Tracheal Breathing Sounds Analysis during Wakefulness and Sleep Using Machine Learning Methods: Characterization of Upper Airway and Daytime Screening of Obstructive Sleep Apnea

By

Farahnaz Hajipour

The University of Manitoba

In partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

Biomedical Engineering Program

University of Manitoba

Winnipeg

# Abstract

Obstructive sleep apnea (OSA) is a prevalent yet underdiagnosed respiratory health disorder. Assessment of OSA is currently based on sleep studies that are time-consuming and expensive. Therefore, developing technologies for quick OSA screening is momentous. The upper airways (UA) structural and physiological changes alter the tracheal breathing sounds (TBS) characteristics. The main objectives of this thesis are to investigate 1) whether the differences in TBS characteristics during wakefulness and their changes from wakefulness to sleep correlate with the severity of OSA and 2) whether different feature selection and classification methods over TBS can enhance the current algorithms. Accordingly, we have implemented advanced signal processing, machine learning, and statistical techniques on TBS data of individuals with non-OSA to severe OSA recorded during wakefulness and a short period of sleep. In TBS analysis, commonly, a considerable number of features are extracted from data. A major challenge in high-dimensional data analysis is related to classification and prediction of the variables of interest by building parsimonious models and removing variables that do not add any information to our model. We evaluated and compared the performance of the Random Forest (RF) and Least Absolute Shrinkage and Selection Operator (LASSO) Regularized Logistic Regression, as non-parametric and parametric feature selection and classification approaches, respectively, for OSA screening during wakefulness. Moreover, we analyzed and compared the spectral characteristics of TBS during wakefulness and sleep, and also from wakefulness to sleep in individuals with different OSA severity levels. The main outcomes of this work are: 1) confirming our hypothesis that the TBS characteristics during wakefulness and their variations from wakefulness to sleep correlate with the OSA severity and could reveal the structural changes of UA regarding OSA and its severity in a

detailed but straightforward manner, and 2) enhancing the current OSA algorithms by providing a

fast, inexpensive and robust method to stratify the severity of OSA patients with high accuracy.

*"Nothing in life is to be feared. It is only to be understood."*

*Marie Curie*

# Acknowledgments

This dissertation would not have been possible without the support of many people who in one way or another contributed to the preparation and completion of this study.

I would like to express my sincere appreciation to my supervisor, Prof. Zahra Moussavi, who guided me to be professional and do the right thing even when the road got tough. Without her invaluable assistance, encouragement, and support, the goal of this project would not have been realized.

Deepest gratitude is also due to the members of the supervisory committee, Prof. Katinka Stecina, Prof. Mohammad Jafari Jozani, and Prof. Brian Lithgow. I would like to express my very great appreciation to Dr. Jafari Jozani for his valuable and constructive suggestions during the planning and development of this research work. I would also like to thank Dr. Stecina for her advice and assistance. My grateful thanks are also extended to Prof. Lithgow for his accurate comments. It is a pleasure to acknowledge and thanks to Dr. Eleni Giannouli for her helpful discussions.

I would also like to acknowledge the Natural Sciences and Engineering Research Council (NSERC) of Canada, Mathematics of Information Technology and Complex Systems (MITACS) accelerate entrepreneur program, and the Faculty of Graduate Studies of the University of Manitoba for providing financial support for this work. I would especially like to thank Sir Gordon Wu for establishing a fund at the University of Manitoba in support of graduate students.

I would like to show my gratitude to all my friends, especially Ahmed Elwali and Mehrangiz Ashiri, whose support and friendship enabled me to finish my research.

*To My Husband Mehrdad,*

*For his sincere love and ever-present encouragement and support.*

# Contributions of Authors

This thesis is a "sandwich thesis," consisting of four individual peer-reviewed published manuscripts (Chapter II, Chapter III, Chapter IV.I, and Chapter IV.II). In addition, three peer-reviewed conference proceedings and abstracts on the findings of Chapter III and Chapter IV.I have also been published. The list of publications is below.

Ms. Hajipour has been the main contributor and first author of all the manuscripts presented in this thesis. Her contribution to this work includes developing the research questions, designing the studies, pre/processing and refining the extracted data, conducting the analyses, writing up all manuscripts, submitting three of the manuscripts, and responding to reviewers' comments. Dr. Moussavi and Dr. Jafari Jozani contributed to the conception and design of the study, drafting of the articles, and reviewing process.

List of journal publications:

- **Chapter II:** Farahnaz Hajipour, Mohammad Jafari Jozani, Ahmed Elwali and Zahra Moussavi "Regularized Logistic Regression for Obstructive Sleep Apnea Screening during Wakefulness Using Daytime Tracheal Breathing Sounds and Anthropometric Information" Med Biol Eng Comput (2019) 57: 2641. https://doi.org/10.1007/s11517-019-02052-4.

- **Chapter III:** Farahnaz Hajipour, Mohammad Jafari Jozani and Zahra Moussavi, "A Comparison of Regularized Logistic Regression and Random Forest for Daytime Diagnosis of Obstructive Sleep Apnea", Med Biol Eng Comput (2020) 58: 10. 2517–2529. https://doi-org.uml.idm.oclc.org/10.1007/s11517-020-02206-9

- **Chapter IV.I:** Farahnaz Hajipour and Zahra Moussavi, "Spectral and Higher Order Statistical Characteristics of Expiratory Tracheal Breathing Sounds During Wakefulness and Sleep in People with Different Levels of Obstructive Sleep Apnea," Journal of Medical & Biological Engineering (2019) 39: 2. 244-250. https://doi-org.uml.idm.oclc.org/10.1007/s40846-018-0409-7

- **Chapter IV.II:** Farahnaz Hajipour, Eleni Gionouli and Zahra Moussavi, " Acoustic Characterization of Upper Airway Variations from Wakefulness to Sleep in relation to Obstructive Sleep Apnea", Med Biol Eng Comput (2020) 58: 10. 2375–2385. https://doi.org/10.1007/s11517-020-02234-5.

List of conference and abstract publications:

- Farahnaz Hajipour, Mohammad Jafari Jozani and Zahra Moussavi "Random Forest Analysis of Tracheal Breathing Sounds for Predicting Obstructive Sleep Apnea" World Sleep conference, 2019.

- Farahnaz Hajipour and Zahra Moussavi. "Spectral Changes of Breathing Sounds from Wakefulness to Sleep in People with Different Severity of Sleep Apnea." C109. NEW TECHNOLOGY IN SLEEP: DIAGNOSTICS AND THERAPEUTICS. American Thoracic Society, 2017. A6961-A6961.

- Farahnaz Hajipour and Zahra Moussavi. "Expiratory Breathing Sounds Characteristics During Wakefulness and Sleep in Mild and Severe Apneic Groups." CMBES Proceedings 40.1 (2018).

# Table of Contents

# List of Tables

# List of Figures

# List of Abbreviations

| | |
|---|---|
| AHI | Apnea/hypopnea index |
| ANOVA | Analysis of variance |
| AUC | The area under the curve |
| BMI | Body mass index |
| CI | Confidence interval |
| CSA | Central sleep apnea |
| ECG | Electrocardiogram |
| EEG | Electroencephalogram |
| EMG | Electromyogram |
| EOG | Electrooculogram |
| FFT | Fast Fourier transform |
| FRC | Functional residual capacity |
| HSS | Home sleep study |
| KFD | Katz fractal dimension |
| LASSO | Least absolute shrinkage and selection operator |
| LDA | Linear discriminant analysis |
| LG | Loop Gain |
| LogVar | The logarithm of the sound's variance |
| LR | Logistic regression |
| MANOVA | Multivariate analysis of variance |
| MRI | Magnetic resonance imaging |
| NC | Neck circumference |
| NN | Neural network |
| OSA | Obstructive sleep apnea |
| OOB | Out-of-bag |
| $P_{ALV}$ | alveolar pressure |
| $PCO_2$ | The partial pressure of $CO_2$ |
| Pcrit | Critical collapsing pressure |
| $PO_2$ | The partial pressure of $O_2$ |
| $P_{PI}$ | Intra-plural pressure |
| PSD | Power spectrum density |
| $PSD_{avg}$ | Average Power spectrum density |
| PSG | Polysomnography |
| ROC | Receiver operating characteristic |
| REM | Rapid-eye-movement |
| RF | Random Forest |
| $S_aO_2$ | The oxygen saturation level of blood |
| SE | Standard error |
| SVM | Support vector machine |
| TBS | Tracheal breathing sounds |
| UA | Upper airway |

# Chapter I.   Introduction

Sleep apnea is a common and serious respiratory disorder. It is characterized by repetitive episodes of apnea and/or hypopnea during sleep [1]. Apnea is defined as a cessation of airflow for ≥10 sec, while hypopnea is a reduction of the peak airflow by ≥30% from pre-event baseline if it lasts ≥10 sec, and is associated with more than 3% oxygen desaturation [2]. The number of apnea and hypopnea events per hour of sleep is called the apnea-hypopnea index (AHI), which is one of the most important measure of the sleep apnea severity. AHI and clinical symptoms of sleep apnea are currently used for the diagnosis of this disorder. Clinical signs and symptoms of sleep apnea include daytime fatigue, excessive daytime sleepiness, morning headaches, impaired concentration, restless sleep, , gasping and choking during sleep, habitual snoring, and observed episodes of breathing cessation [2]–[4].

The repetitive nature of apneas and hypopneas cause intermittent oxygen deprivation in the blood that stimulates arousal from sleep and sleep fragmentation. Sleep apnea is associated with many deficits including excessive daytime sleepiness [5], [6], increased risk of cardiovascular and cerebrovascular disease [7]–[10], motor vehicle accidents [11], [12] and a significant decrease in quality of life [13], [14].

There are three types of sleep apnea syndrome according to the underlying mechanism: obstructive sleep apnea (OSA), central sleep apnea (CSA), and mixed apnea [15]. OSA is the most common type of sleep apnea in the general population; it accounts for >84% of the sleep apnea population [16]. As the name suggests, OSA is caused by intermittent complete (apnea) or partial (hypopnea) collapse of the upper airway (UA) during sleep. In OSA, the respiratory effort (rib cage and abdominal movement) exists, but there is no resultant respiratory airflow. CSA, in

contrast to OSA, is less common (about 0.4% of the sleep apnea population) [16]. In CSA, the effort to breathe is reduced (hypopnea) or absent (apnea) temporarily due to the dysfunction of central drive mechanisms. Mixed apnea is a combination of both OSA and CSA and accounts for about 15% of the sleep apnea population [16]. An apneic event is considered as mixed if it initiates with CSA, but closer to the termination, there is an effort to breathe without airflow [15]. The focus of this dissertation is to evaluate acoustical UA changes due to OSA and screening OSA acoustically. As the primary site of collapse in patients with OSA is in the upper respiratory tract [17], we will first elaborate on this part of the respiratory system.

## I.1 Anatomy and Physiology of the UA

The respiratory system is divided into two components, the lower respiratory track and the upper respiratory track. The anatomy and physiology of the lower airway is elaborated in Appendix A.

The human UA is a dynamic and complex structure. The principal components of UA are 1) the nasopharynx, which is behind the nose between nasal turbinate to the superior part of the soft palate, 2) the retropalatal oropharynx (velopharynx), between the hard palate to the caudal margin of the soft palate, 3) the retroglossal oropharynx, between the caudal margin of the soft palate to the base of the epiglottis, and 4) the hypopharynx, from the epiglottis to the larynx [18], Figure 1.
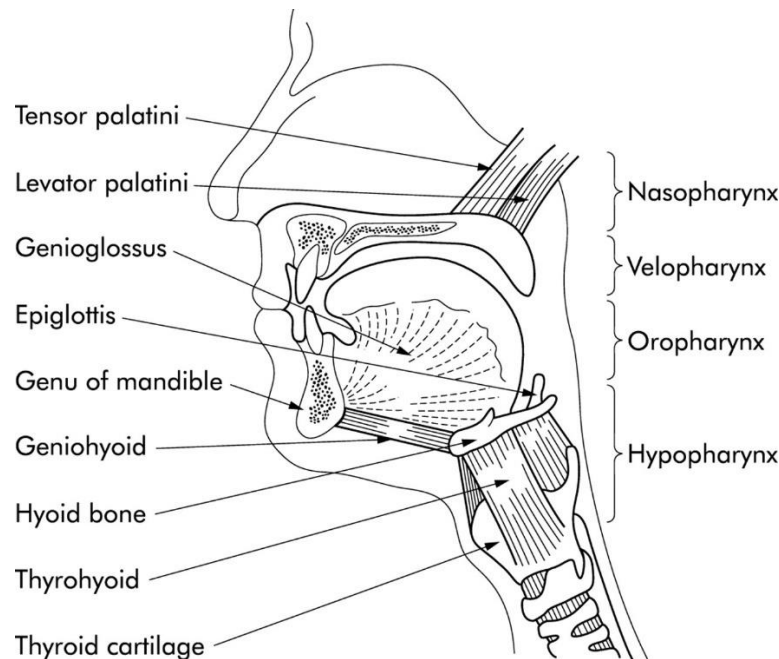
Figure 1 Anatomical representation of the upper airway and the important muscles controlling airway patency. In patients with apnea, airway collapse typically occurs behind the palate (velopharynx), the tongue (oropharynx), or both. Reproduced from [19] Copyright © 2020 Thorax with permission from BMJ Publishing Group Ltd.

The pharyngeal airway is involved in not only respiratory functions but also deglutition and vocalization. This variability in functionality requires a complex structure of tissues, bones, and muscles for controlling this portion of the airway and maintaining its patency during breathing or keeping it close for swallowing. The soft tissue structures form the walls of the UA and include the tongue, soft palate, uvula, tonsillar pillars, parapharyngeal fat pads, epiglottis, blood vessels of the neck, and lateral pharyngeal walls [18]. The bony structures of the UA consist of the nasal turbinate, the hard palate of maxilla, mandible, cervical vertebrae, and the hyoid bone [20]. The hyoid bone is the critical anchoring sites for pharyngeal muscles and soft tissues to attach [20].

There are more than 20 muscles in the UA that actively constrict and dilate to maintain the patency of the UA; which are classified into four groups of muscles that regulating the position of [20]:

- Soft palate (alai nasi, tensor palatini, levator palatini)

- Tongue (genioglossus, geniohyoid, hyoglossus, styloglossus)

- Hyoid bone (hyoglossus, genioglossus, digastric, geniohyoid, sternohyoid)

- The posterolateral pharyngeal walls (palatoglossus, pharyngeal constrictors).

Over the last years, numerous imaging studies have assessed UA structural and anatomical characteristics [21]–[25]. It has been shown that in healthy awake subjects, the UA has an ellipsoid shape, and the lateral dimension is greater than the anteroposterior dimension [21], [23]. The soft palate, tongue, and lingual tonsils mainly form the anterior wall of the oropharynx. Its posterior wall is made up of the superior, middle, and inferior constrictor muscles that lie in front of the cervical spine. The lateral pharyngeal walls are more complex and consist of muscles (hypoglossus, styloglossus, stylohyoid, stylopharyngeus, palatoglossus, palatopharyngeus, the pharyngeal constrictors), lymphoid tissue, and pharyngeal mucosa.

## I.2 Anatomy and Configuration of the UA in OSA Population

The evolution of speech in human leading to a hyoid bone without rigid attachment to other bony or cartilaginous structure. Accordingly, the shape and size of the UA are dependent on the position of the skeletal and soft tissue structures like the mandible, soft palate, tongue, and the walls of the oropharynx. It has been suggested that the passive pharyngeal airway size is a product of the two independent variables [26]:

- Size of the bony structure surrounding the UA

- Quantity of soft tissue placed in this enclosure.

As Figure 2 shows, the UA collapse may happen if the bony compound is small relative to the quantity of tissue that it must contain or vice versa [17]. Generally, in OSA patients, the ratio of

the UA soft tissue mass (lateral pharyngeal walls, tongue, and total soft tissue) is disproportionately high for the pharynx bony structures space [27], which impinges on the pharyngeal lumen in most OSA patients.
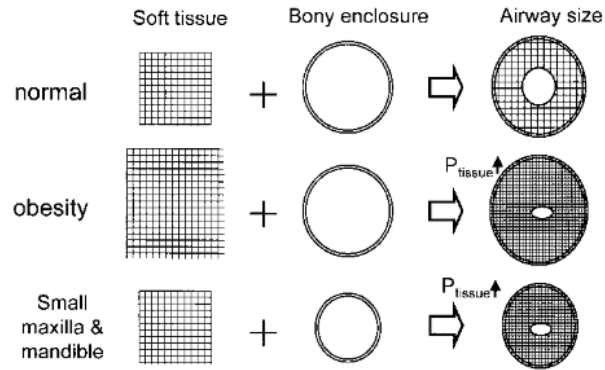


Figure 2 "Schematic explanations for the mechanical model of the pharyngeal airway. $P_{tissue}$ = pressure surrounding the collapsible tube." Originally Published in [26]. Reprinted with permission of the American Thoracic Society. Copyright © 2020 American Thoracic Society. All rights reserved. The American Journal of Respiratory and Critical Care Medicine is an official journal of the American Thoracic Society

Abnormalities of the craniofacial structures enveloping the oropharyngeal cavity are frequently observed in OSA patients. Apnea patients compared to the normal controls have shown to have a greater inferior displacement of hyoid bone and shortening and retro-positioning of the mandible and maxilla, even in the absence of distinct craniofacial abnormalities (hypoplasia and retrodisplacement of the maxilla and mandible) [26], [28]. These deficits result in retrodisplacement of the UA soft tissues; hence, restricting space in the retropalatal and retroglossal oropharyngeal cavity [26]. The mandible shortening is the most common skeletal abnormality predisposing to OSA [27].

On the other hand, an increase of the soft tissues through edema, hypertrophy, or inflammation, may decrease the UA diameter. It has been shown that the total volume of UA soft tissue is more considerable in OSA patients compared with the normal controls, resulted in a reduction in the UA size in the retropalatal and, slightly, the retroglossal area of OSA patients [29]. The lateral pharyngeal walls of the apnea patient are thicker and have higher volume while contains more non-fat soft tissue than those of normal controls [18], [29]. The soft palate, tongue, and parapharyngeal fat pad of the OSA patients are also more massive than the healthy controls [18], [29]. OSA patients may also have the adenotonsillar hypertrophy and increased edema caused by vascular congestion or inflammation secondary to the chronic trauma caused by low-frequency vibration of tissues during snoring [27].

In addition to the bony and soft tissue structures, various imaging studies have investigated the shape and length of the UA and have shown that in contrast to the normal subjects, the pharyngeal airway of snorers and OSA patients is circular or elliptical, with the long axis oriented in the anterior-posterior dimension as a result of medial displacement of the lateral pharyngeal walls [18], [21], [23]. Regarding the length of UA, it was demonstrated that the UA of men with OSA is longer, compared with those without OSA, generally measured from hard palate to the epiglottis [30].

The minimum caliber of the UA during the wakefulness and, therefore, the potential location of collapse during sleep is found to be in the retropalatal oropharynx region [18]. It has been reported that the collapse usually starts at retropalatal oropharynx regions in most (56–75%) OSA patients. Then, due to the dynamic process of airway narrowing and since the narrowing is markedly varying within and between individuals, the initial collapse is followed by caudal

6

extension to the base of the tongue in 25–44% of patients and to the hypopharyngeal region in 0–33% of patients [27], [31].

