   J. R. Orozco-Arroyave, E. A. Belalcazar-Bolanos, J. D. Arias-Londoño, J. F. Vargas-Bonilla, S. Skodda, J. Rusz, K. Daqrouq, F. Hönig, and E. Nöth. "Characterization methods for the detection of multiple voice disorders: neurological, functional, and laryngeal diseases". *IEEE journal of biomedical and health informatics*, Vol. 19, No. 6, pp. 1820–1828, 2015.

[Oroz 16a]   J. Orozco-Arroyave, F. Hönig, J. Arias-Londoño, J. Vargas-Bonilla, K. Daqrouq, S. Skodda, J. Rusz, and E. Nöth. "Automatic detection of Parkinson's disease in running speech spoken in three different languages". *The Journal of the Acoustical Society of America*, Vol. 139, No. 1, pp. 481–500, 2016.

[Oroz 16b]   J. R. Orozco-Arroyave. *Analysis of speech of people with Parkinson's disease*. Vol. 41, Logos Verlag Berlin GmbH, 2016.

[Oroz 18]   J. R. Orozco-Arroyave, J. C. Vásquez-Correa, J. F. Vargas-Bonilla, R. Arora, N. Dehak, P. S. Nidadavolu, H. Christensen, F. Rudzicz, M. Yancheva, H. Chinaei, *et al.* "NeuroSpeech: An open-source software for Parkinson's speech analysis". *Digital Signal Processing*, Vol. 77, pp. 207–221, 2018.

[Osis 17]   F. Osisanwo, J. Akinsola, O. Awodele, J. Hinmikaiye, O. Olakanmi, and J. Akinjobi. "Supervised machine learning algorithms: classification and comparison". *International Journal of Computer Trends and Technology (IJCTT)*, Vol. 48, No. 3, pp. 128–138, 2017.

[Ozsu 20]   C. Özsürekci, S. S. Arslan, N. Demir, H. Çalışkan, G. Şengül Ayçiçek, H. E. Kılınç, Ö. F. Yaşaroğlu, C. Kızılarslanoğlu, R. Tuna Doğrul, C. Balcı, *et al.* "Timing of dysphagia screening in Alzheimer's dementia". *Journal of Parenteral and Enteral Nutrition*, Vol. 44, No. 3, pp. 516–524, 2020.

[Palm 00]   J. Palmer, J. Drennan, and M. Baba. "Evaluation and treatment of swallowing impairments". *Am Fam Physician*, Vol. 61, No. 8, pp. 2453–2462, 2000.

[Palm 89]   J. B. Palmer. "Electromyography of the muscles of oropharyngeal swallowing: basic concepts". *Dysphagia*, Vol. 3, No. 4, pp. 192–198, 1989.

[Palm 99]   P. M. Palmer, E. S. Luschei, D. Jaffe, and T. M. McCulloch. "Contributions of individual muscles to the submental surface electromyogram during swallowing". *Journal of Speech, Language, and Hearing Research*, Vol. 42, No. 6, pp. 1378–1391, 1999.

[Papa 16]   V. Papapanagiotou, C. Diou, L. Zhou, J. van den Boer, M. Mars, and A. Delopoulos. "A novel chewing detection system based on ppg, audio, and accelerometry". *IEEE journal of biomedical and health informatics*, Vol. 21, No. 3, pp. 607–618, 2016.

[Park 22]   H.-Y. Park, D. Park, H. S. Kang, H. Kim, S. Lee, and S. Im. "Post-stroke respiratory complications using machine learning with voice features from mobile devices". *Scientific Reports*, Vol. 12, No. 1, pp. 1–11, 2022.

[Pate 17]   D. Patel, S. Krishnaswami, E. Steger, E. Conover, M. Vaezi, M. Ciucci, and D. Francis. "Economic and survival burden of dysphagia among inpatients in the United States". *Diseases of the Esophagus*, 2017.

[Pere 22]   P. A. Pérez-Toro, D. Rodríguez-Salas, T. Arias-Vergara, P. Klumpp, M. Schuster, E. Nöth, J. R. Orozco-Arroyave, and A. K. Maier. "Interpreting acoustic features for the assessment of Alzheimer's disease using ForestNet". *Smart Health*, p. 100347, 2022.

[Perl 99]   A. Perlman, P. Palmer, T. McCulloch, and D. Vandaele. "Electromyographic activity from human laryngeal, pharyngeal, and submental muscles during swallowing". *Journal of applied physiology*, Vol. 86, No. 5, pp. 1663–1669, 1999.

[Pfei 04]   R. F. Pfeiffer. "Neurogenic dysphagia". *Neurology in Clinical Practice: Principles of diagnosis and management*, Vol. 1, p. 165, 2004.

[Pfei 16]   R. F. Pfeiffer. "Neurogenic Dysphagia". In: R. B. Daroff, J. Jankovic, J. C. Mazziotta, and S. L. Pomeroy, Eds., *Bradley's Neurology in Clinical Practice*, pp. 148–157.e2, Elsevier Health Sciences, 2016.

[Phin 09]   A. Phinyomark, C. Limsakul, and P. Phukpattaranont. "A comparative study of wavelet denoising for multifunction myoelectric control". In: *2009 international conference on computer and automation engineering*, pp. 21–25, IEEE, 2009.

[Phin 12a] A. Phinyomark, P. Phukpattaranont, and C. Limsakul. "Feature reduction and selection for EMG signal classification". *Expert Systems with Applications*, Vol. 39, No. 8, pp. 7420–7431, 2012.

[Phin 12b] A. Phinyomark, P. Phukpattaranont, and C. Limsakul. "Fractal analysis features for weak and single-channel upper-limb EMG signals". *Expert Systems with Applications*, Vol. 39, No. 12, pp. 11156–11163, 2012.

[Phin 13]   A. Phinyomark, F. Quaine, S. Charbonnier, C. Serviere, F. Tarpin-Bernard, and Y. Laurillau. "EMG feature evaluation for improving myoelectric pattern recognition robustness". *Expert Systems with Applications*, Vol. 40, No. 12, pp. 4832–4840, 2013.

[Pir 15]   D. Pir and T. Brown. "Acoustic Group Feature Selection Using Wrapper Method for Automatic Eating Condition Recognition". In: *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[Poor 14]   M. Poorjavad, S. Talebian Moghadam, N. Nakhostin Ansari, and M. Daemi. "Surface electrical stimulation for treating swallowing disorders after stroke: a review of the stimulation intensity levels and the electrode placements". *Stroke research and treatment*, Vol. 2014, 2014.

[Poor 17]   M. Poorjavad, S. Talebian, N. N. Ansari, and Z. Soleymani. "Surface electromyographic assessment of swallowing function". *Iranian journal of medical sciences*, Vol. 42, No. 2, p. 194, 2017.

[Powe 11]   D. Powers. "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation". *J. Mach. Learn. Technol.*, Vol. 2, No. 1, pp. 37–63, 2011.

[Rami 90]   L. O. Ramig, R. C. Scherer, E. R. Klasner, I. R. Titze, and Y. Horii. "Acoustic analysis of voice in amyotrophic lateral sclerosis: A longitudinal case study". *Journal of Speech and Hearing Disorders*, Vol. 55, No. 1, pp. 2–14, 1990.

[Ramo 20]   V. M. Ramos, H. A. K. Hernandez-Diaz, M. E. H.-D. Huici, H. Martens, G. Van Nuffelen, and M. De Bodt. "Acoustic features to characterize sentence accent production in dysarthric speech". *Biomedical Signal Processing and Control*, Vol. 57, pp. 1–8, 2020.