## I.3  Pathophysiology of OSA

The human UA is a complex and collapsible structure [17]. The patency of the UA is suggested to be dependent on the on the balance of pressure exerted by the UA heterogeneous surrounding soft tissue, the pressure inside the airway, and the compliance of pharyngeal walls [27]. Compliance is expressed as the change in cross-sectional area per unit change in pressure. It indicates the ease with which an airway can be deformed. It has been shown that the increased pharyngeal compliance in addition to the combination of the increased soft tissue mass and craniofacial abnormalities may increase the critical collapsing pressure (Pcrit), the pressure at which the UA collapses [26], [32]. In normal subjects, the Pcrit is noticeably negative; hence, their UA requires a negative intraluminal pressure for closure. Contrarily, in severe OSA patients, the Pcrit is positive, indicating a more collapsible UA even at positive airway pressure.

The pharynx is susceptible to collapse at the beginning of inspiration or at the end of expiration. At the beginning of inspiration, in an individual with narrowed and/or highly compliant pharynx, a greater intra-pharyngeal negative pressure is necessary to maintain breathing airflow [33]. However, due to the normal reduction in pharyngeal dilator muscle tone at the onset of sleep, the increased negative pressure will further narrow the UA; resulting in a defective cycle leading to complete UA collapse [33], [34]. On the other hand, the collapse of the UA at the end of expiration is either due to absent/decreased pharyngeal dilator muscle activity, or it is due to the mechanical factors including declining lung volume throughout the expiratory phase [27], [35].

The sleep-related narrowing and increased compliance/collapsibility of the UA are critical contributors to the pathogenesis of OSA [27]. Therefore, conditions that influence UA narrowing and collapsibility, including obesity, aging, male gender, and decreased lung volume, can predispose individuals to OSA syndrome. A detailed description on how these factors affect UA narrowing and collapsibility is provided in Appendix A.

Although anatomical factors and obesity are major determinants of apnea development, approximately 20-40% of OSA individuals are not obese [34]. In these individuals, non-anatomic factors such as impaired UA dilator muscle activity, instability of ventilatory control (loop gain), and low arousal threshold are major contributors in facilitating UA collapse [34], [36]. It has been shown that 69% of OSA individuals have one or more of the non-anatomical pathophysiological characteristics [37].

The UA dilator muscles have an important role in maintaining airway patency. It has been reported that compliance of the airway is decreased by the presence of UA muscles activities [20]. During wakefulness, OSA patients have increased dilator muscles activities, especially for the genioglossus muscle, as compared with healthy subjects, to compensate for their abnormal anatomy and/or more collapsible pharyngeal airway [38]. Therefore, OSA individuals do not experience any difficulty in breathing while awake. However, the activity of the UA dilator muscles is decreased at the sleep onset due to a reduction in central respiratory drive [27]. In OSA individuals, a greater reduction in dilator muscles activities, especially in the genioglossus muscle's activities, has been observed; that is superimposed on an anatomically vulnerable UA. Hence, it would lead to UA collapse during sleep [27], [29], [39], [40]. Over the course of an apnea/hypopnea event, to compensate for the rising of partial pressure of $CO_2$ ($P_aCO_2$), the genioglossus activity increases until the patient awakens. Isono et al. suggest that there is a

"seesaw" balance between the negative pressure within the airway and the opposing dilating force generated by the UA muscles (Figure 3) [41]. During wakefulness, the balance is maintained in favor of patency in both OSA and healthy groups. During sleep, however, collapse occurs in OSA patients due to the imbalance between the negative intraluminal pressure and dilating muscle forces [20].
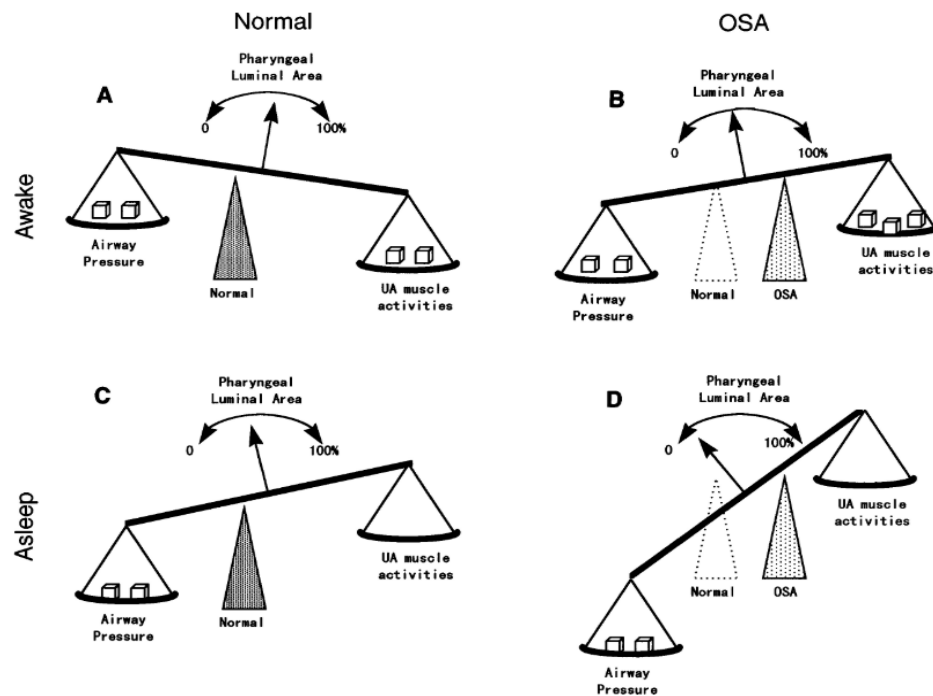


Figure 3 schematic model of upper airway patency and balance of pressures in normal and OSA patients during wakefulness and sleep. Originally Published in [41]. Copyright © 2020 American Physiological Society

Despite the importance of the control of dilator muscles in OSA pathogenesis, more variables are also likely to be important. Multiple feedback loops, including the chemo- and mechano-receptors, regulate the quantity and pattern of ventilation to maintain the $CO_2$ and $O_2$ levels of the blood [42]. In general, when feedback loops regulate a mechanical system, the system has the potential to become unstable [42]. Loop gain (LG) expresses the gain or sensitivity of the feedback-

controlled system that controls ventilation. Mathematically, LG is defined as the ratio of the corrective response to the disturbance (e.g., hyperpnea) over the disturbance itself (e.g., apnea or hypopnea) [42]. The LG greater than 1 means the respiratory disturbance will lead to a quick and intense response such that ventilation will oscillate for an unlimited time; hence, the system could become unstable (Figure 4).
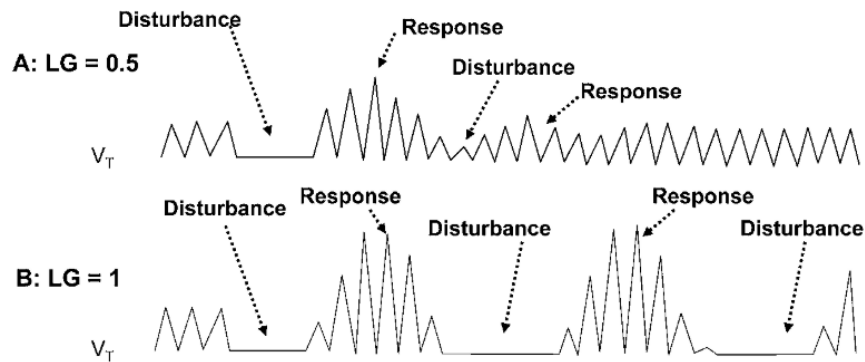


Figure 4 The ventilatory response to an apnea (first disturbance in both figures) is demonstrated for (A) an individual with a loop gain (LG) of 0.5 and (B) an individual with an LG of 1. In A, ventilation quickly returns to a regular pattern, whereas in B, a sustained oscillation is established. Originally published in [42]: White, D. P. Pathogenesis of obstructive and central sleep apnea. American journal of respiratory and critical care medicine. 2005; 172:1363-1370. OI: 10.1164/rccm.200412-1631SO. Reprinted with permission of the American Thoracic Society. Copyright © 2020 American Thoracic Society. All rights reserved. The American Journal of Respiratory and Critical Care Medicine is an official journal of the American Thoracic Society.

As the UA muscles' activities depend on the respiratory drive, the unstable ventilatory control would lead to unstable activity of pharyngeal musculature and UA collapse [42]. Studies have been shown that the LG of individuals with severe OSA is much higher than that in those with mild OSA or in healthy controls; this implies of an inherently less stable ventilatory control system among OSA individuals during sleep [37], [43].

Another non-anatomical factor leading to UA collapse is one's low arousal threshold. Arousal (a temporary sate change from sleep to wakefulness) is a defensive mechanism that often triggered by apnea and hypopnea events, helping to terminate them [44]. It is an important factor contributing to the accumulation of various respiratory stimuli, which are helpful in recruiting dilator muscles' activities and reopening of the constricted UA [45]. However, in the majority of OSA individuals, the transient responses that occur with frequent transitions between sleep and arousal include excessive excitatory drives to the respiratory system that cause a surge of ventilation and consequently transient hypocapnia [46], [47]. These changes might contribute to further UA collapse [47]. Accordingly, arousal has a paradoxical role in sleep apnea as it terminates the existing apnea/hypopnea events but can also trigger new ones. It has been reported that if sleep can be maintained for an adequate duration, the accumulation of sufficient physiological stimuli can activate UA dilator muscles without an arousal [47], [48]. Therefore, premature arousal as a result of a low respiratory arousal threshold (i.e., the ease with which an individual wake up to respiratory stimuli) may lead to repetitive apnea and hypopnea due to inadequate accumulation of physiological stimuli to enable UA muscles recruitment [49].

## I.4   Prevalence and Diagnosis of OSA

OSA is a relatively common breathing disorder that can affect the health of all age groups [50], [51]. It has been reported that more than 10% of North American society and nearly 1 billion adults aged 30-69 years around the world are affected by OSA [52]–[54].  However, OSA is still massively underdiagnosed. Available data suggest that >80% of OSA remain undiagnosed and untreated [55], [56]. Untreated moderate to severe OSA (AHI>15) is associated with increased morbidity and mortality [53]. Thus, characterizing the severity of OSA is an important part of the OSA diagnosis as it can lead to more appropriate therapeutic decisions [57].

## I.4.1  Polysomnography (PSG)

Currently, the severity of sleep apnea is determined by the full-night Polysomnography (PSG) assessment, the gold standard for determining the AHI and diagnosing OSA [58]. For the PSG assessment, individuals must spend the entire night in a sleep laboratory, while they are being connected to many monitoring electrodes and are under the supervision of a skilled technician. Various physiological signals including electroencephalogram (EEG), electrocardiogram (ECG), electrooculogram (EOG), submental and leg electromyograms (EMG), nasal airflow, body position, pulse oximetry, and abdominal/thoracic respiratory effort to provide a full assessment of sleep quality  are recorded during a PSG study [58], [59].

Electroencephalography is a method to record the EEG signal, the spontaneous electrical activity of the brain over a period of time, from electrodes placed on the individual's scalp. It is performed to monitor the sleep/wakefulness status of individuals (whether the patient is awake or asleep), identify different sleep stages, and record arousals from sleep [59]. The electrooculography is a technique to record EOG signal, the eye movement. The EOG signal, along with the chin EMG is mandatory for distinguishing wakefulness and rapid-eye-movement (REM) sleep from other sleep stages [59]. Electrocardiography is the process of producing ECG signals. It records the electrical activity of the heart to detect the heart rate and nocturnal cardiac arrhythmias that may link to sleep disorders [59]. Abdominal and chest movements are used to assess the breathing effort and breathing pattern. Pulse oximetry is used to assess oxygen desaturation of blood and is an important parameter for scoring respiratory disturbance events, especially the hypopnea for which the reduction in oxygen saturation is a criterion. Using the recorded signals during PSG assessment, apneas and hypopneas are scored, and AHI is measured (using the standard AASM Scoring Manual definitions of events (version 2.4) [60]). AHI is the

most commonly used metric to measure the severity of sleep apnea. Individuals with AHI<5 are considered as non-OSA, 5<AHI<15 as mild OSA, 15<AHI<30 as moderate OSA, and AHI>30 as severe OSA [61].

Although an overnight complete PSG assessment is a comprehensive and reliable means for diagnosing OSA, PSG studies are labor-intensive, expensive, and time-consuming. There are long waiting lists reaching up to 2 years in some places in Canada, particularly small towns and remote areas [62]. Thus, in emergencies that the OSA status of a patient is needed, it may not be feasible to perform a quick OSA assessment using the PSG approach. As a result, sleep apnea status of around 80% of the patients proceeding for operations requiring full anaesthesia are unknown at the time of surgery [63]. Inadequate preoperative assessment of OSA of those patients may increase their postoperative complications risks [63]. Besides, it has been reported that the estimated prevalence of sleep-related breathing disorders increased by about 14-55% over the last two decades [64]. The increased number of people in need of OSA diagnosis will further increment the waiting time of undergoing PSG assessment. Therefore, it may disrupt the timely diagnosis of OSA, especially for severe cases in need of quick treatment.

Accordingly, PSG assessment is not suitable as an early-stage diagnostic tool. The daytime OSA screening tools can provide the prioritization of patients for PSG studies based on the severity of their condition, thus increasing the social benefits and cost-effectiveness of medical resources. It can also facilitate a well-timed perioperative risk stratification.

## I.4.2  Questionnaires

The current quick OSA assessment in hospitals that is commonly used for patients undergoing surgery requiring full anesthesia is subjective questionnaires, such as the STOP-BANG

Questionnaire [65]. Questionnaires are simple, fast, and inexpensive assessments to identify patients with a high risk of OSA syndrome who may benefit from receiving early diagnosis and management [66]. These methods have shown to have a high sensitivity ( 90%) but at the cost of a very low specificity (<40%) [65]. Any pre-screening tool with low specificity could potentially result in identifying a significant number of participants as high-risk patients, hence, increases the false positive and referral rate to the full PSG study. Therefore, it reduces the cost-effectiveness of such assessments and could impose a substantial impact on health care costs.

Accordingly, there is a need for a reliable and prompt objective tool with high sensitivity and specificity for OSA screening to provide efficient preoperative management. Such a solution will help in diminishing the PSG waiting list, the plausible mortality rate, and the cost of perioperative OSA diagnosis [67]. Additionally, the development of an objective technology with the ability to providing a physiological interpretation of the cause of the OSA disorder will be an important step in improving treatment.

## I.4.3    Acoustic Diagnosis of OSA

Tracheal sounds heard at the suprasternal notch, are a measure of tracheal wall vibration set into motion by the passage of turbulent airflow from the UA, including vocal cords, trachea, and pharynx [68]. The structural and physiological properties of the UA, such as its patency, shape, and size, affect the resonance frequency of the UA, which is detected by tracheal sounds analysis [20], [69], [70]. As mentioned in Section I.3, anatomical and functional abnormalities of the UA play an important role in the pathogenesis of OSA. These abnormalities are expected to change the acoustical properties of the tracheal speech and respiratory sounds, including snoring and tracheal breathing sounds (TBS). It has been shown that there is a dynamic variation in physiological and structural properties of UA but with different degrees concerning the

14

sleep/wakefulness status and OSA severity [29], [39], [40], [71]–[74]. Consequently, a dynamic change in the acoustical properties of tracheal sounds regarding these factors is predicted.

Advances in techniques for non-invasive sound measurement and signal analysis have increased the role of tracheal sound analysis in the diagnosis of respiratory disorders. Studies have been shown that tracheal sounds are rich in features that could provide information about snoring and the site from which it arises [75] as well as respiratory flow and UA abnormalities [76]. Tracheal sound analysis has also been used to detect the OSA mechanism and to distinguish individuals with OSA from non-snores and healthy persons during wakefulness and sleep [76]–[81]. Consequently, tracheal sound analysis is a simple and non–invasive way to study the pathophysiology of the UA and is a potentially valuable tool that could be used for early OSA screening and diagnosis with comparable outcomes to that of current standards.

In a study by Azarbarzin et al. [82], the diagnostic ability of snoring sounds was investigated in 57 individuals (42 OSA and 15 non-OSA). They extracted snoring sound segments from the recorded tracheal sounds of individuals, then, using PSG information, divided them into three classes of non-apneic, hypopneic, and post-apneic. They extracted several features from the snoring sound segments and compared them among three classes within each subject. Their results showed a significant change in snoring sound characteristics of hypopneic and post-apneic classes over time. Using this finding, they discriminated non-OSA snorers from OSA through the linear discriminant analysis (LDA) classifier. Their results showed an accuracy of 96.4% with the sensitivity (specificity) of 92.2% (100%). One of the limitations of this study was the relatively small dataset with unbalanced groups.

In a study by Yadollahi et al. [83], pulse oximetry and TBS data of 66 participants during sleep were used to develop a fully automated acoustic technology for OSA monitoring. After the

detection of snoring and breathing episodes, they extracted various features from these parts to estimate the participants' AHI values. The detected apnea-hypopnea events were highly correlated to those identified using PSG. The sensitivity and specificity of their proposed method for diagnosis of OSA with AHI more than 5, 10, 15, and 20 were in the range of [74.3%, 91.6%] and [82.4%-97.8%], respectively.

More recently, in another study by Nakano et al. [84], the tracheal sounds spectrogram images were estimated from the data of 1,852 individuals to detect the apnea/hypopnea events. They analyzed the constructed images using a convolutional deep neural network. The diagnostic sensitivity and specificity of their method for different AHI cut-off values, i.e., 5, 15, and 30, were in the range of [76%-99%], [99%-99%], respectively. In that study, mainly data of moderate and severe OSA individuals were used; therefore, the applicability to the general population was not investigated. In addition, their dataset included only the Japanese population. Hence, it is not known whether the accuracy of their method would remain the same on other ethnic groups as anatomical factors do affect the tracheal breathing sounds.

The mentioned studies, while using a smaller number of measured signals compared to PSG, are still nocturnal methods that rely on signals recorded during overnight sleep. Such nocturnal nature makes these methods time-consuming. As elaborated in Section I.3, due to an increase in the dilator muscles activities of the OSA individuals, they do not experience any difficulty in breathing while awake [38]. Therefore, it might appear counterintuitive that such a disorder that only manifests during sleep can be screened (or predicted) while a person is awake. However, studies have shown that OSA is associated with chronic physiological and anatomical changes that persist during wakefulness [21], [23], [29], [72]. Thus, our team, as the pioneer, and a few other researchers around the world have used daytime breathing sounds or speech sounds analysis for

OSA monitoring and screening to predict the individuals' OSA condition with better accuracy than the commonly used questionnaires and a comparable accuracy of those overnight methods [79]–[81], [85], [86].

In a study by Solà-Soler et al. [80], the formant frequencies of TBS and their variation from inspiration to expiration were investigated in mild-moderate (AHI<30, n=10) and severe (AHI>=30, n=13) OSA patients during wakefulness. Their LDA classification method over the formant features and anthropometric information resulted in testing sensitivity (specificity) and the accuracy of 88.9% (84.6%) and 86.4%, respectively. Despite the high classification results, as their features were extracted from the entire dataset, those results are considered as biased. Another limitation of that study is the small sample size, given the heterogeneity of their dataset.

In another recent study [79], the breathing sounds and speech signals of 398 individuals during wakefulness were studied for OSA detection and severity estimation. Their system comprised three different sub-systems based on analysis of breathing sounds within continuous speech signals, sustained vowels, and continuous speech signals using deep neural networks (NN). The combination of AHIs, estimated by each sub-system, with age and body mass index (BMI), produced a composite system that estimates AHI using linear regression. Their classification decision using an AHI threshold of 15 achieved an average accuracy, sensitivity, and specificity of 77.14%, 75%, and 79%, respectively. In that study, due to the black-box nature of NN systems, contributions of the selected features to OSA diagnosis and their interpretation were missed.

In the study reported in [81], speech sounds of two unbalanced groups of OSA (AHI>5, n=67) and non-OSA (AHI<5, n=26) were recorded. One hundred acoustic features were extracted. After feature reduction, discriminative features, including linear prediction coefficients, were selected.

Their Gaussian mixture model-based classification system resulted in testing sensitivity (specificity) of 79% (83%) for male OSA and 84% (86%) for female OSA, respectively.

In one of our team's previous study [85], TBS of 130 individual were recorded during wakefulness before PSG assessment. Various acoustic features were extracted from 105 randomly selected participants (56 non-OSA (AHI ≤ 5) and 49 OSA (AHI≥10)). After conducting a multi-stage feature reduction using an exhaustive leave-two-out routine (by leaving one subject from each of the two groups as testing and the remaining as training), two spectral features with the most discriminative power between the two OSA groups were selected. Using a 2-class support vector machine (SVM) classifier, they achieved the testing sensitivity, specificity, and accuracy of 82.61%, 85.22%, and 83.92%, respectively. Despite the high accuracy of the classification results, their proposed procedure was complicated and computationally expensive. Moreover, the random selection of training participates for feature extraction was done only once; thus, the test results might have some bias.

## I.5   Goals and Objectives

Despite several promising steps towards the development of an alternative tool for the diagnosis of OSA, the currently available methods suffer from considerable limitations such as small sample size and lack of a blind test dataset; thus, their reported results are not completely unbiased. The aim of this dissertation is to enhance one of the current technologies for diagnosis of OSA to address the limitations of its current diagnostic algorithm. More specifically, the main goals of this thesis are:

- Investigating whether different feature selection and classification methods over TBS and anthropometric information during wakefulness can enhance the outcomes of the current AwakeOSA algorithm [87] while keeping the computational cost low.

- Finding the TBS characteristics that reveal the structural changes of UA in relation to OSA and its severity in a detailed manner using TBS during wakefulness and a short period of sleep.

The detailed objectives of the thesis are elaborated as the followings.

## Objective 1– **Investigating the Regularized Logistic Regression for Wakefulness OSA Screening**

A problem associated with the acoustic screening of OSA is that one could extract a variety of potential features from TBS, which would lead to a large number of predictors for screening OSA; thus, making the classification of data high-dimensional. In high-dimensional settings, it is useful to build parsimonious models that are easy to understand and interpret; this is usually done by removing the variables that are redundant and do not add any information to the model [88]. By focusing on a selected subset of features, one might also be able to improve the prediction performance [89]. In many studies, the manual selection of features is inappropriate. Therefore, relying on automatic feature selection and classification techniques is recommended [89]. In Chapter II, we propose to apply a famous regression-based model, the logistic regression with the least absolute shrinkage and selection operator (LASSO) or simply LASSO regularized LR, for feature reduction and classification of OSA individuals using the combination of the anthropometric information and TBS during daytime while the subjects are awake. The main sub-objectives of Chapter II are to:

19

1. use a linear and parametric regression-based model for developing a simple and fast technique for feature reduction and classification of non-OSA (AHI<5, n=74) and moderate/severe OSA (AHI $\geq$ 15, n=90) individuals by selecting the discriminative features from the vast number of potential features, which could be extracted from anthropometric information and power spectra of the TBS recorded during daytime while the subjects are awake.

2. discuss the physiological interpretation of the selected TBS features and anthropometric information of the same algorithm in relation to the structural changes of UA due to OSA.

The work presented in Chapter II has been published in [90].

## Objective 2− **Comparing Two Machine Learning Models for Fast Daytime Screening of OSA**

The main goals of analyzing any data are to make the prediction/classification and finding the association between the inputs and response variables. There are two different approaches to achieve these goals: data modeling and algorithmic modeling [91]. In data modeling culture, a stochastic data model is assumed for the data generation process, i.e., a model is fit to data to draw conclusions. The LASSO regularized LR model that has been elaborated in Chapter II is a popular model-based method [90]. On the other hand, in algorithmic modeling the data generation process is assumed to be complicated and unknown, and the approach is to find an algorithm that operates on inputs to generate the response [91]. One of our objectives in Chapter III is to elaborate on the Random Forest (RF), a famous algorithmic-based model, and specify its parameters in a way to be applicable for real-time OSA screening during wakefulness. Other objectives in Chapter III are to find a metric to explain the importance of its selected features for clinical use, and to clarify the association between these features and OSA severity. It is known that for each data problem, the

best model depends on the underlying problem and the trade-off between interpretability and the complexity of the desired model [92]. Therefore, another objective of Chapter III is to compare the RF and the LASSO regularized LR models with respect to the selected features, the physiological interpretation of their selected features regarding OSA, any noticeable differences in their classification results, and their computational complexities. In summary, the main sub-objectives of Chapter III are to:

1. use a non-linear and non-parametric tree-based model as an automatic feature selection and classification technique for analysis of anthropometric information and TBS data collected during daytime (wakefulness) to predict the likelihood of OSA and its severity in non-OSA (AHI<5) and moderate/severe OSA (AHI ≥ 15) individuals.

2. compare the findings of the tree-based and regression-based (presented in Chapter II) models with respect to selected features, the physiological interpretation of those features regarding the OSA, any noticeable differences in their classification results, their computational complexities, and their performances on real-time screening of OSA.

The work presented in Chapter III has been published in [93]. In addition, three severity group classification using RF modeling was also investigated that is presented in Appendix B.

## Objective 3− **Acoustic Characterization of Upper Airway Variations with Respect to OSA**

Another major application of TBS analysis is the non-invasive and quick investigation of UA anatomy and physiology. As mentioned in Section I.4.3, the physiological properties of the UA change dynamically, but this change is dependent on the sleep/wakefulness status and severity of OSA [40], [71], [74]. In Chapter IV, we hypothesize TBS characteristics during wakefulness and

sleep, and their changes from wakefulness to sleep are correlated with the OSA severity and they may have the potential to reveal the physiological and structural changes of UA including the narrowing, thickness, and stiffness due to OSA. In Chapter IV.I we study the spectral and higher-order statistical characteristics of TBS during a short period in stage 2 of sleep and during wakefulness to investigate any difference between the TBS characteristics of mild-OSA (AHI <15) and moderate/severe-OSA (AHI >15) groups. Then, in Chapter IV.II, we investigate their pattern of changes from wakefulness to sleep in each of the two OSA groups and analyze whether those changes are correlated to the severity of OSA and have any classification power for identifying the mild-OSA and moderate/severe OSA individuals. We assess the correlation between the TBS features as well as their within-group (from wakefulness to sleep) and between-group (between mild and moderate/severe OSA groups) differences to the AHI. To the best of our knowledge, this study is the first to explore the potential role of TBS analysis in assessing UA changes from wakefulness to sleep with respect to OSA and its severity. The main expected outcomes of these works are finding the TBS characteristics that reveal the structural changes of the UA regarding OSA and its severity in a detailed but straightforward manner. Also, enhancing the current OSA screening algorithms by providing a non-time-consuming and less expensive method to stratify the severity of OSA patients in a fast but more precise way. The main sub-objectives of Chapter IV are to:

1. investigate the spectral and higher-order statistical differences between TBS of mild-OSA (AHI<15) and moderate/severe OSA (AHI $\geq$ 15) individuals during either wakefulness or a short period of sleep.

2. investigate the correlation between the OSA severity and the pattern of changes in spectral characteristics of TBS from wakefulness to stage 2 of sleep within and between the mild-OSA and moderate/severe OSA individuals using a short period of sleep.

3. investigate the role of such spectral changes from wakefulness to sleep in screening and classification of mild-OSA and moderate/severe OSA individuals in a non-laborious, non-time-consuming, and less expensive manner using a short period of sleep instead of the full overnight sleep study.

4. investigate whether these changes have the potential to indirectly and non-invasively reveal the changes in physiological characteristics of the UA (including narrowing, and thickness) due to OSA in a detailed but straightforward manner.

The works presented in Chapter IV have been published in [32], [94].

## I.6   Thesis Organization

This dissertation document is consisting of 5 chapters and two appendices. In this current chapter, Chapter I- Introduction, we introduced the key concepts, objectives, and scope of the thesis. The following chapters consist of 4 individual manuscripts, published in peer-reviewed journals (Chapters II, III, IV. I, and IV.II). Chapter IChapter II deals with objective 1, Chapter III deals with objective 2, and Chapter IV deals with objective 3. In Chapter V, a summary of the findings presented in previous chapters, limitations of the studies, as well as suggestions and recommendations for future works related to this research are provided. Appendix A presents details of the anatomy and physiology of the respiratory system. Appendix B investigates the use of TBS recorded during wakefulness for multi-class OSA classification using the RF approach explained in objective 2.

# References

[1]     A. Malhotra and D. P. White, "Obstructive sleep apnoea," *Lancet*, vol. 360, no. 9328, pp. 237–245, Jul. 2002.

[2]     B. RB, A. CL, and H. SM, *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.5*. 2018.

[3]     D. J. Eckert, A. Malhotra, and A. S. Jordan, "Mechanisms of Apnea," *Prog. Cardiovasc. Dis.*, vol. 51, no. 4, pp. 313–323, Jan. 2009.

[4]     L. J. Epstein *et al.*, "Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults.," *J. Clin. sleep Med.  JCSM*, vol. 5, no. 3, pp. 263–76, 2009.

[5]     D. Rizzo *et al.*, "The role of fatigue and sleepiness in drivers with obstructive sleep apnea," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 62, pp. 796–804, 2019.

[6]     J. F. Pagel, "Excessive daytime sleepiness," *Am. Fam. Physician*, vol. 79, no. 5, pp. 391–396, 2009.

[7]     J. D. L. Lattimore, D. S. Celermajer, and I. Wilcox, "Obstructive sleep apnea and cardiovascular disease," *J. Am. Coll. Cardiol.*, vol. 41, no. 9, pp. 1429–1437, 2003.

[8]     A. Yoshihisa and Y. Takeishi, "Sleep disordered breathing and cardiovascular diseases," *J. Atheroscler. Thromb.*, vol. 26, no. 4, pp. 315–327, 2019.

[9]     J. R. D. Espiritu, "Health Consequences of Obstructive Sleep Apnea Journal of Sleep Disorders and Therapy," *J. Sleep Disord. Ther.*, vol. 8, no. 5, pp. 1–17, 2019.

[10]   T. D. Bradley, J. S. Floras, and editors, *Sleep apnea: implications in cardiovascular and cerebrovascular disease.* CRC Press, 2016.

[11]   D. J. Gottlieb, J. M. Ellenbogen, M. T. Bianchi, and C. A. Czeisler, "Sleep deficiency and motor vehicle crash risk in the general population: A prospective cohort study," *BMC Med.*, vol. 16, no. 1, pp. 1–10, 2018.

[12]   A. Noda, F. Yasuma, S. Miyata, K. Iwamoto, Y. Yasuda, and N. Ozaki, "Sleep Fragmentation and Risk of Automobile Accidents in Patients with Obstructive Sleep Apnea—Sleep Fragmentation and Automobile Accidents in OSA," *Health (Irvine. Calif).*, vol. 11, no. 02, pp. 171–181, 2019.

[13]   C. A. Moyer, S. S. Sonnad, S. L. Garetz, J. I. Helman, and R. D. Chervin, "Quality of life in obstructive sleep apnea: A systematic review of the literature," *Sleep Med.*, vol. 2, no. 6, pp. 477–491, 2001.

[14]   J. F. Garvey, M. F. Pengo, P. Drakatos, and B. D. Kent, "Epidemiological aspects of obstructive sleep apnea," *J. Thorac. Dis.*, vol. 7, no. 5, pp. 920–929, 2015.

[15]   J. Chesson *et al.*, "Practice parameters for the indications for polysomnography and related procedures," *Sleep*, vol. 20, no. 6, pp. 406–422, 1997.

[16] T. I. Morgenthaler, V. Kagramanov, V. Hanak, and P. A. Decker, "Complex sleep apnea syndrome: Is it a unique clinical syndrome?," *Sleep*, vol. 29, no. 9, pp. 1203–1209, 2006.

[17] D. P. White and M. K. Younes, "Obstructive Sleep Apnea," in *International Encyclopedia of Public Health*, vol. 2, no. October, Elsevier, 2012, pp. 308–314.

[18] R. J. Schwab, K. B. Gupta, W. B. Gefter, L. J. Metzger, E. A. Hoffman, and A. I. Pack, "Upper airway and soft tissue anatomy in normal subjects and patients with sleep-disordered breathing. Significance of the lateral pharyngeal walls.," *Am. J. Respir. Crit. Care Med.*, vol. 152, no. 5, pp. 1673–1689, 1995.

[19] R. B. Fogel, A. Malhotra, and D. P. White, "Sleep . 2: Pathophysiology of obstructive sleep apnoea/hypopnoea syndrome," *Thorax*, vol. 59, no. 2, pp. 159–163, Feb. 2004.

[20] I. Ayappa and D. M. Rapoport, "The upper airway in sleep: physiology of the pharynx," *Sleep Med. Rev.*, vol. 7, no. 1, pp. 9–33, Feb. 2003.

[21] K.-H. Liu *et al.*, "Sonographic measurement of lateral parapharyngeal wall thickness in patients with obstructive sleep apnea.," *J. Sleep*, vol. 30, no. 11, pp. 1503–8, 2007.

[22] H.-M. Lun, S.-Y. Zhu, R.-C. Liu, J.-G. Gong, and Y.-L. Liu, "Investigation of the Upper Airway Anatomy With Ultrasound," vol. 32, no. 1, pp. 86–92, 2016.

[23] G. C. Barkdull, C. a Kohl, M. Patel, and T. M. Davidson, "Computed Tomography Imaging of Patients With Obstructive Sleep Apnea," *Laryngoscope*, vol. 118, no. 8, pp. 1486–1492, Aug. 2008.

[24] R. J. Schwab *et al.*, "Identification of Upper Airway Anatomic Risk Factors for Obstructive Sleep Apnea With Volumetric Mri," *ON-LINE DATA Suppl. Blue-200208-866OC.R2*.

[25] A. Javed, Y.-C. Kim, M. Khoo, S. Ward, and K. Nayak, "Dynamic 3D MR Visualization and Detection of Upper Airway Obstruction during Sleep using Region Growing Segmentation.," *IEEE Trans. Biomed. Eng.*, vol. 9294, no. c, pp. 431–437, 2015.

[26] T. Watanabe, S. Isono, A. Tanaka, H. Tanzawa, and T. Nishino, "Contribution of body habitus and craniofacial characteristics to segmental closing pressures of the passive pharynx in patients with sleep-disordered breathing," *Am. J. Respir. Crit. Care Med.*, vol. 165, no. 2, pp. 260–265, 2002.

[27] C. M. Ryan and T. D. Bradley, "Pathogenesis of obstructive sleep apnea," *J. Appl. Physiol.*, vol. 99, no. 6, pp. 2440–2450, 2005.

[28] R. L. Riha, P. Brander, M. Vennelle, and N. J. Douglas, "A Cephalometric Comparison of Patients With the Sleep Apnea/Hypopnea Syndrome and Their Siblings," *Sleep*, Mar. 2005.

[29] R. J. Schwab *et al.*, "Identification of Upper Airway Anatomic Risk Factors for Obstructive Sleep Apnea with Volumetric Magnetic Resonance Imaging," *Am. J. Respir. Crit. Care Med.*, vol. 168, no. 5, pp. 522–530, Sep. 2003.

[30] A. Malhotra *et al.*, "The male predisposition to pharyngeal collapse: Importance of airway length," *Am. J. Respir. Crit. Care Med.*, vol. 166, no. 10, pp. 1388–1395, 2002.

[31] N. Bhattacharyya, S. P. Blake, and M. P. Fried, "Assessment of the airway in obstructive

sleep apnea syndrome with 3-dimensional airway computed tomography," *Otolaryngol. - Head Neck Surg.*, vol. 123, no. 4, pp. 444–449, 2000.

[32]   F. Hajipour and Z. Moussavi, "Spectral and Higher Order Statistical Characteristics of Expiratory Tracheal Breathing Sounds During Wakefulness and Sleep in People with Different Levels of Obstructive Sleep Apnea," *J. Med. Biol. Eng.*, vol. 39, no. 2, pp. 244–250, Apr. 2019.

[33]   F. Lopez-Jimenez, F. H. Sert Kuniyoshi, A. Gami, and V. K. Somers, "Obstructive sleep apnea: Implications for cardiac and vascular disease," *Chest*, vol. 133, no. 3, pp. 793–804, 2008.

[34]   S. Javaheri *et al.*, "Sleep Apnea: Types, Mechanisms, and Clinical Cardiovascular Consequences," *J. Am. Coll. Cardiol.*, vol. 69, no. 7, pp. 841–858, 2017.

[35]   H. Schneider *et al.*, "Modulation of upper airway collapsibility during sleep: Influence of respiratory phase and flow regimen," *J. Appl. Physiol.*, vol. 93, no. 4, pp. 1365–1376, 2002.

[36]   D. J. Eckert, A. Malhotra, and A. S. Jordan, "Mechanisms of Apnea," *Progresas Cardiovasc. Dis.*, vol. 51, no. 4, pp. 313–323, Jan. 2009.

[37]   D. J. Eckert, D. P. White, A. S. Jordan, A. Malhotra, and A. Wellman, "Defining phenotypic causes of obstructive sleep apnea: Identification of novel therapeutic targets," *Am. J. Respir. Crit. Care Med.*, vol. 188, no. 8, pp. 996–1004, 2013.

[38]   M. Younes, "Role of respiratory control mechanisms in the pathogenesis of obstructive sleep disorders," *J. Appl. Physiol.*, vol. 105, no. 5, pp. 1389–1405, Nov. 2008.

[39]   R. B. Fogel *et al.*, "The effect of sleep onset on upper airway muscle activity in patients with sleep apnoea versus controls," *J. Physiol.*, vol. 564, no. 2, pp. 549–562, Apr. 2005.

[40]   A. S. Jordan *et al.*, "Airway Dilator Muscle Activity and Lung Volume During Stable Breathing in Obstructive Sleep Apnea," *Sleep*, vol. 32, no. 3, pp. 361–368, Mar. 2009.

[41]   S. Isono, J. E. Remmers, A. Tanaka, Y. Sho, J. Sato, and T. Nishino, "Anatomy of pharynx in patients with obstructive sleep apnea and in normal subjects.," *J. Appl. Physiol.*, vol. 82, no. 4, pp. 1319–1326, 1997.

[42]   D. P. White, "Pathogenesis of obstructive and central sleep apnea," *Am. J. Respir. Crit. Care Med.*, vol. 172, no. 11, pp. 1363–1370, 2005.

[43]   M. Younes, M. Ostrowski, W. Thompson, C. Leslie, and W. Shewchuk, "Chemical control stability in patients with obstructive sleep apnea," *Am. J. Respir. Crit. Care Med.*, vol. 163, no. 5, pp. 1181–1190, 2001.