[Rebr 18]   C. Rebrion, Z. Zhang, Y. Khalifa, M. Ramadan, A. Kurosu, J. L. Coyle, S. Perera, and E. Sejdić. "High-resolution cervical auscultation signal features reflect vertical and horizontal displacements of the hyoid bone during swallowing". *IEEE journal of translational engineering in health and medicine*, Vol. 7, pp. 1–9, 2018.

[Rest 17]   S. Restrepo-Agudelo, S. Roldan-Vasco, L. Ramirez-Arbelaez, S. Cadavid-Arboleda, E. Perez-Giraldo, and A. Orozco-Duque. "Improving surface EMG burst detection in infrahyoid muscles during swallowing using digital filters and discrete wavelet analysis". *Journal of Electromyography and Kinesiology*, Vol. 35, pp. 1–8, 2017.

[Rich 00]   J. S. Richman and J. R. Moorman. "Physiological time-series analysis using approximate entropy and sample entropy". *American Journal of Physiology-Heart and Circulatory Physiology*, Vol. 278, No. 6, pp. H2039–H2049, 2000.

[Rieb 19]   B. Riebold, H. Nahrstaedt, T. Schauer, and R. O. Seidl. "Self-adapting Classification System for Swallow Intention Detection in Dysphagia Therapy". *Current Directions in Biomedical Engineering*, Vol. 5, No. 1, pp. 49–52, 2019.

[Rode 13]   D. F. Roden and K. W. Altman. "Causes of dysphagia among different age groups: a systematic review of the literature". *Otolaryngologic Clinics of North America*, Vol. 46, No. 6, pp. 965–987, 2013.

[Rold 18]   S. Roldan-Vasco, S. Restrepo-Agudelo, Y. Valencia-Martinez, and A. Orozco-Duque. "Automatic detection of oral and pharyngeal phases in swallowing using classification algorithms and multichannel EMG". *Journal of Electromyography and Kinesiology*, Vol. 43, pp. 193–200, 2018.

[Rold 20]   S. Roldan-Vasco, E. Perez-Giraldo, and A. Orozco-Duque. "Scalogram-energy based segmentation of surface electromyography signals from swallowing related muscles". *Computer methods and programs in biomedicine*, Vol. 194, p. 105480, 2020.

[Rose 93]   M. T. Rosenstein, J. J. Collins, and C. J. De Luca. "A practical method for calculating largest Lyapunov exponents from small data sets". *Physica D: Nonlinear Phenomena*, Vol. 65, No. 1-2, pp. 117–134, 1993.

[Rose 96]   J. C. Rosenbek, J. A. Robbins, E. B. Roecker, J. L. Coyle, and J. L. Wood. "A penetration-aspiration scale". *Dysphagia*, Vol. 11, No. 2, pp. 93–98, 1996.

[Rusz 11]   J. Rusz, R. Cmejla, H. Ruzickova, and E. Ruzicka. "Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease". *The journal of the Acoustical Society of America*, Vol. 129, No. 1, pp. 350–367, 2011.

[Rusz 13]   J. Rusz, R. Cmejla, T. Tykalova, H. Ruzickova, J. Klempir, V. Majerova, J. Picmausova, J. Roth, and E. Ruzicka. "Imprecise vowel articulation as a potential early marker of Parkinson's disease: Effect of speaking task". *The Journal of the Acoustical Society of America*, Vol. 134, No. 3, pp. 2171–2181, 2013.

[Ryu 04]   J. S. Ryu, S. R. Park, and K. H. Choi. "Prediction of laryngeal aspiration using voice analysis". *American journal of physical medicine & rehabilitation*, Vol. 83, No. 10, pp. 753–757, 2004.

[Samp 14]   M. Sampaio, N. Argolo, A. Melo, and A. C. Nóbrega. "Wet voice as a sign of penetration/aspiration in Parkinson's disease: does testing material matter?". *Dysphagia*, Vol. 29, No. 5, pp. 610–615, 2014.