[44]   A. Malhotra and A. Jordan, "The importance of arousal in obstructive sleep apnea—updates from the American Thoracic Society 2016," *J. Thorac. Dis.*, vol. 8, no. S7, pp. S542–S544, 2016.

[45]   M. L. Stanchina *et al.*, "Genioglossus muscle responsiveness to chemical and mechanical stimuli during non-rapid eye movement sleep," *Am. J. Respir. Crit. Care Med.*, vol. 165, no. 7, pp. 945–949, 2002.

[46]   N. Deacon and A. Malhotra, "Potential protective mechanism of arousal in obstructive sleep apnea," *J. Thorac. Dis.*, vol. 8, no. S7, pp. S545–S546, 2016.

[47]   D. J. Eckert and M. K. Younes, "Arousal from sleep: implications for obstructive sleep apnea pathogenesis and treatment.," *J. Appl. Physiol.*, vol. 116, no. 3, pp. 302–13, 2014.

[48]   M. Younes, M. Ostrowski, R. Atkar, J. Laprairie, A. D. Siemens, and P. Hanly, "Mechanisms of breathing instability in patients with obstructive sleep apnea," *J. Appl. Physiol.*, vol. 103, no. 6, pp. 1929–1941, 2007.

[49]   D. J. Eckert *et al.*, "Eszopiclone increases the respiratory arousal threshold and lowers the apnoea/hypopnoea index in obstructive sleep apnoea patients with a low arousal threshold," *Clin. Sci.*, vol. 120, no. 12, pp. 505–514, 2011.

[50]   I. E. Gabbay and P. Lavie, "Age- and gender-related characteristics of obstructive sleep apnea," *Sleep Breath.*, vol. 16, no. 2, pp. 453–460, Jun. 2012.

[51]   J. Chan, J. C. Edman, and P. J. Koltai., "Obstructive Sleep Apnea in Children," *Am. Fam. Physician*, vol. 69, no. 5, pp. 1147–1160, 2004.

[52]   L. Grote, "The global burden of sleep apnoea," *Lancet Respir. Med.*, vol. 7, no. 8, pp. 645–647, 2019.

[53]   T. Young *et al.*, "Sleep disordered breathing and mortality: eighteen-year follow-up of the Wisconsin sleep cohort.," *Sleep*, vol. 31, no. 8, pp. 1071–1078, 2008.

[54]   A. V Benjafield *et al.*, "Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis," *Lancet Respir. Med.*, vol. 7, no. 8, pp. 687–698, Aug. 2019.

[55]   F. Chung *et al.*, "Society of Anesthesia and Sleep Medicine Guidelines on Preoperative Screening and Assessment of Adult Patients with Obstructive Sleep Apnea," *Anesth. Analg.*, vol. 123, no. 2, pp. 452–473, 2016.

[56]   American Academy of Sleep Medicine, *Hidden Health Crisis Costing America Billions: Underdiagnosing and Undertreating Obstructive Sleep Apnea Draining Healthcare System.* Mountain View, CA: Sullivan & Frost, 2016.

[57]   A. N. Rama, S. H. Tekwani, and C. A. Kushida, "Sites of Obstruction in Obstructive Sleep Apnea," *Chest*, vol. 122, no. 4, pp. 1139–1147, Oct. 2002.

[58]   C. A. Kushida *et al.*, "Practice Parameters for the Indications for Polysomnography and Related Procedures: An Update for 2005," *Sleep*, vol. 28, no. 4, pp. 499–523, 2005.

[59]   A. S. BaHammam, D. E. Gacuan, S. George, K. L. Acosta, S. R. Pandi-Perumal, and R. Gupta, "Chapter 25 Polysomnography I: Procedure and Technology," in *Synopsis of Sleep Medicine*, no. March 2018, Apple Academic Press, 2016, pp. 443–453.

[60]   R. B. Berry *et al.*, "AASM Scoring Manual Updates for 2017 (Version 2.4)," vol. 13, no. 5, pp. 665–666, 2017.

[61]   J. W. Quan, S. F., Gillin, J. C., Littner, M. R., & Shepard, "Sleep–Related Breathing Disorders in Adults: Recommendations for Syndrome Definition and Measurement

Techniques in Clinical Research," *Sleep*, vol. 22, no. 5, pp. 667–689, Aug. 1999.

[62]   M. Povitz *et al.*, "Clinical pathways and wait times for OSA care in Ontario, Canada: A population cohort study," *Can. J. Respir. Crit. Care, Sleep Med.*, vol. 3, no. 2, pp. 91–99, 2019.

[63]   S. G. Memtsoudis, M. C. Besculides, and M. Mazumdar, "A Rude Awakening — The Perioperative Sleep Apnea Epidemic," *N. Engl. J. Med.*, vol. 368, no. 25, pp. 2352–2353, Jun. 2013.

[64]   P. E. Peppard, T. Young, J. H. Barnet, M. Palta, E. W. Hagen, and K. M. Hla, "Increased prevalence of sleep-disordered breathing in adults," *Am. J. Epidemiol.*, vol. 177, no. 9, pp. 1006–1014, 2013.

[65]   H. Y. Chiu *et al.*, "Diagnostic accuracy of the Berlin questionnaire, STOP-BANG, STOP, and Epworth sleepiness scale in detecting obstructive sleep apnea: A bivariate meta-analysis," *Sleep Med. Rev.*, vol. 36, pp. 57–70, 2017.

[66]   J. Corral-Peñafiel, J. L. Pepin, and F. Barbe, "Ambulatory monitoring in the diagnosis and management of obstructive sleep apnoea syndrome," *Eur. Respir. Rev.*, vol. 22, no. 129, pp. 312–324, 2013.

[67]   S. Articles, "Practice Guidelines for the Perioperative Management of Patients with Obstructive Sleep Apnea," *Anesthesiology*, vol. 120, no. 2, pp. 268–286, Feb. 2014.

[68]   K. N. Priftis, L. J. Hadjileontiadis, and M. L. Everard, *Breath Sounds From Basic Science to Clinical Practice*. Cham: Springer International Publishing, 2018.

[69]   T. Penzel and A. Sabil, "The use of tracheal sounds for the diagnosis of sleep apnoea," *Breathe*, vol. 13, no. 2, pp. e37–e45, Jun. 2017.

[70]   I. Sanchez and H. Pasterkamp, "Tracheal sound spectra depend on body height.," *Am. Rev. Respir. Dis.*, vol. 148, no. 4 Pt 1, pp. 1083–1087, 1993.

[71]   Z. Lan, A. Itoi, M. Takashima, M. Oda, and K. Tomoda, "Difference of pharyngeal morphology and mechanical property between OSAHS patients and normal subjects," *Auris Nasus Larynx*, vol. 33, no. 4, pp. 433–439, Dec. 2006.

[72]   Y. Finkelstein *et al.*, "Velopharyngeal Anatomy in Patients With Obstructive Sleep Apnea Versus Normal Subjects," *J. Oral Maxillofac. Surg.*, vol. 72, no. 7, pp. 1350–1372, Jul. 2014.

[73]   Z. Wu, W. Chen, M. C. K. Khoo, S. L. Davidson Ward, and K. S. Nayak, "Evaluation of upper airway collapsibility using real-time MRI," *J. Magn. Reson. Imaging*, vol. 44, no. 1, pp. 158–167, 2016.

[74]   F. J. TRUDO, W. B. GEFTER, K. C. WELCH, K. B. GUPTA, G. MAISLIN, and R. J. SCHWAB, "State-related Changes in Upper Airway Caliber and Surrounding Soft-Tissue Structures in Normal Subjects," *Am. J. Respir. Crit. Care Med.*, vol. 158, no. 4, pp. 1259–1270, Oct. 1998.

[75]   A. Azarbarzin, "Higher Order Statistics , and its Application for by," 2012.

[76] A. Yadollahi, "Respiratory Sound Analysis for Flow Estimation During Wakefulness and Sleep , and its Applications for Sleep Apnea Detection and Monitoring," University of Manitoba, 2011.

[77] A. Azarbarzin and Z. Moussavi, "Intra-subject variability of snoring sounds in relation to body position, sleep stage, and blood oxygen level," *Med. Biol. Eng. Comput.*, vol. 51, no. 4, pp. 429–439, 2013.

[78] A. Elwali and Z. Moussavi, "A Novel Decision Making Procedure during Wakefulness for Screening Obstructive Sleep Apnea using Anthropometric Information and Tracheal Breathing Sounds," *Sci. Rep.*, vol. 9, no. 1, pp. 1–12, 2019.

[79] R. Simply, E. Dafna, and Y. Zigel, "Diagnosis of Obstructive Sleep Apnea using Speech Signals from Awake Subjects," *IEEE J. Sel. Top. Signal Process.*, vol. 4553, no. 1403, 2019.

[80] J. Sola-Soler, J. A. Fiz, A. Torres, and R. Jane, "Identification of Obstructive Sleep Apnea patients from tracheal breath sound analysis during wakefulness in polysomnographic studies," *Conf. Proc. ... Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2014, pp. 4232–4235, 2014.

[81] E. Goldshtein, A. Tarasiuk, and Y. Zigel, "Detection of obstructive sleep apnea using speech signal analysis," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 5, pp. 1373–1382, 2010.

[82] A. Azarbarzin and Z. Moussavi, "Snoring sounds variability as a signature of obstructive sleep apnea," *Med. Eng. Phys.*, vol. 35, no. 4, pp. 479–485, Apr. 2013.

[83] A. Yadollahi, E. Giannouli, and Z. Moussavi, "Sleep apnea monitoring and diagnosis based on pulse oximetery and tracheal sound signals," *Med. Biol. Eng. Comput.*, vol. 48, no. 11, pp. 1087–1097, Nov. 2010.

[84] H. Nakano, T. Furukawa, and T. Tanigawa, "Tracheal sound analysis using a deep neural network to detect sleep apnea," *J. Clin. Sleep Med.*, vol. 15, no. 8, pp. 1125–1133, 2019.

[85] A. Elwali and Z. Moussavi, "Obstructive Sleep Apnea Screening and Airway Structure Characterization During Wakefulness Using Tracheal Breathing Sounds," *Ann. Biomed. Eng.*, vol. 45, no. 3, pp. 839–850, Mar. 2017.

[86] A. Montazeri, E. Giannouli, and Z. Moussavi, "Assessment of Obstructive Sleep Apnea and its Severity during Wakefulness," *Ann. Biomed. Eng.*, vol. 40, no. 4, pp. 916–924, Apr. 2012.

[87] Zahra Moussavi and A. Montazeripouragha, "System and methods of acoustical screening for obstructive sleep apnea during wakefulness," US 9,931,073 B2, 2018.

[88] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, vol. 103. New York, NY: Springer New York, 2013.

[89] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *J. Mach. Learn. Res.*, vol. 3, no. 3, pp. 1157–1182, 2003.

[90] F. Hajipour, M. Jafari Jozani, A. Elwali, and Z. Moussavi, "Regularized logistic regression for obstructive sleep apnea screening during wakefulness using daytime tracheal breathing

sounds and anthropometric information," *Med. Biol. Eng. Comput.*, vol. 57, no. 12, pp. 2641–2655, Nov. 2019.

[91]   L. Breiman, "Statistical modeling: The two cultures," *Stat. Sci.*, vol. 16, no. 3, pp. 199–215, 2001.

[92]   T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY: Springer New York, 2009.

[93]   F. Hajipour, M. J. Jozani, and Z. Moussavi, "A comparison of regularized logistic regression and random forest machine learning models for daytime diagnosis of obstructive sleep apnea," *Med. Biol. Eng. Comput.*, vol. 58, no. 10, pp. 2517–2529, Oct. 2020.

[94]   F. Hajipour, E. Giannouli, and Z. Moussavi, "Acoustic characterization of upper airway variations from wakefulness to sleep with respect to obstructive sleep apnea," *Med. Biol. Eng. Comput.*, vol. 58, no. 10, pp. 2375–2385, Oct. 2020.

# Chapter II.Regularized Logistic Regression for Wakefulness OSA Screening

## Regularized Logistic Regression for Obstructive Sleep Apnea Screening during Wakefulness Using Daytime Tracheal Breathing Sounds and Anthropometric Information

Farahnaz Hajipour, Mohammad Jafari Jozani, Ahmed Elwali and Zahra Moussavi

*Abstract* — Obstructive sleep apnea (OSA) is a prevalent health problem. Developing a technology for quick OSA screening is momentous. In this study, we used regularized logistic regression to predict the OSA severity level of 199 individuals (116 males) with apnea/hypopnea index (AHI) ≥15 (moderate/severe OSA) and AHI<5 (non-OSA) using their tracheal breathing sounds (TBS) recorded during daytime, while they were awake. The participants were guided to breathe through their nose, and then through their mouth at their deep breathing rate. The least absolute shrinkage and selection operator (LASSO) feature selection approach was used to select the discriminative features from the power spectra of the TBS and the anthropometric information. Using a five-fold cross-validation procedure, five different training-sets and their corresponding blind-testing sets were formed. The average blind-testing classification accuracy over the five different folds was found to be 79.3%±6.1 with the sensitivity (specificity) of 82.2%±7.2 (75.8%±9.9). The accuracy for the entire dataset was found to be 81.1% with sensitivity (specificity) of 84.4% (77.0%). The feature selection and classification procedures were intelligible and fast. The selected features were physiologically meaningful. Overall, the results show that TBS analysis can be used as a quick and reliable prediction of the presence and severity of OSA during wakefulness without a sleep study.

## II.1  Introduction

Obstructive sleep apnea-hypopnea disorder (OSA) is characterized by repetitive narrowing or complete closure of the upper airway (UA) during sleep, leading to complete cessation (apnea) and/or partial ($\geq$50%) reduction (hypopnea) of airflow that lasts at least 10sec and is associated with a minimum of 4% drop in oxygen saturation level of blood ($S_aO_2$) [1]. The apneic episodes often contribute to frequent arousals or awakenings in order to restore airway functioning. These frequent arousals reduce sleep continuity and quality [2]. Untreated OSA is associated with many deficits, including excessive daytime sleepiness, increased risk of motor vehicle accidents [1], and increased risk of cardiovascular disease [3].

OSA is a common health problem which affects all age groups. Between 9% to 38% of the general adult population is suffering from OSA, while it is higher in men compared to women and much higher in the elderly groups [4]. Since OSA is still underdiagnosed, these statistics are believed to underestimate the actual numbers [5]. The severity of OSA is currently measured by the number of apneic episodes per hour of sleep using the apnea/hypopnea index (AHI). AHI<5 considered as non-OSA, and 5$\leq$ AHI<15, 15$\leq$ AHI<30, AHI$\geq$30 considered as mild, moderate, and severe OSA, respectively [6]. AHI is measured by Polysomnography (PSG) assessment over the night sleep. PSG records various signals including electroencephalogram, electrocardiogram, electrooculogram, electromyography of chins and leg, body position, nasal airflow, $S_aO_2$, as well as abdominal and thoracic movements in order to provide a full assessment of sleep quality [7]. While full-nocturnal PSG is considered as the gold standard for sleep apnea assessment, it is time-consuming, laborious, expensive, and not easily accessible, particularly in small towns and remote areas. Thus, in emergency situations that the OSA status of a patient is needed it might not be feasible to perform a quick assessment of OSA status using PSG approach, hence underdiagnose

OSA. For example, around 80% of patients presenting for operation are undiagnosed at the time of surgery [5]. Inadequate preoperative assessment of OSA in these patients may increase their postoperative complications risks. Therefore, a quick and reliable screening OSA assessment during daytime without a full-nocturnal sleep study is highly desirable.

The current quick pre-operative OSA screening method in hospitals is based on using subjective questionnaires, such as Berlin Questionnaire, Epworth sleepiness scale, STOP, and STOP-BANG [8], that collect anthropometric information (gender, neck-circumference, age, etc.) [9]. These methods have shown to have a high sensitivity ( 90%) but at the cost of a very low specificity (<40%) [8]. Consequently, they could potentially result in identifying a much higher number of participants as high-risk patients and reducing the cost-effectiveness of such assessment. On the other hand, our team and a few other researchers around the world have used tracheal breathing sounds (TBS) for OSA monitoring and screening during daytime to predict their OSA condition with better accuracy than the commonly used method based on questionnaires [10]–[12]. The premise of our technique to screen OSA during wakefulness is based on the fact that the TBS change as the UA structure varies in OSA individuals [13], [14]. It has been shown that these structural changes are present not only during sleep [15], [16] but also during daytime when patients are awake (wakefulness) [17]–[20].

In a study reported in [12], the formant frequencies of TBS and their variation from inspiration to expiration were investigated for 10 mild-moderate (AHI<30) and 13 severe (AHI>=30) OSA patients during wakefulness before getting asleep. They used an LDA classification approach using formant features and some anthropometric information. They reported sensitivity (specificity) of 88.9% (84.6%) and accuracy of 86.4%. Since the features were extracted from the entire dataset (due to the small sample size), those results are considered as biased. Also, LDA is a generative

approach that makes strong and sometimes unrealistic assumptions such as normality as well as equal variance structure of all extracted features for each class. In our team's previous study [10], many features were extracted from the TBS' power spectra of 105 randomly selected participants (56 non-OSA, AHI ≤ 5 and 49 OSA, AHI≥10) out of a total of 130 participants. Next, using data of all participants and within an exhaustive leave-two-out routine (by leaving one subject from each of the two groups as testing and the remaining as training) six features with the best discriminative power were selected. Afterward, using a support vector machine (SVM) classifier with linear kernel the classification accuracy of all combination of three-feature sets were calculated. This procedure was repeated until each participant's data was used at least once as a test. Finally, from the most repeated 3 features, 2 of them with the lowest correlation were selected. Then, through another exhaustive leave-two-out cross-validation, the 2-class SVM classifier resulted in 83.83% and 83.92% accuracies for training and test datasets, respectively. The sensitivity (specificity) was 83.92% (85.20%) for training and 82.61% (85.22%) for test datasets [10]. Despite the high accuracy of the classification results, the proposed procedure in [10] is complicated and computationally expensive. Moreover, the random selection of training participates for feature extraction was done only once; thus, the reported results might have some bias.

In this paper, we propose to apply the logistic regression with least absolute shrinkage and selection operator (LASSO) or $l_1$-penalty for classification of apneic individuals using the combination of the anthropometric information and TBS during daytime while the subjects are awake. LASSO is a powerful regularization method that performs the feature selection for statistical models [21] by shrinking and sometimes setting some of the coefficients of the regression variables to zero [21], [22]. Removing some of the coefficients can reduce the variance

without a substantial increase of the bias, hence, increase the prediction accuracy. Moreover, by eliminating the irrelevant variables that are not associated with the response variable, the LASSO helps to increase the model interpretability and reduce the overfitting [21], [22]. This method has been widely used in variable selection and classification in many clinical fields [23]–[25]. We also use LASSO logistic regression for our classification problem, as it is a discriminant approach that does not require normality assumption or homogeneity of features in OSA classes. To the best of our knowledge, this is the first research where these techniques are used to assist in OSA diagnosis from TBS signals recorded during the daytime. We have validated our approach on a larger number of participants compared to that of previous studies. The physiological interpretation of the selected features in relation to the structural changes of UA due to OSA are also discussed.

## II.2 Method

### II.2.1 Data

Data of this study were collected from 199 participants suspected of OSA, prior to overnight PSG assessment at Sleep Disorder Lab at Misericordia Health Centre (Winnipeg, MB, Canada). The Biomedical Research Ethics Board of the University of Manitoba approved the study, and all participants signed an informed consent form prior to participating in the study. After our recording, participants proceeded to overnight PSG preparation and sleep assessment. We collected their anthropometric information and AHI (from PSG assessment) afterward to compare with our proposed method's AHI prediction and calculate its accuracy. Our collected data included 74 (29 males) non-OSA with AHI<5, and 90 (66 males) moderate/severe OSA with AHI≥15. The remaining 35 (21 males) mild-OSA participants with 5≤AHI<15 were dealt with separately. Table 1 presents the anthropometric information of the participants.

Table 1 Anthropometric information's mean and their corresponding standard deviations (std) for the non-OSA, moderate/severe-OSA and mild-OSA groups. AHI is apnea/hypopnea index, NC is neck circumference, BMI is body mass index and MP is Mallampati-score.

| | non-OSA (AHI<5, n=74) | Moderate/ severe-OSA (AHI≥15, n=90) | Mild-OSA (5≤AHI<15, n=35) |
|---|---|---|---|
| **AHI±std** | 1.2±1.3 | 42.8±32.7 | 8.7±2.6 |
| **NC ($cm$) ±std** | 38.7±4.1 | 44.1±3.7 | 42.2±6.3 |
| **Age (year) ±std** | 46.8±12.9 | 52.2±11.6 | 52.3±11.7 |
| **BMI ($kg/m^2$) ±std** | 30.6±6.3 | 36.4±8.0 | 34.3±8.4 |
| **Sex (Male: Female)** | 29: 45 | 66: 24 | 21: 14 |
| **Height ($cm$) ±std** | 167.5±9.4 | 170±10.3 | 170.6±9.9 |
| **Weight ($kg$) ±std** | 85.1±19.3 | 106±23.2 | 100.8±20.8 |
| **MP (I-II-III-IV)** | 41-19-6-8 | 22-30-22-16 | 19-6-9-1 |

## II.2.2 Recording Procedure

TBS data were recorded during the daytime, while the participants were awake, and prior to their PSG assessment. Participants' TBS were recorded by a miniature microphone (Sony ECM-77B) embedded in a chamber with 2mm cone-shaped space with skin placed over the suprasternal notch of trachea using a double-sided adhesive ring tape. The chamber was embedded in a soft neckband wrapped around the participants' neck to prevent plausible movement and comfort of the participant. The recorded breathing sounds were amplified using a Biopac (DA100C) amplifier, band-pass filtered in the frequency range of (0.05-5000 Hz) and digitized at 10240 Hz sampling rate. In supine position with head resting on a pillow, the participants were instructed to breathe deeply in two maneuvers: first through their mouth with a nose clip in place; second through their nose with their mouth closed; we recorded 5 full breath cycles in each maneuver.

## II.2.3 Pre-processing and signal analysis

All of our recordings were done in a hospital setting in a relatively noisy background. Thus, the recorded sounds contained various types of noises (e.g. vocal noises, air conditioner sounds, etc.) in addition to the desired breathing sounds. We applied different processes to remove unwanted noises from the collected TBS. As the first step, we examined the recorded sounds in the time-and-frequency domain by visual and auditory means using the spectrogram program of Matlab$^{TM}$ software; this step was done manually to ensure 100% accuracy in selecting noise-free breathing sounds segments. Consequently, all of the breathing cycles that included artifacts, vocal noises, swallowing, and low signal to noise ratio compared to the background noise were excluded from the analysis. Figure 5 shows a sample spectrogram that contains breathing sounds in addition to various type of noises including swallowing, coughing, vocal noises, and the microphone movement sounds due to the excessive neck tissues of the participants that leads to attachment of the microphone to the skin.
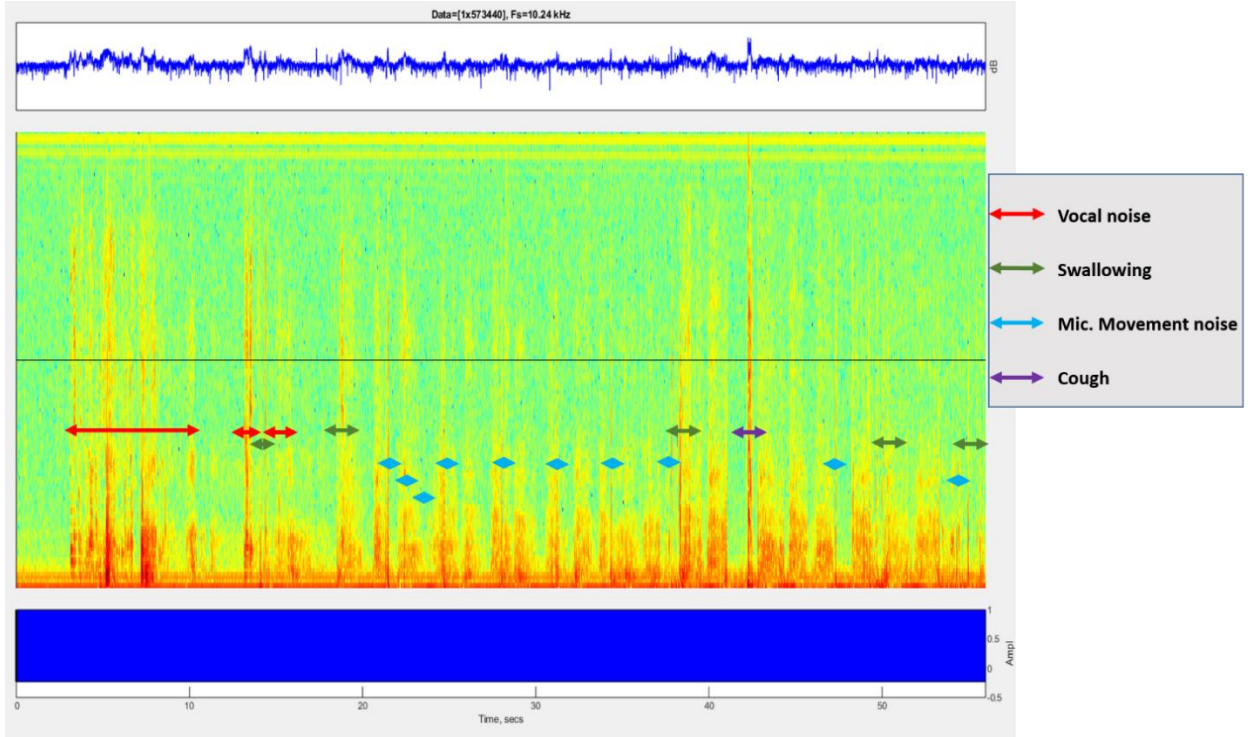
Figure 5 Spectrogram of a sample recording with vocal noises, swallowing, coughing, and microphone movement sounds due to attachment of the microphone to the participant's skin.

As the second step of noise reduction, we used a 5th-order Butterworth band-pass filter in the range of [75-2500 Hz] to keep the main frequency components of the breathing phases of each selected sound and to suppress the effect of low and high-frequency noises (including highest amplitude component of the heartbeats (which is less than 75 Hz [26]), neck muscle artefact, fundamental frequency of the power line (60 Hz in Canada), and background noises). The reason for choosing the 2500 Hz as the high frequency bound is that most of the TBS energy is claimed to be below 2500 Hz [27]. Next, similar to our previous studies [10], [28], we normalized the filtered sounds first by their variance envelope (64-sample sequence moving average filter of the signal), and secondly, by their energy to compensate for probable different flow rates in different breathing cycles.

In this study, we did not record the respiratory flow of participants, however, to ensure the respiratory phases all recording procedures started at the inspiration phase and marked by the voice of the experimenter. Using that auditory marker, the inspiratory and expiratory phases were separated manually and analyzed independently. Then, we used the method introduced in [14] to determine the approximately stationary portion of each normalized sound. In brief, the logarithm of the sound's variance (LogVar) was calculated, and the sound segments corresponding to the middle part (50% duration around the maximum) were extracted as the stationary portion of each breathing phase. Finally, we estimated the power spectrum density (PSD) of the stationary portion of each TBS signal using Welch's method [29] in windows of 205-point (~20 ms) with 50% overlap between the successive windows. The values of optimum window size and overlap for segmenting the TBS signal were selected based on the results of our previous study [28] and our various pilot studies that showed these values estimates the mostly differentiable PSD signals between the non-OSA and OSA subjects.

Following the above routine, for each participant, we estimated the PSD in both linear and logarithmic scale for TBS recorded in each mouth and nasal maneuver. The PSD estimation was done for each inspiratory/expiratory phase separately and also for their combination (summation/subtraction). This routine resulted in 16 different PSD signals for each participant; they were then averaged over the breathing cycles of each participant's data. Figure 6 outlines the above-mentioned pre-processing and signal analysis.
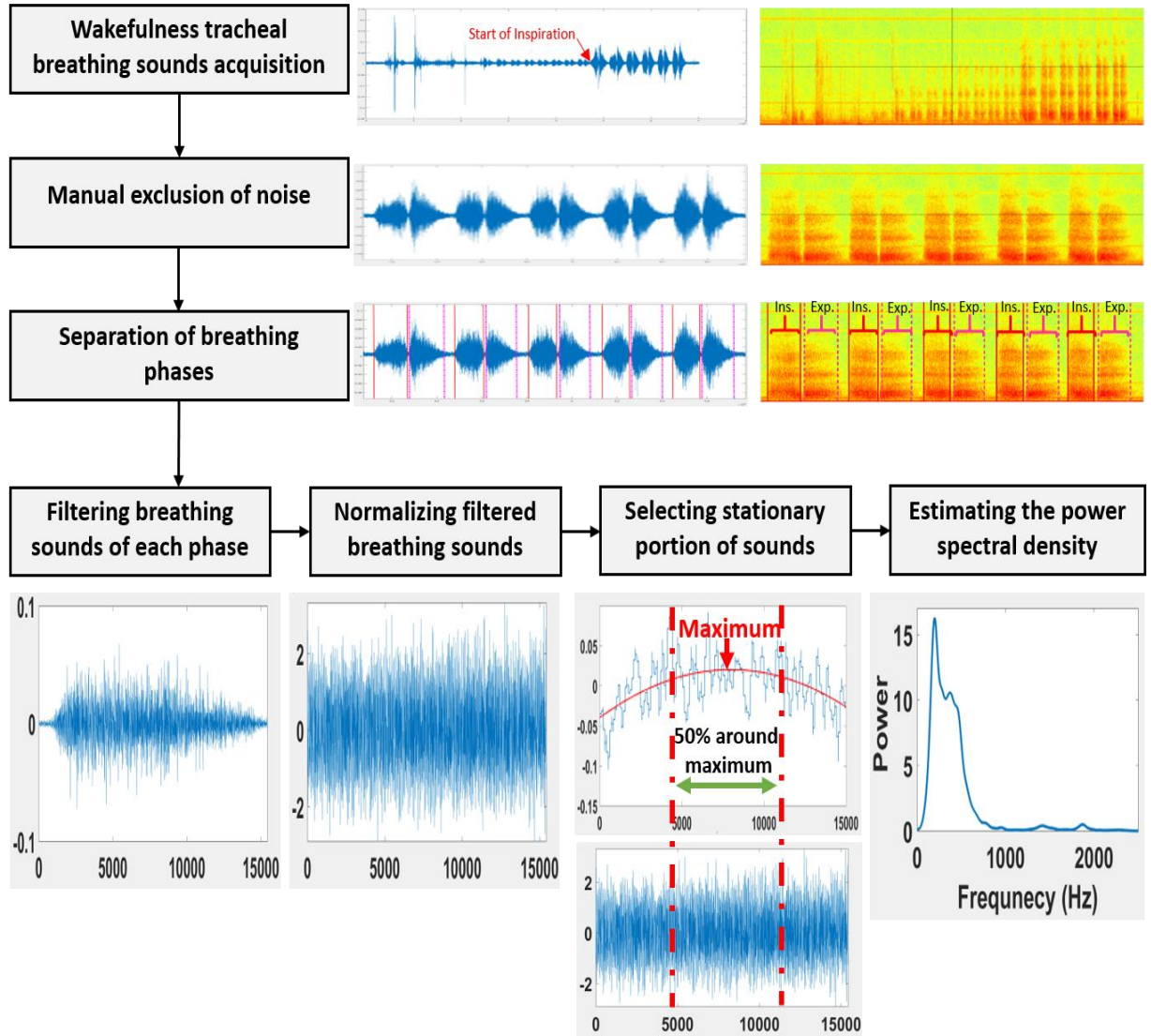
Figure 6 Pre-processing and signal analysis framework. Tracheal breathing sounds were extracted during daytime while the participants were awake. After manual exclusion of noisy sounds, the breathing sounds were separated as inspiratory and expiratory phases. Each individual phase of breathing sounds was filtered and normalized to purify tracheal breathing sounds. Afterward the power spectrum density was estimated for the approximately stationary portion of each normalized sound.

## II.2.4 Feature Extraction

We used the 5-fold cross-validation routine to randomly split the set of 164 participants associated with the non-OSA and moderate/severe-OSA groups into five non-overlapping groups (folds), each consisting of approximately 33 participants (20% non-OSA and 20% moderate/severe

OSA). Each time, participants' TBS data of one-fold were left out and considered as blind-testing data, and data of the remaining 4 folds (131 participants) were considered as training data and used for feature extraction. The test error was estimated by averaging the five resulting error estimates. Data of the mild-OSA group (5<AHI<15, n=35) were not used during feature selection, and they were dealt separately as described later in Section II.2.5.

As we mentioned in Section II.2.3, within each training set, we estimated the PSD for all participants separately. In order to extract the characteristic features between the non-OSA and moderate/severe OSA groups of this study, we separately averaged the PSD signals of non-OSA group's participants and the moderate/severe OSA group's participants within each training dataset and for each of the 16 different PSD signals and calculated their 95% confidence intervals (CI). The 95% CI for either non-OSA or moderate/severe OSA groups were calculated according to the *Eq. II.1*.

$$CI = \overline{PSD} \pm Critical\_value \times \frac{std(PSD)}{\sqrt{n}}, \qquad\qquad Eq.\ II.1$$

where $\overline{PSD}$ is the average of PSD for each OSA group, $std$ is the standard deviation, $n$ is the number of participants in each OSA groups, and $Critical\_value$ is the t-statistics (values of t-distributions) with $n - 1$ degree of freedom.

We considered the mean and slope of the regions with different slopes and the non-overlapped area between the average power spectra of two groups in association with their corresponding 95% CIs as characteristic features to be further selected for classification. As an example, Figure 7 shows the average power spectra that calculated from the summation of inspiratory and expiratory mouth TBS signals in linear and logarithmic scale for the non-OSA and moderate/severe-OSA groups in one of the training sets as well as their corresponding 95% CIs. TBS features were

extracted from the regions between the solid, dotted, and dashed lines. Overall, we extracted 78 TBS features. Combining these TBS features with seven anthropometric features of each participant (sex, age, height, weight, body mass index (BMI), neck circumference (NC), and Mallampati score (MP)) resulted in a total of eighty-five features.



(a)



(b)
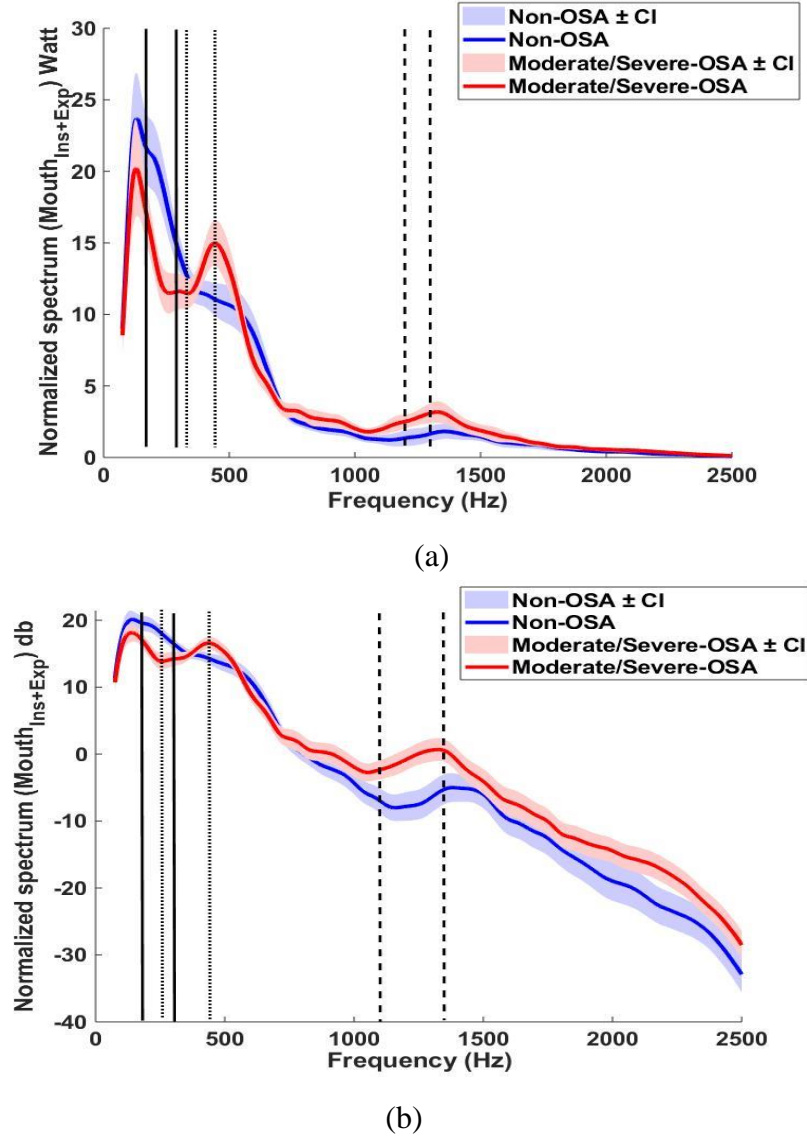
Figure 7 Average power spectra of the summation of mouth inspiratory and expiratory breathing sounds with their 95% confidence intervals (CI; shadows) in non-OSA (blue) and moderate/severe-OSA (red) groups, using one of the training sets with 131 participants in a) linear scale, b) logarithmic scale with base of 10. The area between solid lines, dotted lines and dashed lines shows the regions where the features were extracted.

## II.2.5 Feature reduction and classification

Due to the high number of extracted features, we used a feature selection/reduction approach to find the most characteristic features for classification of the two moderate/severe OSA and non-OSA groups. The first step of feature reduction was to apply an unpaired t-test on the features of the participants in each training set and to select the statistically significantly different features (p<0.01) between non-OSA and moderate/severe-OSA groups within that training set. In order to meet the statistical significance of 99% for the overall test, the significance level of each test was considered as $1 - (1 - 0.01)^{1/85} \cong 1.2 \times 10^{-4}$ [30], [31]. Within each training set, on average 11 features with p-value$> 1.2 \times 10^{-4}$ were omitted, and the features with the discrimination potential between the two groups were kept for further analysis.