[Sanc 18]   Y. Sánchez-Cardona, A. Orozco-Duque, and S. Roldán-Vasco. "Caracterización y Clasificación de Señales de Auscultación Cervical Adquiridas con Estetoscopio para Detección Automática de Sonidos Deglutorios". *Revista mexicana de ingeniería biomédica*, Vol. 39, No. 2, pp. 205–216, 2018.

[Sant 15]   K. W. d. Santos, B. Scheeren, A. C. Maciel, and M. Cassol. "Vocal variability post swallowing in individuals with and without oropharyngeal dysphagia". *International archives of otorhinolaryngology*, Vol. 19, No. 1, pp. 61–66, 2015.

[Sapi 10]   S. Sapir, L. O. Ramig, J. L. Spielman, and C. Fox. "Formant centralization ratio: A proposal for a new acoustic measure of dysarthric speech". *Journal of Speech, Language, and Hearing Research*, 2010.

[Sase 17]   A. Sasegbon and S. Hamdy. "The anatomy and physiology of normal and abnormal swallowing in oropharyngeal dysphagia". *Neurogastroenterology & Motility*, Vol. 29, No. 11, p. e13100, 2017.

[Sazo 10]   E. S. Sazonov, O. Makeyev, S. Schuckers, P. Lopez-Meyer, E. L. Melanson, and M. R. Neuman. "Automatic detection of swallowing events by acoustical means for applications of monitoring of ingestive behavior". *IEEE Transactions on Biomedical Engineering*, Vol. 57, No. 3, pp. 626–633, 2010.

[Scha 15]   M. Schartner, A. Seth, Q. Noirhomme, M. Boly, M.-A. Bruno, S. Laureys, and A. Barrett. "Complexity of multi-dimensional spontaneous EEG decreases during propofol induced general anaesthesia". *PloS one*, Vol. 10, No. 8, p. e0133532, 2015.

[Schu 14]   C. Schultheiss, T. Schauer, H. Nahrstaedt, and R. O. Seidl. "Automated detection and evaluation of swallowing using a combined EMG/bioimpedance measurement system". *The Scientific World Journal*, Vol. 2014, 2014.

[Sejd 09]   E. Sejdic, C. M. Steele, and T. Chau. "Segmentation of dual-axis swallowing accelerometry signals in healthy subjects with analysis of anthropometric effects on duration of swallowing activities". *IEEE Transactions on Biomedical Engineering*, Vol. 56, No. 4, pp. 1090–1097, 2009.

[Sejd 10]   E. Sejdić, T. H. Falk, C. M. Steele, and T. Chau. "Vocalization removal for improved automatic segmentation of dual-axis swallowing accelerometry signals". *Medical engineering & physics*, Vol. 32, No. 6, pp. 668–672, 2010.

[Sejd 12]   E. Sejdić, C. M. Steele, and T. Chau. "A method for removal of low frequency components associated with head movements from dual-axis swallowing accelerometry signals". *PLoS One*, Vol. 7, No. 3, p. e33464, 2012.

[Sejd 13]   E. Sejdić, C. M. Steele, and T. Chau. "Classification of Penetration–Aspiration Versus Healthy Swallows Using Dual-Axis Swallowing Accelerometry Signals in Dysphagic Subjects". *IEEE Transactions on Biomedical Engineering*, Vol. 60, No. 7, pp. 1859–1866, 2013.

[Sejd 18]   E. Sejdic, G. A. Malandraki, and J. L. Coyle. "Computational deglutition: Using signal-and image-processing methods to understand swallowing and associated disorders [life sciences]". *IEEE signal processing magazine*, Vol. 36, No. 1, pp. 138–146, 2018.

[Sfor 11]   C. Sforza, R. Rosati, M. De Menezes, F. Musto, and M. Toma. "EMG analysis of trapezius and masticatory muscles: experimental protocol and data reproducibility". *Journal of oral rehabilitation*, Vol. 38, No. 9, pp. 648–654, 2011.