To classify the two non-OSA and moderate/severe-OSA groups, we considered a linear logistic regression model for the severity level of participants:

$$log(\frac{\text{Pr }(Y=1|X=x)}{1-\text{Pr }(Y=1|X=x)}) = \beta_0 + \beta^T x, \qquad\qquad Eq.\ II.2$$

where $Y \epsilon \{0,1\}$ is the true severity label of participants, $X$ is the matrix of $(n \times p)$ for $n$ participants and $p$ features, $\beta$ is the coefficient vector, and $\text{Pr}(Y = 1|X = x)$ represents the probability of assigning a participant to the moderate/severe group.

To further reduce the number of features and obtain a parsimonious model, we used the LASSO approach [21] as a powerful and stable feature selection algorithm to automatically select significant features by shrinking the coefficient of unimportant features to zero. This is done by estimating the parameters of the logistic regression via minimizing the negative log-likelihood with an $l_1$- regularization defined as

$$\arg min_\beta \left(-\frac{1}{n}\sum_{i=1}^{n}\{y_i(\beta_0 + \beta^T x_i) - \log(1 + e^{\beta_0 + \beta^T x_i})\} + \lambda \sum_{j=1}^{p}|\beta_j|\right), \qquad \text{Eq. II.3}$$

where $\lambda$ is the penalty parameter that is selected in a way to minimize the out of sample error (generalized error) of the model. For each value of $\lambda$, the above optimization problem is solved, and the optimal value of $\lambda$ is tuned using the cross-validation [32]. As the objective of this study was to balance the accuracy and computational cost, we selected $\lambda$ such that it provides the smallest number of coefficients with a reasonable accuracy. Therefore, the $\lambda$ that occurred within one standard error of the optimal $\lambda$ was selected as the optimum value. The LASSO penalized logistic regression model was implemented using the Glmnet package of R statistical software version 3.5.1.

We applied this method to the dataset of entire non-OSA and moderate/severe-OSA participants (164 participants) and also to the data of different training sets to investigate the robustness of the procedure. From the selected features of each dataset, the features with a high correlation (>50%) to other features were excluded, and the remaining were used for classification.

To classify each participant into one of the two groups of non-OSA or moderate/severe-OSA, we used the features selected by the lasso feature selection method as input to another LASSO logistic regression method trained for the classification. Once the $\beta$ coefficients determined, the probability of assigning a participant to either non-OSA or moderate/severe OSA is determined using the *Eq. II.2*. The accuracy, specificity, and sensitivity were calculated using test and training data of each dataset as well as data of entire non-OSA and moderate/severe-OSA participants. Data of the mild-OSA participants with 5≤AHI<15 were also classified at this stage with the same procedure using the models formed based on data of the entire non-OSA and moderate/severe-

OSA groups as well as data of each training dataset. Figure 8 shows the flowchart of this method for feature selection and classification.
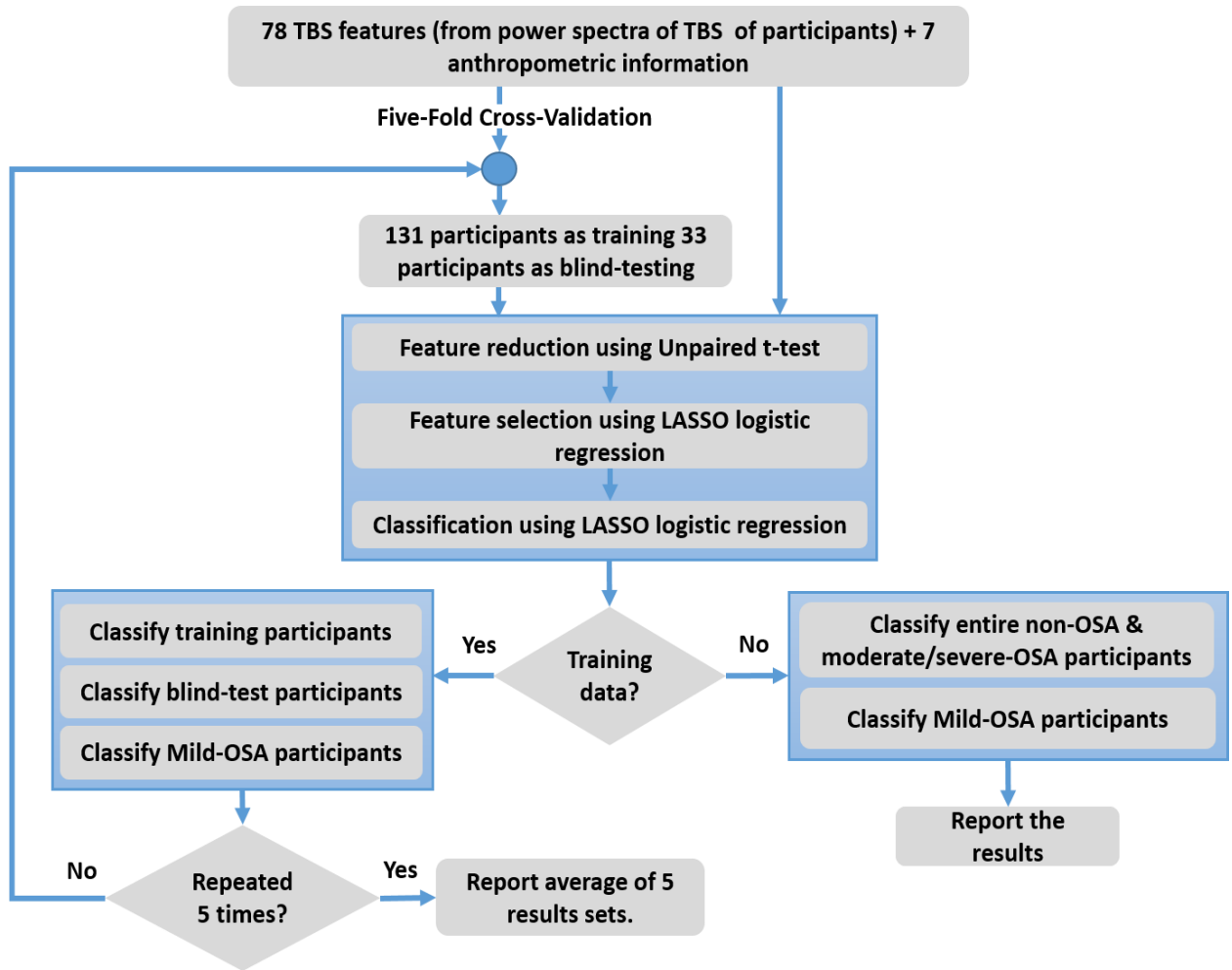


Figure 8 The flowchart of feature selection and classification methods, using the LASSO logistic regression.

The total computational cost of our proposed algorithm was also calculated; it is detailed in Appendix II. B.

## II.3 Results

In this study, features were mainly extracted from the areas with different slopes and the non-overlapped regions of the average power spectra of non-OSA and moderate/severe-OSA participants. As these regions were slightly changed among different training sets, for robustness, we chose the common area among the five folds training datasets. Our feature selection procedure resulted in the following five features as described in *Eq. II.4- Eq. II.7*; F represents a feature.

Table 2 presents the features that were selected using data of each training set.

$F1$: $neck\ circumference$

$F2$: $slope(10log10(PSD_{mouth-Ins} + PSD_{mouth-exp}(260:440\ Hz)))$      *Eq. II.4*

$F3$: $mean(10log10(PSD_{mouth-Ins} + PSD_{mouth-exp}(1100:1350\ Hz)))$      *Eq. II.5*

$F4$: $slope((\ PSD_{mouth-exp}(250:350\ Hz)))$      *Eq. II.6*

$F5$: $mean(10log10(PSD_{nose-Ins} + PSD_{nose-exp}(1000:1300\ Hz)))$      *Eq. II.7*

Table 2 Selected features for different training sets and the dataset containing entire non-OSA and moderate/severe-OSA participants, captured using the regularized logistic regression with the LASSO penalty feature selection method. $PSD_{ME}$ is power spectrum density of mouth expiratory sounds; $PSD_{MI}$ is power spectrum density of mouth inspiratory sounds; $PSD_{NE}$ is power spectrum density of nose expiratory sounds; $PSD_{NI}$ is power spectrum density of nose inspiratory sounds

| Dataset | F1 | F2 | F3 | F4 | F5 |
|---|---|---|---|---|---|
| Training set 1 | * | * | * | | |
| Training set 2 | * | * | | | * |
| Training set 3 | * | * | * | | |
| Training set 4 | * | * | | | * |

| | F1 | F2 | F3 | F4 | F5 |
|---|---|---|---|---|---|
| **Training set 5** | * | | | * | * |
| **Entire non-OSA and moderate/severe-OSA** | * | * | * | | |
| F1: Neck circumference<br>F2: $slope\ (10\ log10(PSD_{MI} + PSD_{ME}[260{:}440\ Hz]))$<br>F3: $mean\ (10\ log10(PSD_{MI} + PSD_{ME}[1100{:}1350\ Hz]))$<br>F4: $slope\ (PSD_{ME}\ [250{:}350\ Hz])$<br>F5: $mean(10log10(PSD_{NI} + PSD_{NE}[1000{:}1300\ Hz]))$ | | | | | |

Table 3 presents the coefficients of each feature for different training sets and also the entire dataset of non-OSA and moderate/severe-OSA participants.

Table 3 Coefficients of selected features of different training sets and the dataset containing entire non-OSA and moderate/severe-OSA participants, captured using the regularized logistic regression with the LASSO penalty feature selection method. $\beta_0$: Intercept; $\beta_1$: coefficient of the neck circumference; $\beta_2$: coefficient of the feature related to the slope of the power spectrum around second formant; $\beta_3$: coefficient of the feature related to the average power spectrum in the high frequency range.

| Dataset | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_3$ |
|---|---|---|---|---|
| **Training set 1** | -6.5 | 0.2 | 4.8 | 0.05 |
| **Training set 2** | -7.2 | 0.2 | 10.5 | 0.05 |
| **Training set 3** | -7.5 | 0.2 | 11.6 | 0.03 |
| **Training set 4** | -4.5 | 0.1 | 11.6 | 0.08 |
| **Training set 5** | -6.8 | 0.2 | 10.2 | 0.08 |
| **Entire non-OSA and moderate/severe-OSA** | **-7.1** | **0.2** | **10.2** | **0.04** |

Figure 9 shows the correlation between the AHI and the selected features for separating non-OSA and moderate/severe-OSA participants within one of the training sets as an example.
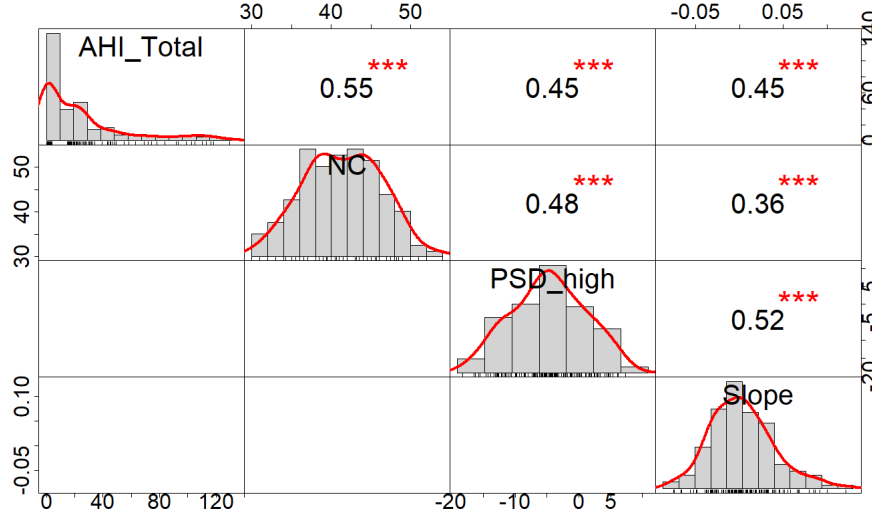
Figure 9 Visualization of the correlation between AHI and the selected features of one of the training datasets, using LASSO feature selection method. On top, the values of the correlation between features plus their significance level as star (***: p<0.001). The overall distribution of features showed as histograms with a fitted curve. The selected features are neck circumference, the average of the power spectrum calculated from summation of mouth inspiratory and expiratory breathing sounds in the high frequency range of (1100-1350 Hz), and the slope of the power spectrum calculated from summation of mouth inspiratory and expiratory breathing sounds in the frequency range of (260-440 Hz).

The selected features of different datasets were investigated using the unpaired t-test to check whether they were statistically significantly different from each other. All the p-values were found to be highly significant ($p< 7.9 \times 10^{-6}$). The discriminative power of the selected features between non-OSA and moderate/severe OSA participants within different datasets were also investigated using other sets of unpaired t-tests. Compared to moderate/severe OSA participants, the non-OSA participants showed to have lower NC, lower high-frequency power, and lower slope around the second formant. All the p-values were found statistically significant ($p < 4.6 \times 10^{-6}$).

Different multivariate analysis of variance (MANOVA) tests were run over the selected features of the five different training datasets and also the entire non-OSA and moderate/severe-OSA

dataset to ensure the combination of the selected features in each feature set were (statistically) significantly different between the two groups; all were found to be highly significant ($p < 6.5 \times 10^{-7}$). Therefore, the selected features were used for classification of the two groups within each training sets as well as the entire dataset (all 5 folds' data).

Table 4 provides the classification specificity, sensitivity, and accuracy of applying the regularized logistic regression with LASSO penalty over data of different training datasets and their corresponding blind testing sets, as well as data of the entire 5 folds' dataset. The classification of the entire 164 non-OSA and moderate/severe participants showed 81.1% accuracy with sensitivity and specificity of 84.4% and 77.0%, respectively. The average accuracies of the 5 training sets were found to be 82.3%±0.7 and 79.3%±6.1 for training and testing, respectively. The sensitivity (specificity) were 85.0%±2.0 (79.1%±1.0) for training and 82.2%±7.2 (75.8%±9.9) for testing, respectively.

Table 4 Regularized logistic regression classification results for different training sets and their corresponding blind-testing sets, the average of all five datasets and their standard deviations (std), and also for the dataset containing entire non-OSA and moderate/severe-OSA participants, using their corresponding features selected by the LASSO penalized logistic regression feature selection method.

| Dataset | | Specificity (%) | Sensitivity (%) | Accuracy (%) | AUC (%) |
|---|---|---|---|---|---|
| **Training set 1** | Training | 78 | 87.5 | 83.2 | 92.1 |
| | Blind-Testing | 60 | 83.3 | 72.7 | |
| **Training set 2** | Training | 78 | 86.1 | 82.4 | 90 |
| | Blind-Testing | 73.3 | 72.2 | 72.7 | |
| **Training set 3** | Training | 80 | 82 | 81.1 | 87.3 |
| | Blind-Testing | 85.7 | 77.8 | 81.3 | |

| | | | | | |
|---|---|---|---|---|---|
| **Training set 4** | Training | 79.7 | 84.7 | 82.4 | 89.4 |
| | Blind-Testing | 80 | 88.8 | 84.8 | |
| **Training set 5** | Training | 79.7 | 84.7 | 82.4 | 90.5 |
| | Blind-Testing | 80 | 88.9 | 84.8 | |
| **Average** | **Training±std** | **79.1±1.0** | **85±2.0** | **82.3±0.7** | **89.9±1.7** |
| | **Blind-Testing±std** | **75.8±9.9** | **82.2±7.2** | **79.3±6.1** | |
| **Entire non-OSA and moderate/severe-OSA** | **-** | **77.0** | **84.4** | **81.1** | **89.02** |

Table 4 also reports the area under the curve (AUC) values associated with the average receiver operating characteristic (ROC) curves of the regularized logistic regression classifier for classification of non-OSA and moderate/severe-OSA participants in different datasets, using their selected features. The AUC value associated with the ROC curve calculated using data of all non-OSA and moderate/severe-OSA participants was shown to be 89.02. In addition, the average of the AUC values between the 5 training datasets was found to be 89.9±1.7. Figure 10 shows the ROC curve calculated using the non-OSA and moderate/severe-OSA participants for one of the training sets.
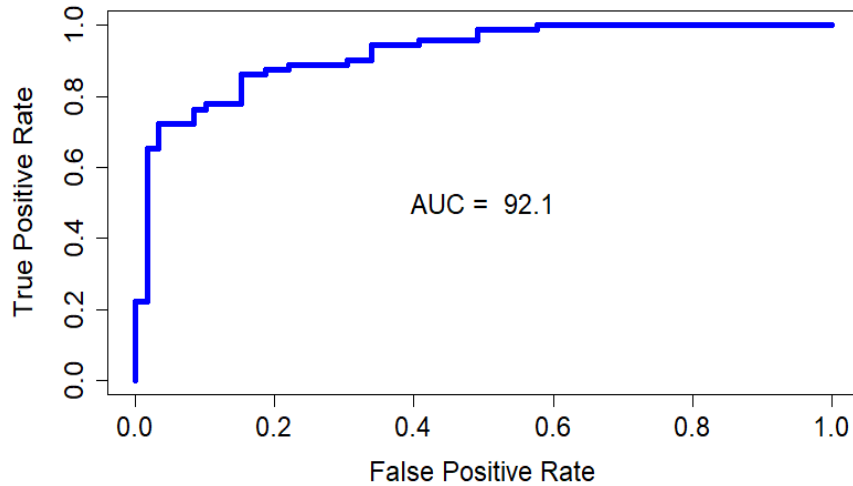
Figure 10 ROC plot of the regularized logistic regression classification over selected features of one of the training datasets, found using the LASSO feature selection method. The selected features are neck circumference, the average of the power spectrum calculated from summation of mouth inspiratory and expiratory breathing sounds in the high frequency range of (1100-1350 Hz), and the slope of the power spectrum calculated from summation of mouth inspiratory and expiratory breathing sounds in the frequency range of (260-440 Hz).

Table 5 shows the number of misclassified participants in each fold of training and testing as well as the entire dataset. Table 6 details the anthropometric information of the misclassified participants when applying the LASSO logistic regression model to the entire dataset.