[Shaw 13]   S. M. Shaw and R. Martino. "The normal swallow: muscular and neurophysiological control". *Otolaryngologic Clinics of North America*, Vol. 46, No. 6, pp. 937–956, 2013.

[Sher 99]   B. Sherman, J. M. Nisenboum, B. L. Jesberger, C. A. Morrow, and J. A. Jesberger. "Assessment of dysphagia with the use of pulse oximetry". *Dysphagia*, Vol. 14, No. 3, pp. 152–156, 1999.

[Shir 14]   S. S. Shirazi, A. H. Birjandi, and Z. Moussavi. "Noninvasive and automatic diagnosis of patients at high risk of swallowing aspiration". *Medical & biological engineering & computing*, pp. 1–7, 2014.

[Sifr 14]   D. Sifrim, N. Vilardell, and P. Clavé. "Oropharyngeal dysphagia and swallowing dysfunction". In: *Functional and GI Motility Disorders*, pp. 1–13, Karger Publishers, 2014.

[Smit 90]   D. Smith, S. Hamlet, and L. Jones. "Acoustic technique for determining timing of velopharyngeal closure in swallowing". *Dysphagia*, Vol. 5, No. 3, pp. 142–146, 1990.

[Soln 10]   S. Solnik, P. Rider, K. Steinweg, P. DeVita, and T. Hortobágyi. "Teager–Kaiser energy operator signal conditioning improves EMG onset detection". *European journal of applied physiology*, Vol. 110, No. 3, pp. 489–498, 2010.

[Spad 09]   A. A. Spadotto, A. R. Gatto, R. C. Guido, A. N. Montagnoli, P. C. Cola, J. C. Pereira, and A. O. Schelp. "Classification of normal swallowing and oropharyngeal dysphagia using wavelet". *Applied Mathematics and Computation*, Vol. 207, No. 1, pp. 75–82, 2009.

[Spey 13]   R. Speyer. "Oropharyngeal dysphagia: screening and assessment". *Otolaryngologic Clinics of North America*, Vol. 46, No. 6, pp. 989–1008, 2013.

[Stee 13]   C. M. Steele, E. Sejdić, and T. Chau. "Noninvasive detection of thin-liquid aspiration using dual-axis swallowing accelerometry". *Dysphagia*, Vol. 28, No. 1, pp. 105–112, 2013.

[Stee 19]   C. M. Steele, R. Mukherjee, J. M. Kortelainen, H. Pölönen, M. Jedwab, S. L. Brady, K. B. Theimer, S. Langmore, L. F. Riquelme, N. B. Swigert, *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*, Vol. 34, No. 5, pp. 698–707, 2019.

[Stei 81]   C. M. Stein. "Estimation of the mean of a multivariate normal distribution". *The annals of Statistics*, pp. 1135–1151, 1981.

[Step 12]   C. E. Stepp. "Surface electromyography for speech and swallowing systems: measurement, analysis, and interpretation". *Journal of Speech, Language, and Hearing Research*, Vol. 55, No. 4, pp. 1232–1246, 2012.

[Suar 18]   J. C. Suárez-Escudero, Z. V. Rueda Vallejo, and A. F. Orozco. "Disfagia y neurología:¿ una unión indefectible". *Acta Neurol Colomb*, Vol. 34, No. 1, pp. 92–100, 2018.

[Suar 22]   L. V. Suárez-Patiño, A. Orozco-Duque, E. Pérez-Giraldo, S. Roldán-Vasco, J. C. Suárez-Escudero, and L. Martínez-Moreno. "Sincronización entre la videodeglución y la electromiografía de superficie en pacientes con afectación neurológica y síntomas de disfagia". *Biomédica*, Vol. 42, No. 4, pp. 650–664, 2022.

[Suba 12]   A. Subasi. "Medical decision support system for diagnosis of neuromuscular disorders using DWT and fuzzy support vector machines". *Computers in Biology and Medicine*, Vol. 42, No. 8, pp. 806–815, 2012.

[Suzu 20]   M. Suzuki, M. Sasaki, K. Kamata, A. Nakayama, I. Shibamoto, and Y. Tamada. "Swallowing pattern classification method using multichannel surface EMG signals of suprahyoid and infrahyoid muscles". *Advanced Biomedical Engineering*, Vol. 9, pp. 10–20, 2020.

[Taki 16]   C. Takizawa, E. Gemmell, J. Kenworthy, and R. Speyer. "A systematic review of the prevalence of oropharyngeal dysphagia in stroke, Parkinson's disease, Alzheimer's disease, head injury, and pneumonia". *Dysphagia*, Vol. 31, No. 3, pp. 434–441, 2016.

[Tomc 14]   M. Tomczak and E. Tomczak. "The need to report effect size estimates revisited. An overview of some recommended measures of effect size". *Trends in Sport Sciences*, Vol. 1, No. 21, pp. 19–25, 2014.

[Trav 17]   C. M. Travieso, J. B. Alonso, J. R. Orozco-Arroyave, J. Vargas-Bonilla, E. Nöth, and A. G. Ravelo-García. "Detection of different voice diseases based on the nonlinear characterization of speech signals". *Expert Systems with Applications*, Vol. 82, pp. 184–195, 2017.

[Vaba 19]   A. Vabalas, E. Gowen, E. Poliakoff, and A. J. Casson. "Machine learning algorithm validation with a limited sample size". *PloS one*, Vol. 14, No. 11, p. e0224365, 2019.

[Vaim 04a] M. Vaiman, E. Eviatar, and S. Segal. "Evaluation of normal deglutition with the help of rectified surface electromyography records". *Dysphagia*, Vol. 19, No. 2, pp. 125–132, 2004.

[Vaim 04b] M. Vaiman, E. Eviatar, and S. Segal. "Surface electromyographic studies of swallowing in normal subjects: a review of 440 adults. Report 1. Quantitative data: timing measures". *Otolaryngology-Head and Neck Surgery*, Vol. 131, No. 4, pp. 548–555, 2004.

[Vaim 04c] M. Vaiman, E. Eviatar, and S. Segal. "Surface electromyographic studies of swallowing in normal subjects: a review of 440 adults. Report 2. Quantitative data: amplitude measures". *Otolaryngology-Head and neck Surgery*, Vol. 131, No. 5, pp. 773–780, 2004.

[Vaim 04d] M. Vaiman, E. Eviatar, and S. Segal. "Surface electromyographic studies of swallowing in normal subjects: a review of 440 adults. Report 3. Qualitative data". *Otolaryngology—Head and Neck Surgery*, Vol. 131, No. 6, pp. 977–985, 2004.

[Vaim 06]   M. Vaiman. "Surface electromyography in preoperative evaluation and postoperative monitoring of Zenker's diverticulum". *Dysphagia*, Vol. 21, No. 1, pp. 14–20, 2006.

[Vaim 07]   M. Vaiman. "Standardization of surface electromyography utilized to evaluate patients with dysphagia". *Head & face medicine*, Vol. 3, No. 1, p. 26, 2007.

[Vaim 08]   M. Vaiman, G. Shoval, and H. Gavriel. "The electrodiagnostic examination of psychogenic swallowing disorders". *European Archives of Oto-Rhino-Laryngology*, Vol. 265, No. 6, pp. 663–668, 2008.

[Vaim 09]   M. Vaiman and E. Eviatar. "Surface electromyography as a screening method for evaluation of dysphagia and odynophagia". *Head & face medicine*, Vol. 5, No. 1, p. 9, 2009.