Table 5 Number of misclassified participants found using the regularized logistic regression classification for different training sets and their corresponding blind-testing sets, the average of all five datasets, and also for the dataset containing entire non-OSA and moderate/severe-OSA participants.

| Dataset | | Total Misclassified | non-OSA Misclassified | moderate/severe-OSA Misclassified |
|---|---|---|---|---|
| Training set 1 | Training | 22 | 13/59 | 9/72 |
| | Blind-Testing | 9 | 6/15 | 3/18 |
| Training set 2 | Training | 23 | 13/59 | 10/72 |

51

| | | | | |
|---|---|---|---|---|
| | Blind-Testing | 9 | 4/15 | 5/18 |
| **Training set 3** | Training | 25 | 12/60 | 13/72 |
| | Blind-Testing | 6 | 2/14 | 4/18 |
| **Training set 4** | Training | 23 | 12/59 | 11/72 |
| | Blind-Testing | 5 | 3/15 | 2/18 |
| **Training set 5** | Training | 24 | 13/59 | 11/72 |
| | Blind-Testing | 4 | 3/15 | 1/18 |
| **average** | **Training** | **24** | **13** | **11** |
| | **Blind-Testing** | **11** | **4** | **7** |
| **Entire non-OSA and moderate/severe OSA** | | **31** | **17/74** | **14/90** |

Table 6 Anthropometric information's mean and their corresponding standard deviations (std) for the misclassified non-OSA and moderate/severe-OSA participants of the dataset containing entire non-OSA and moderate/severe OSA participants, found using the regularized logistic regression with LASSO penalty classifier. AHI is apnea/hypopnea index; NC is neck circumference; BMI is body mass index; MP is Mallampati-score and the numbers in that row show the number of subjects with MP level I, II, II and IV.

| | **non-OSA (AHI<5, n=74)** | **Moderate/ severe-OSA (AHI≥15, n=90)** |
|---|---|---|
| **AHI**±std | 1.1±1.1 | 29.5±20.2 |
| **NC** ($cm$) ±std | 43.6±3.1 | 39.04±2.1 |
| **Age** (year) ±std | 46.6±13.4 | 57.1±10.2 |
| **BMI** ($kg/m^2$) ±std | 35.3±7.3 | 33.6±4.7 |
| **Sex** (Male: Female) | 13:4 | 5:9 |
| **Height** ($cm$) ±std | 170.9±9.4 | 164.3±12.1 |
| **Weight** ($kg$) ±std | 105.7±21.7 | 88.3±15.5 |

| **MP** (I-II-III-IV) | 8-6-1-2 | 5-7-1-1 |

The computational cost for each step of our proposed method is detailed in Appendix II. B. The overall computational cost of our proposed approach is shown in *Eq. II.8*.

$$16n \times O\big(Nlog(L)\big) + p\, O(n) + O(pn) + O(3n) \approx 16n \times O\big(Nlog(L)\big) + O(pn) \quad \textit{Eq. II.8}$$

where $n$ is the number of participants, $N$ is the length of the TBS after normalization, $L$ is the window length, and $p$ is the number of features. The computational cost of our proposed method is linear in terms of the number of features and the number of participants; hence, the method is very fast.

## II.4  Discussion

In this study, we proposed to use a regularized logistic modeling approach as a simple and fast method for analyzing TBS recorded during daytime when the subjects were awake (wakefulness) with the purpose of predicting their OSA severity by a few minutes breathing sound recording and data analysis. Our proposed approach is discriminative in nature and does not rely on strong model-based assumptions. We developed a robust feature selecting approach and used it in classifying subjects with AHI<5 and AHI≥15. The first set of features considered for this study (p=85) were a combination of the TBS features and the anthropometric information of participants. As there were two breathing maneuvers and we investigated respiratory phases both separately and in combination, the number of features was large. Thus, we used a feature selection approach based on regularized logistic regression methodology with a LASSO penalty to select the best subset of features.

In LASSO feature selection method, shrinking and removing some of the model's coefficients lead to reducing the variance without a considerable increase in bias; hence, it provides desirable prediction accuracy [32]. This is useful when the number of participants is not much larger than the number of features, as it was the case in our study. The LASSO method is very fast and provides an interpretable model by weighting the selected features. Therefore, it is easy to explain which features are more important for predicting the true class labels; the coefficients of the selected features show their importance ranking (Table 3).

As the LASSO method uses a cross-validation procedure to automatically estimate the optimal value of $\lambda$ (the penalty parameter), there is no need to perform another cross-validation procedure for obtaining the training and testing results. However, we generated five sets of randomly selected training and blind testing data to ensure the unbiased test classification. Our findings (Table 2 and Table 3) show that in different training sets the selected features and their corresponding coefficient values were different; however, in all of the training sets and also in the entire 5 folds' dataset three selected feature types were common. Those features were: 1) the neck circumference, 2) the slope of the average PSD around the second formant, and 3) the average of the high-frequency portion of the average PSD. These features are congruent with the findings of our previous study on a separate dataset [28]. In all training and testing datasets of this study, the unpaired t-test showed that the selected features were statistically significantly different from each other and also different between the two groups of non-OSA and moderate/severe OSA. These features also showed high correlations with AHI. In addition, the ROC curves of the different LASSO logistic regression classifiers over different training sets and also on the entire dataset showed high performance with high values of AUC (Table 4). These high AUCs indicate the robustness of the LASSO logistic regression classifiers with respect to our selected features.

Appendix II. A reports the results of using a simple logistic regression for classification as opposed to the LASSO logistic regression method. The simple logistic regression classification of the entire 164 non-OSA and moderate/severe participants showed 80.5% accuracy with sensitivity (specificity) of 82.2% (78.4%), which are lower than those of LASSO logistic regression classification (Table 4). The reason that LASSO logistic regression classification resulted in better than the simple logistic regression is that regularization term of LASSO logistic regression classification method ($\lambda$) improves the generalization performance (performance on new data) by reducing the variance of parameter estimates (shrinks the coefficients that contribute most to the error).

It is also important to investigate the characteristics of misclassified participants. For the dataset containing all non-OSA and moderate/severe OSA participants, the overall accuracy of the LASSO logistic regression was 81.1%; out of 164 subjects, 31 were misclassified, out of which, 17 were from non-OSA and 14 were from moderate/severe OSA group (Table 5). Further investigation on the anthropometric information of misclassified subjects (Table 6) and comparing them with the non-OSA and moderate/severe participants of this study, revealed that overall, the non-OSA misclassified individuals are characterized with higher NC, BMI, height, weight and being male, while the moderate/severe OSA misclassified individuals are characterized with lower NC, BMI, height, weight and being female. This finding is not surprising, as it is known that these anthropometric parameters affect the breathing sounds while being correlated with AHI [33]. Thus, ideally, we should have the two non-OSA and moderate/severe OSA groups of the training and testing matched in these characteristics. However, that requires a much larger dataset.

In this study, similar to all other similar studies, we used the AHI values as the true label of each individual to be either in non-OSA or moderate/severe OSA group. Although AHI is one of

the most important and commonly used indicators for diagnosis of OSA, there are many individuals with similar AHI but different daytime symptoms and different levels of OSA severity. The clinical diagnosis of OSA is not only based on AHI; it is based on a combination of frequency of arousals, daytime symptoms, lack of deep sleep, etc. [34]. We speculate that the TBS features are indeed good representatives of the OSA severity as they are affected by the pathophysiology of OSA. However, the UA pathology of two individuals with similar AHI (e.g. AHI of 14 and 16) may not be that much different to affect the TBS differently enough to be detected by sound analysis. For this reason, to avoid the misclassification in the borderline ranges with artificially crisp nature, we allowed a small gap ($5 \leq AHI < 15$) in continuous AHI values of the non-OSA and moderate/severe OSA groups to form the two groups. Once the classifiers were trained, the participants with AHI in this gap range were classified using our proposed method. Using the model obtained based on data of 164 non-OSA and moderate/severe participants, 11 out of 35 individuals in the gap range were assigned to the non-OSA group and the remaining 24 individuals assigned to the moderate/severe group. Further investigation of these participants revealed that the individuals assigned to the moderate/severe group had higher NC, higher slope around the second formant and higher average PSD in high-frequency ranges compared to those assigned to the non-OSA group, while there was not much difference in their AHI values ($8.8 \pm 0.6$ and $8.4 \pm 0.8$, respectively). This finding confirms the shortcoming of using AHI as the only indicator for OSA severity. Finding a better OSA severity indicator still remains a challenge. Perhaps a better indicator would be a combination of several factors including AHI, duration of apneic events, number of arousals, etc. and also the breathing sounds features.

Last but not the least advantage of the proposed method is that the most repeated features among the selected features of all datasets were extracted from the combination of inspiratory and

expiratory TBS. This finding implies that there may be no need to separate the breathing phases in the pre-processing step. Although there are various methods for automatic separation of breathing phases into inspiration and expiration [35], [36], however, none of them are 100% accurate. Therefore, there is always a need for manual investigation of breathing sounds. Eliminating this part from the pre-processing procedure will speed up the data analysis significantly; hence, providing a faster on-line screening tool for OSA.

## II.4.1 Physiological interpretation of the selected features

It is worth mentioning that selected features using the regularized logistic regression approach are physiologically meaningful and highly correlated with AHI (Figure 9).

Obesity is a risk factor in patients with sleep apnea [37], [38]. The relationship between obesity and OSA has been demonstrated to be totally explained by variation in neck size [39], [40]. It was suggested that the increases in neck circumference and fat deposited around the UA might narrow the UA [38]. Therefore, we can conclude that NC, a surrogate measure of neck fat, can be considered as a predictor of the presence of OSA.

A highly collapsible/compliant pharyngeal airway is known as the main anatomical abnormality in the UA of OSA patients [41]. A reason for increased collapsibility of OSA subjects compared to that of controls is their smaller cross-sectional area [42], [43]. According to the Helmholtz equation [44], an UA with a smaller cross-sectional area reflects a lower resonant frequency ($f$)

$$f = \frac{c}{2\pi}\sqrt{\frac{A}{VL}}$$

*Eq. II.9*

where $c$ is the speed of sound, $A$ is cross-section area, $V$ is the cavity volume and $L$ is the length of the bottleneck. The resonant frequencies of the UA are represented by formant frequencies, the

peaks in the spectrum of sound signals. Hence, we expect to see lower formant frequencies for moderate/severe-OSA participants. This is congruent with what is observed in Figure 7 especially for the second formant (F2). In our study, the higher slope in PSD of the moderate/severe-OSA participants' around their F2 implied that their F2 occurred in lower frequencies compared to that of non-OSA participants.

On the other hand, stiffness is the property of a tissue that resists the change in shape in response to an applied force. It has been shown that during wakefulness, the stiffness of the tongue muscles and soft palate tissues of OSA patients is higher compared to those of controls [45]. The stiffening of the narrower portion of the UA of OSA individuals was also supported by showing a predominantly bottle shape for their narrowed UA [20]. As an increase in stiffness is presented by an increase in high-frequency components of the breathing sounds, we expected to see higher energy in the high-frequency portion of the average PSD of moderate/severe-OSA participants compared to that of non-OSA; this is what is observed in Figure 7.

## II.5 Conclusion

In this study, we investigated the application of regularized logistic regression model using a LASSO penalty for prediction of OSA severity during wakefulness using TBS analysis and anthropometric information. The selected features by this regression modeling approach are congruent with those selected by other methods in our previous studies; they are physiologically meaningful and highly correlated with AHI. The unbiased blind-testing classification accuracy, sensitivity, and specificity over the non-OSA and moderate/severe OSA groups of five different folds were found to be 79.3%±6.1, 82.2%±7.2 and 75.8%±9.9, respectively. The accuracy for the entire dataset was 81.1% with sensitivity and specificity of 84.4% and 77.0%, respectively. Although results of this study are similar but not necessarily superior to previous studies' results,

they are important from different aspects. The most important contributions of this work are: 1) the method presented in this work does not need respiratory phase identification as the best features do not depend on that; 2) the method is simple, quick and computationally more effective than SVM and other previous methods for OSA screening during wakefulness; thus, suitable for online applications; 3) the method can be used for screening OSA individuals with severity as low as 15 with high sensitivity and a reasonable specificity. The main limitation of this study is the lack of a large sample size to have a better way to handle OSA severity prediction of individuals with $5 < AHI < 15$.

## Appendix II. A Classification using simple logistic regression

In this study, as we had 2 classes of participants (non-OSA and moderate/severe OSA). Thus, we also applied the binomial logistic regression classifier to the features selected by the LASSO penalized logistic regression for classification purpose. The classification of the entire 164 non-OSA and moderate/severe participants showed 80.5% accuracy with sensitivity and specificity of 82.2% and 78.4%, respectively. The average of the results between the 5 training sets showed accuracies of 82.0%±1.8 and 78.1%±7.7 for training and testing, respectively. The sensitivity (specificity) was 83.0%±3.1 (80.8%±2.2) for training and 78.9%±11.4 (77.1%±9.9) for testing, respectively. The AUC associated with ROC curve calculated using all non-OSA and moderate/severe participants was shown to be 80.3. In addition, the average of the AUC values between the 5 training-sets was shown to be 81.9±1.7. For the dataset containing all 164 non-OSA and moderate/severe OSA participants, 32 were misclassified, out of which, 16 were from non-OSA and 16 were from moderate/severe OSA group. In addition, 17 out of 35 subjects in the gap range were assigned to the non-OSA group and the remaining 18 subjects to the moderate/severe group.

The reason for achieving the better result for regularized logistic regression classification than the simple logistic regression is that as there is still some correlation between the selected features, the regularization term can improve the generalization performance (performance on new data) by reducing the variance of parameter estimates.

## Appendix II. B Computational complexities

The computation cost of our proposed algorithm is relatively low. For the feature extraction part, the computational cost was estimated for a typical TBS with length $N$ after normalization. The feature extraction phase consisted of 2 main parts: LogVar estimation, and PSD estimation. The LogVar and PSD calculated for each segment, defined as the window with length $L$ which moves along the signal with 50% overlap. Thus, the total number of overlapping windows within each TBS is approximately $\frac{2N}{L}$. The computational cost of calculating LogVar within one segment is $L$; thus, the total number of operations to calculate the LogVar of each TBS is

$$L \times \frac{2N}{L} = O(N). \qquad \text{Eq. II.10}$$

For the PSD estimation, the fast Fourier transform (FFT) of each segment was calculated and averaged. The computational cost of calculating FFT of each segment is $L \log (L)$, thus the total cost of PSD estimation for each TBS is

$$O(\frac{2N}{L} \times L \log (L)) = O(Nlog(L)). \qquad \text{Eq. II.11}$$

The feature extraction part is from $O(k)$, where $K$ is the number of extracted features from each TBS signal. As we have 16 signals for each participant (for mouth/nose breathing maneuvers, inspiratory/expiratory phases, summation/subtraction of phases, and normal/logarithmic scale), therefore, the total cost of feature extraction from $n$ participant is

$$16n \times \left(O(N) + O\big(Nlog(L)\big) + O(k)\right) \approx 16n \times O\big(Nlog(L)\big). \qquad Eq.\ II.12$$

The feature reduction part of this study consists of applying t-test and LASSO. As each t-test has the computational cost of $O(n)$, therefore, the overall t-tests for all $p$ features have the computational cost of $p \times O(n)$. On the other hand, in this study, we ran the LASSO with the Glmnet package of R, which uses the coordinate descent algorithm to find the optimum solution. The computational cost of this method is reported to be $O(pn)$ for each iteration of the optimization [32]. Besides, as 3 features were selected as the best feature set of each training set, thus, the computational cost of logistic regression classification using the LASSO penalty is $O(3n)$. Therefore, the total computational cost of our proposed method can be written as

$$16n \times O\big(Nlog(L)\big) + p\,O(n) + O(pn) + O(3n) \approx 16n \times O\big(Nlog(L)\big) + O(pn) \qquad Eq.\ II.13$$

which is linear in terms of $n, p$.

## References

[1]     A. Malhotra and D. P. White, "Obstructive sleep apnoea," *Lancet*, vol. 360, no. 9328, pp. 237–245, Jul. 2002.

[2]     D. J. Eckert and M. K. Younes, "Arousal from sleep: implications for obstructive sleep apnea pathogenesis and treatment.," *J. Appl. Physiol.*, vol. 116, no. 3, pp. 302–13, 2014.

[3]     Terry Young, M. Palta, Jerome Dempsey, P. E. Peppard, F. J. Nieto, and K. Mae Hla, "Burden of Sleep Apnea: Rationale, Design, and Major Findings of the Wisconsin Sleep Cohort Study," *WMJ Off. Publ. State Med. Soc. Wisconsin*, vol. 108, no. 5, pp. 246–249, Nov. 2010.

[4]     C. V. Senaratna *et al.*, "Prevalence of obstructive sleep apnea in the general population: A systematic review," *Sleep Med. Rev.*, vol. 34, no. October, pp. 70–81, Aug. 2017.

[5]     S. G. Memtsoudis, M. C. Besculides, and M. Mazumdar, "A Rude Awakening — The Perioperative Sleep Apnea Epidemic," *N. Engl. J. Med.*, vol. 368, no. 25, pp. 2352–2353, Jun. 2013.

[6]     American Academy of Sleep Medicine, *Hidden Health Crisis Costing America Billions: Underdiagnosing and Undertreating Obstructive Sleep Apnea Draining Healthcare System*. Mountain View, CA: Sullivan & Frost, 2016.

[7]     C. A. Kushida *et al.*, "Practice Parameters for the Indications for Polysomnography and Related Procedures: An Update for 2005," *Sleep*, vol. 28, no. 4, pp. 499–523, 2005.

[8]     H. Y. Chiu *et al.*, "Diagnostic accuracy of the Berlin questionnaire, STOP-BANG, STOP, and Epworth sleepiness scale in detecting obstructive sleep apnea: A bivariate meta-analysis," *Sleep Med. Rev.*, vol. 36, pp. 57–70, 2017.

[9]     T. Young *et al.*, "Predictors of Sleep-Disordered Breathing in Community-Dwelling Adults<subtitle>The Sleep Heart Health Study</subtitle>," *Arch. Intern. Med.*, vol. 162, no. 8, p. 893, Apr. 2002.

[10]    A. Elwali and Z. Moussavi, "Obstructive Sleep Apnea Screening and Airway Structure Characterization During Wakefulness Using Tracheal Breathing Sounds," *Ann. Biomed. Eng.*, vol. 45, no. 3, pp. 839–850, Mar. 2017.

[11]    A. Montazeri, E. Giannouli, and Z. Moussavi, "Assessment of Obstructive Sleep Apnea and its Severity during Wakefulness," *Ann. Biomed. Eng.*, vol. 40, no. 4, pp. 916–924, Apr. 2012.

[12]    J. Sola-Soler, J. A. Fiz, A. Torres, and R. Jane, "Identification of Obstructive Sleep Apnea patients from tracheal breath sound analysis during wakefulness in polysomnographic studies," *Conf. Proc. ... Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2014, pp. 4232–4235, 2014.

[13]    A. Yadollahi, A. Montazeri, A. Azarbarzin, and Z. Moussavi, "Respiratory Flow–Sound Relationship During Both Wakefulness and Sleep and Its Variation in Relation to Sleep Apnea," *Ann. Biomed. Eng.*, vol. 41, no. 3, pp. 537–546, Mar. 2013.

[14]    A. Yadollahi and Z. M. K. Moussavi, "Acoustical flow estimation: Review and validation," *IEEE Eng. Med. Biol. Mag.*, vol. 26, no. 1, pp. 56–61, Jan. 2007.

[15]    Y. Wang, J. P. Mcdonald, Y. Liu, K. Pan, X. Zhang, and R. Hu, "Dynamic alterations of the tongue in obstructive sleep apnea-hypopnea syndrome during sleep: analysis using ultrafast MRI," *Genet. Mol. Res.*, vol. 13, no. 2, pp. 4552–4563, 2014.

[16]    Z. Wu, W. Chen, M. C. K. Khoo, S. L. Davidson Ward, and K. S. Nayak, "Evaluation of upper airway collapsibility using real-time MRI," *J. Magn. Reson. Imaging*, vol. 44, no. 1, pp. 158–167, 2016.

[17]    G. C. Barkdull, C. a Kohl, M. Patel, and T. M. Davidson, "Computed Tomography Imaging of Patients With Obstructive Sleep Apnea," *Laryngoscope*, vol. 118, no. 8, pp. 1486–1492, Aug. 2008.

[18]    K.-H. Liu *et al.*, "Sonographic measurement of lateral parapharyngeal wall thickness in patients with obstructive sleep apnea.," *J. Sleep*, vol. 30, no. 11, pp. 1503–8, 2007.