[Vasq 18]   J. Vásquez-Correa, J. Orozco-Arroyave, T. Bocklet, and E. Nöth. "Towards an automatic evaluation of the dysarthria level of patients with Parkinson's disease". *Journal of communication disorders*, 2018.

[Vasq 20]   J. C. Vasquez-Correa, T. Arias-Vergara, M. Schuster, J. R. Orozco-Arroyave, and E. Nöth. "Parallel representation learning for the classification of pathological speech: studies on Parkinson's disease and cleft lip and palate". *Speech Communication*, Vol. 122, pp. 56–67, 2020.

[Vasq 21]   J. C. Vásquez-Correa, C. D. Rios-Urrego, T. Arias-Vergara, M. Schuster, J. Rusz, E. Nöth, and J. R. Orozco-Arroyave. "Transfer learning helps to improve the accuracy to classify patients with different speech disorders in different languages". *Pattern Recognition Letters*, Vol. 150, pp. 272–279, 2021.

[Voge 17]   A. P. Vogel, N. Rommel, A. Oettinger, M. Horger, P. Krumm, E.-M. Kraus, L. Schöls, and M. Synofzik. "Speech and swallowing abnormalities in adults with POLG associated ataxia (POLG-A)". *Mitochondrion*, Vol. 37, pp. 1–7, 2017.

[Vyas 16]   G. Vyas, M. K. Dutta, J. Prinosil, and P. Harár. "An automatic diagnosis and assessment of dysarthric speech using speech disorder specific prosodic features". In: *Telecommunications and Signal Processing (TSP), 2016 39th International Conference on*, pp. 515–518, IEEE, 2016.

[Wain 21]   J. Wainer and G. Cawley. "Nested cross-validation when selecting classifiers is overzealous for most practical applications". *Expert Systems with Applications*, Vol. 182, p. 115222, 2021.

[Wait 11]   A. Waito, G. L. Bailey, S. M. Molfenter, D. C. Zoratto, and C. M. Steele. "Voice-quality abnormalities as a sign of dysphagia: validation against acoustic and videofluoroscopic data". *Dysphagia*, Vol. 26, No. 2, pp. 125–134, 2011.

[Walt 18]   J. Walton and P. Silva. "Physiology of swallowing". *Surgery (Oxford)*, 2018.

[Wang 15]   C. M. Wang, J. Y. Chen, C. C. Chuang, W. C. Tseng, A. M. Wong, and Y. C. Pei. "Aging-related changes in swallowing, and in the coordination of swallowing and respiration determined by novel non-invasive measurement techniques". *Geriatrics & gerontology international*, Vol. 15, No. 6, pp. 736–744, 2015.

[Wang 17]   Y.-C. Wang, W. Chou, B.-S. Lin, J.-J. Wang, and B.-S. Lin. "The use of surface electromyography in dysphagia evaluation". *Technology and Health Care*, Vol. 25, No. 5, pp. 1025–1028, 2017.

[Wass 06]   L. Wasserman. *All of nonparametric statistics*. Springer Science & Business Media, 2006.

[Watt 15]   C. R. Watts and B. Kelly. "The effect of bolus consistency and sex on electrophysiological measures of hyolaryngeal muscle activity during swallowing". *Dysphagia*, Vol. 30, No. 5, pp. 551–557, 2015.

[Wils 12]   R. D. Wilson and E. C. Howe. "A cost-effectiveness analysis of screening methods for dysphagia after stroke". *PM&R*, Vol. 4, No. 4, pp. 273–282, 2012.

[Wyth 93]   B. J. Wythoff. "Backpropagation neural networks: a tutorial". *Chemometrics and Intelligent Laboratory Systems*, Vol. 18, No. 2, pp. 115–155, 1993.

[Yagi 17]   N. Yagi, S. Nagami, M.-k. Lin, T. Yabe, M. Itoda, T. Imai, and Y. Oku. "A noninvasive swallowing measurement system using a combination of respiratory flow, swallowing sound, and laryngeal motion". *Medical & biological engineering & computing*, Vol. 55, No. 6, pp. 1001–1017, 2017.

[Yama 18]   S. Yamaguchi, M. Ishida, K. Hidaka, S. Gomi, S. Takayama, K. Sato, Y. Yoshioka, N. Wakayama, K. Sekine, S. Matsune, *et al.* "Relationship between swallowing function and breathing/phonation". *Auris Nasus Larynx*, Vol. 45, No. 3, pp. 533–539, 2018.

[Youm 05]   S. R. Youmans and J. A. Stierwalt. "An acoustic profile of normal swallowing". *Dysphagia*, Vol. 20, No. 3, pp. 195–209, 2005.

[Youm 11]   S. R. Youmans and J. A. Stierwalt. "Normal swallowing acoustics across age, gender, bolus viscosity, and bolus volume". *Dysphagia*, Vol. 26, No. 4, pp. 374–384, 2011.

[Yous 14]   J. Yousefi and A. Hamilton-Wright. "Characterizing EMG data using machine-learning tools". *Computers in biology and medicine*, Vol. 51, pp. 1–13, 2014.

[Zare 17]   E. Zaretsky, P. Pluschinski, R. Sader, P. Birkholz, C. Neuschaefer-Rube, and C. Hey. "Identification of the most significant electrode positions in electromyographic evaluation of swallowing-related movements in humans". *European Archives of Oto-Rhino-Laryngology*, Vol. 274, No. 2, pp. 989–995, 2017.

[Zark 18]   A. Zarkada and J. Regan. "Inter-rater Reliability of the Dysphagia Outcome and Severity Scale (DOSS): Effects of Clinical Experience, Audio-Recording and Training". *Dysphagia*, Vol. 33, No. 3, pp. 329–336, 2018.

[Zenn 95]   P. M. Zenner, D. S. Losinski, and R. H. Mills. "Using cervical auscultation in the clinical dysphagia examination in long-term care". *Dysphagia*, Vol. 10, No. 1, pp. 27–31, 1995.

[Zhan 10]   X. Zhang, Y. Wang, and R. P. Han. "Wavelet transform theory and its application in EMG signal processing". In: *2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery*, pp. 2234–2238, IEEE, 2010.

[Zhan 21]   Z. Zhang, S. Mao, J. Coyle, and E. Sejdić. "Automatic annotation of cervical vertebrae in videofluoroscopy images via deep learning". *Medical Image Analysis*, p. 102218, 2021.

[Zhao 22]   H. Zhao, Y. Jiang, S. Wang, F. He, F. Ren, Z. Zhang, X. Yang, C. Zhu, J. Yue, Y. Li, *et al.* "Dysphagia diagnosis system with integrated speech analysis from throat vibration". *Expert Systems with Applications*, p. 117496, 2022.

[Zhou 09]   X.-H. Zhou, D. K. McClish, and N. A. Obuchowski. *Statistical methods in diagnostic medicine*. Vol. 569, John Wiley & Sons, 2009.

[Zhu 17]   M. Zhu, B. Yu, W. Yang, Y. Jiang, L. Lu, Z. Huang, S. Chen, and G. Li. "Evaluation of normal swallowing functions by using dynamic high-density surface electromyography maps". *Biomedical engineering online*, Vol. 16, No. 1, pp. 1–18, 2017.

[Zora 10]   D. Zoratto, T. Chau, and C. Steele. "Hyolaryngeal excursion as the physiological source of swallowing accelerometry signals". *Physiological measurement*, Vol. 31, No. 6, p. 843, 2010.

[Zwic 80]   E. Zwicker and E. Terhardt. "Analytical expressions for critical-band rate and critical bandwidth as a function of frequency". *The Journal of the Acoustical Society of America*, Vol. 68, No. 5, pp. 1523–1525, 1980.