[19]    R. J. Schwab *et al.*, "Identification of Upper Airway Anatomic Risk Factors for Obstructive Sleep Apnea with Volumetric Magnetic Resonance Imaging," *Am. J. Respir. Crit. Care Med.*, vol. 168, no. 5, pp. 522–530, Sep. 2003.

[20]    Y. Finkelstein *et al.*, "Velopharyngeal Anatomy in Patients With Obstructive Sleep Apnea Versus Normal Subjects," *J. Oral Maxillofac. Surg.*, vol. 72, no. 7, pp. 1350–1372, Jul.

2014.

[21] R. Tibshirani, "Regression shrinkage and selection via the lasso: a retrospective," *J. R. Stat. Soc. Ser. B (Statistical Methodol.*, vol. 73, no. 3, pp. 273–282, Jun. 2011.

[22] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learnin*. 2009.

[23] B. A. Goldstein, A. M. Navar, and R. E. Carter, "Moving beyond regression techniques in cardiovascular risk prediction: Applying machine learning to address analytic challenges," *Eur. Heart J.*, vol. 38, no. 23, pp. 1805–1814, 2017.

[24] Y. Wen *et al.*, "Sleep duration, daytime napping, markers of obstructive sleep apnea and stroke in a population of southern China," *Sci. Rep.*, vol. 6, no. October, pp. 1–9, 2016.

[25] A. Oulhaj, S. Al Dhaheri, B. Bin Su, and M. Al-Houqani, "Discriminating between positional and non-positional obstructive sleep apnea using some clinical characteristics," *Sleep Breath.*, vol. 21, no. 4, pp. 877–884, 2017.

[26] F. Arvin, S. Doraisamy, and E. Safar Khorasani, "Frequency shifting approach towards textual transcription of heartbeat sounds," *Biol. Proced. Online*, vol. 13, no. 1, pp. 1–7, 2011.

[27] H. Pasterkamp, S. S. Kraman, and G. R. Wodicka, "Respiratory sounds, advances beyond the stethoscope," *Am. J. Respir. Crit. Care Med.*, vol. 156, no. 3, pp. 974–987, 1997.

[28] F. Hajipour and Z. Moussavi, "Spectral and Higher Order Statistical Characteristics of Expiratory Tracheal Breathing Sounds During Wakefulness and Sleep in People with Different Levels of Obstructive Sleep Apnea," *J. Med. Biol. Eng.*, vol. 39, no. 2, pp. 244–250, Apr. 2019.

[29] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing*. 1996.

[30] R. J. SIMES, "An improved Bonferroni procedure for multiple tests of significance," *Biometrika*, vol. 73, no. 3, pp. 751–754, 1986.

[31] Z. Šidák, "On probabilities of rectangles in multivariate Student distributions: their dependence on correlations," *Ann. Math. Stat.*, pp. 169–175, 1971.

[32] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical Learning with Sparsity: The Lasso and Generalizations*. 2015.

[33] I. Sanchez and H. Pasterkamp, "Tracheal sound spectra depend on body height.," *Am. Rev. Respir. Dis.*, vol. 148, no. 4 Pt 1, pp. 1083–1087, 1993.

[34] L. J. Epstein *et al.*, "Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults.," *J. Clin. Sleep Med.*, vol. 5, no. 3, pp. 263–76, 2009.

[35] H. Alshaer, A. Pandya, T. D. Bradley, and F. Rudzicz, "Subject independent identification of breath sounds components using multiple classifiers," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, pp. 3577–3581, 2014.

[36] S. Huq and Z. Moussavi, "Acoustic breath-phase detection using tracheal breath sounds," *Med. Biol. Eng. Comput.*, vol. 50, no. 3, pp. 297–308, Mar. 2012.

[37] T. Young, J. Skatrud, and P. E. Peppard, "Risk Factors for Obstructive Sleep Apnea," vol. 291, no. 16, pp. 2013–2016, 2004.

[38] A. R. Schwartz, S. P. Patil, A. M. Laffan, V. Polotsky, H. Schneider, and P. L. Smith, "Obesity and Obstructive Sleep Apnea: Pathogenic Mechanisms and Therapeutic Approaches," *Proc. Am. Thorac. Soc.*, vol. 5, no. 2, pp. 185–192, 2008.

[39] R. J. Davies and J. R. Stradling, "The relationship between neck circumference, radiographic pharyngeal anatomy, and the obstructive sleep apnoea syndrome.," *Eur. Respir. J. Off. J. Eur. Soc. Clin. Respir. Physiol.*, vol. 3, no. 5, pp. 509–514, 1990.

[40] R. J. Davies, N. J. Ali, and J. R. Stradling, "Neck circumference and other clinical features in the diagnosis of the obstructive sleep apnoea syndrome.," *Thorax*, vol. 47, no. 2, pp. 101–105, 1992.

[41] D. J. Eckert, D. P. White, A. S. Jordan, A. Malhotra, and A. Wellman, "Defining phenotypic causes of obstructive sleep apnea: Identification of novel therapeutic targets," *Am. J. Respir. Crit. Care Med.*, vol. 188, no. 8, pp. 996–1004, 2013.

[42] R. J. Schwab, K. B. Gupta, W. B. Gefter, L. J. Metzger, E. A. Hoffman, and A. I. Pack, "Upper airway and soft tissue anatomy in normal subjects and patients with sleep-disordered breathing. Significance of the lateral pharyngeal walls.," *Am. J. Respir. Crit. Care Med.*, vol. 152, no. 5, pp. 1673–1689, 1995.

[43] E. F. Haponik, P. L. Smith, M. E. Bohlman, R. P. Allen, S. M. Goldman, and E. R. Bleecker, "Computerized tomography in obstructive sleep apnea. Correlation of airway size with physiology during sleep and wakefulness.," *Am. Rev. Respir. Dis.*, vol. 127, no. 2, pp. 221–6, 1983.

[44] H. von Helmholtz, "Theorie der Luftschwingungen in Röhren mit offenen Enden.," *J. für die reine und Angew. Math.*, vol. 57, pp. 1–72, 1860.

[45] Veldi, Vasar, Vain, Hion, and Kull, "Computerized endopharyngeal myotonometry (CEM): A new method to evaluate the tissue tone of the soft palate in patients with obstructive sleep apnoea syndrome," *J. Sleep Res.*, vol. 9, no. 3, pp. 279–284, Sep. 2000.

# Chapter III. A Comparison of Two Machine Learning Models for Daytime Diagnosis of OSA

## A comparison of Regularized Logistic Regression and Random Forest for Daytime Diagnosis of Obstructive Sleep Apnea

Farahnaz Hajipour, Mohammad Jafari Jozani, and Zahra Moussavi

*Abstract* — A major challenge in big and high-dimensional data analysis is related to the classification and prediction of the variables of interest by characterizing the relationships between the characteristic factors and predictors. This study aims to assess the utility of two important machine-learning techniques to classify subjects with obstructive sleep apnea (OSA) using their daytime tracheal breathing sounds. We evaluate and compare the performance of the Random Forest (RF) and Regularized Logistic Regression (LR) as feature selection tools and classification approaches for wakefulness OSA screening. Results show that RF, which is a low variance committee-based approach, outperforms the regularized LR in terms of blind-testing accuracy, specificity, and sensitivity with 3.5%, 2.4%, and 3.7% improvement, respectively. However, the regularized LR was found to be faster than the RF and resulted in a more parsimonious model. Consequently, both RF and regularized LR feature reduction and classification approaches are qualified to be applied for the daytime OSA screening studies, depending on the nature of data and applications' purposes.

## III.1 Introduction

Nowadays the world is in the "Big Data" era, as most available data are stored [1]. For example, in medical fields, the stored data includes patients' personal, family and demographic information, history of their diseases, their various medical tests, etc. Big Data analysis requires a fair knowledge of the data being processed and proper use of intelligent algorithms to extract appropriate knowledge from the data regarding the relationships between predictors and variables of interest and perform classification and prediction.

When dealing with large and high-dimensional datasets, it is possible to extract a considerable number of features from data. To build parsimonious models that are easy to understand, variable selection can be performed by removing variables that are redundant and do not add any information to our model [2]. This is especially useful in scientific fields such as medicine or biological signal analysis, where the focus is not only on prediction but also on meaningful physiological interpretation of the outcomes.

In many studies, the manual selection of features is inappropriate, and one should rely on automatic feature selection and classification techniques [3]. As a wide variety of such methods exists [4], it is often prudent to use and compare different approaches. The best feature selection procedure depends on the underlying problem and the trade-off between interpretability and the complexity of the desired model [4]. Regression- and tree-based models are two important families of machine-learning methods that can be used for feature selection and classification purposes. The least absolute shrinkage and selection operator (LASSO) logistic regression (LR) or simply LASSO regularized LR [4], [5] is a popular regression-based method that has been used in our previous study [6]. In this method, a penalty is placed on the summation of the estimated regression coefficients to shrink a subset of coefficients toward, and possibly set some to zero. This shrinkage

reduces the variance (at the cost of an often small increase in bias) and results in a more stable model with better prediction accuracy [5].

Random-Forest (RF) is a popular tree-based method [7]. The idea behind RF is to randomly create a committee of low-bias decision trees to explain the data. In each tree, among the random subset of available features, a feature that best separates the data into two discriminant groups is selected as a node. The node selection procedure is repeated in the new subgroups until data of all groups are of the same kind or some impurity thresholds are met, where impurity measure is a metric to evaluate the goodness of any node [7]. Features that contributed the most in purifying the nodes could be used as selected variables for further analysis. To eliminate the large variance problem of tree-based approaches, the RF method uses the bagging and random variable selection, which results in a low correlation among individual trees [2], [7]. From averaging over a large ensemble of low-bias, high-variance but low-correlated trees, this algorithm yields an ensemble that can achieve both low bias and low variance.

The LASSO regularized LR and RF have been used and compared in various fields, including cardiovascular risk prediction[8], forecasting electricity prices [9], and predicting case duration for robot-assisted surgery [10]. In this study, we compare these two approaches as automatic feature selection and classification techniques for analysis of breathing sounds data collected during daytime (wakefulness) to predict the likelihood of obstructive sleep apnea (OSA) and its severity.

OSA is characterized by repetitive episodes of the complete or partial collapse of the upper airway during sleep [11]. The most popular method for assessing the severity of OSA is to measure the Apnea-Hypopnea Index (AHI), which reports the average number of apnea and hypopnea events per hour of sleep [12]. The gold standard approach in the diagnosis of OSA is the full-

nocturnal in-laboratory polysomnography (PSG) assessment. Although PSG is a comprehensive and reliable means for diagnosing OSA, it requires the technician's expertise and is expensive and time-consuming. It has been reported that the timely access of PSG leads to undiagnosed patients (80%) at the time of surgery [13].

OSA is a major health problem, and lack of its diagnosis and treatment could impose a substantial impact on public health and health care costs [14]. Recently, there has been a growing interest in OSA research in order to find an accurate, fast, simple, and less expensive approach compared to the PSG for early OSA screening and diagnosis. One strategy is using tracheal breathing sounds (TBS) analysis during wakefulness. TBS are generated by the passage of turbulent airflow from the upper airway [15]. The structural and physiological properties of the upper airway affect the generated sounds [16], [17]. Studies have been shown that TBS is rich in features that could distinguish individuals with OSA from non-OSA individuals during wakefulness without a sleep study [6], [18]–[20]. This strategy is especially useful for the online screening of OSA in patients at the time of surgery to reduce their pre/postoperative complication risks.

In our analysis, we used the values of the features extracted from TBS recorded during wakefulness (the subjects were awake). We fitted the LASSO regularized LR and developed the RF model over the large set of candidate variables. Then, we compared these models with respect to the selected features and any noticeable differences in their classification results. The comparative analysis in this study allows us to get insight into the performance of these two methods on online screening of OSA. Even though the data of this study may not be sufficiently large to be considered as 'Big Data', however, it could provide an insight into the feature selection of Big Data analytics.

## III.2 Method

### III.2.1 Data

The dataset for this study is adopted from our previous research [6] and consists of observations taken from 199 individuals that were referred to Misericordia Health Centre (Winnipeg, MB, Canada) for a nocturnal PSG assessment. The inclusion criteria were: 1) age between 18 to 70 years; 2) suspected of OSA and referred to full PSG study by a doctor. All participants did not have any cold or any other respiratory illness, as well as insomnia and restless leg at the time of recording. The study was approved by the Biomedical Research Ethics Board of the University of Manitoba. All participants signed informed consent before data collection.

When the participants were awake and lied in the supine position, they were asked to breathe five times through their mouth followed by five breaths through their nose, with mouth closed. The participants' TBS were recorded using a small Sony ECM-77B microphone embedded in a chamber of 2mm diameter and placed over the suprasternal notch of their trachea. The digital signals were band-pass filtered in the frequency range of [0.05-5000 Hz] and recorded using a sampling rate of 10240 Hz. After the wakefulness TBS recording, each participant underwent an overnight PSG assessment. Based on AHI outcomes of the PSG, we divided our participants into three groups of non-OSA (AHI<5, N=74, 29 males), mild-OSA (5≤AHI<15, N=35, 21 males), and moderate/severe-OSA (AHI≥15, N=90, 66 males). The anthropometric information of the participants is reported in our previous study [6].

### III.2.2 Signal Analysis and Feature Extraction

The preprocessing phase of this study is similar to our previous work [6]. Briefly, considering the marked voice of experimenter at the start of inspiration and by investigating the signals in time-

and-frequency domain, inspiratory and expiratory phases were separated manually, and noisy signals were eliminated. The remaining signals were then band-pass filtered in the range of [75-2500 Hz] and normalized by their variance envelope and energy. Next, the segments corresponding to the 50% duration around the maximum of the logarithm of the sounds' variance signal were considered as their stationary portion [21]. For the stationary part of each respiratory phase (inspiration/ expiration) in each breath of mouth and nasal maneuvers and also for the summation and subtraction of the respiratory phases, we estimated the power spectrum density (PSD-using Welch's method) in windows of 20 ms with 50% overlap between adjacent windows. The PSD signals were calculated in both linear and logarithmic scale and were averaged over the five respiratory phases of the participant's data. Overall, we obtain 16 PSD signals for each participant.

The extracted features are the same as those introduced and elaborated in our former study [6]. In summary, the features were extracted from data of non-OSA and moderate/severe-OSA participants (164 participants). Data of mild-OSA participants were dealt with separately to prevent the misclassification in the crisp-nature borderline ranges [6]. To avoid sampling bias, we randomly divided our participants into five groups, each with 33 participants. Every time data of one group was used as the test and the remaining data (131 participants) was considered as training data for feature extraction. In each training dataset and for each of the pre-mentioned 16 PSD signals, we calculated the average power spectra and their 95% confidence intervals (CI) for the non-OSA and moderate/severe-OSA groups. The slope and average of these power signals in regions with no overlap between the CI's of non-OSA and moderate/severe-OSA groups were considered as features and resulted in 78 characteristics. Combining these features to the anthropometric information of individuals including sex, age, height, weight, body mass index, neck circumference (NC), and Mallampati score led to a total number of 85 features per participant.

70

## III.2.3 Feature Selection and Classification

After feature extraction, feature selection algorithms were used in order to remove redundant and irrelevant features. We first employed one-way analysis of variance (ANOVA) tests on each of the 85 features in every training set and removed variables that were not significantly different between the two OSA groups. We considered a 0.01 level of significance for the overall test. Using a modified version of the Bonferroni approach, on average a total number of 45 features with the p-value $< 1 - (1 - 0.01)^{1/85}$ remained in each dataset [6], [22]. The remaining features were on average 41 acoustic features and sex, weight, BMI, and NC. Since the number of discriminative features was still substantial, as the second step of feature reduction, we evaluated and compared the LASSO regularized LR [4], [5] and RF [7] approaches to find the most appropriate features for classification of two OSA groups within each training dataset. Below we describe each method and explain how they fit into the general template for feature selection and classification.

### III.2.3.1    LASSO regularized LR

The LASSO regularized LR approach has already been used for modeling the OSA severity level of participants in our previous research [6]. Briefly, after optimizing the LASSO optimization equation and finding the LASSO coefficients, among the characteristics with non-zero coefficients, we excluded those with more than 50% correlation to other features and kept the remaining features for classification purpose. Subsequently, within each training dataset, we applied another LASSO regularised LR method on our selected features to find the final model's coefficient vector. When the coefficients were estimated, the probability of assigning a participant to the non-OSA or moderate/severe-OSA groups were determined as follows:

$$Pr(Y = 1 | X_1, ..., X_{p*}) = \frac{e^{\beta_0 + \sum_{j=1}^{p^*} \beta_j X_j}}{1 + e^{\beta_0 + \sum_{j=1}^{p^*} \beta_j X_j}},$$

<div align="right">*Eq. III.1*</div>

where $Y$ is the actual class label ($Y = 0$ for non-OSA and $Y = 1$ for moderate/severe-OSA), $p^*$ is the number of selected features, $X$ is an ($n \times p^*$) matrix associated with $n$ participants and $p^*$ features and $\boldsymbol{\beta}$ denotes the LASSO coefficients corresponding to an intercept and $p^*$ features defined as

$$\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_{p*})^T.$$

<div align="right">*Eq. III.2*</div>

In order to map the LR probability into a binary category {0, 1}, we considered the decision threshold to be 0.5, that is

$$Y = \begin{cases} 1 \text{ if } Pr(Y = 1 | X_1, ..., X_p^*) \geq 0.5 \\ 0 \text{ if } Pr(Y = 1 | X_1, ..., X_p^*) < 0.5 \end{cases}.$$

We reported the accuracy, specificity, and sensitivity associated with training and their corresponding blind-testing sets. We also classified the mild-OSA participants using the predicted models of each training dataset and reported the number of mild-OSA participants assigning to each of the OSA groups. Figure 11-a shows the flowchart of the LASSO regularized LR method for feature selection and classification.

### III.2.3.2 RF

In RF, a large ensemble of classification trees are used to explain the data [7]. RF uses approximately 63% of the data to construct each tree; the remaining ~37% of data considered as out-of-bag (OOB) and is used to evaluate the performance of the corresponding tree. To construct each tree in RF, instead of all available variables, a small random subset of variables is considered as candidate set of features for node splitting. To find the best candidate variable for node splitting,

we used the Gini index criterion [2], which is a measure of total variance across two OSA classes and is defined as:

$$GI = p_{m,non-OSA}(1 - p_{m,non-OSA}) + p_{m,m/s\_OSA}(1 - p_{m,m/s\_OSA})$$   Eq. III.3

where $p_{m,non-OSA}$ and $p_{m,m/s-OSA}$ are the proportion of observations in the $m^{th}$ node of the tree belonging to class non-OSA and moderate/severe-OSA, respectively. According to *Eq.III.3*, pure nodes will create regions with low $GI$ [2]. Thus, every time a split of a node is made on a specific feature, the $GI$ for the two descendent nodes is less than that of the parent node. Accordingly, for each feature, we calculated the average decrease in the $GI$ by applying the OOB data to all individual trees of the RF and provided the *mean-decrease-Gini* criteria; a measure of the importance of that feature in RF classification[7].

In a study by Louppe et al. [23], the asymptotic behavior of the RF algorithm was investigated, and it has been shown that the number of operations required for building RF in the best case has the upper bound of

$$O(kB\tilde{n} \, log^2\tilde{n}),$$   Eq. III.4

where $\tilde{n}$ is the size of bootstrapped data ($\sim$63% of the participants), $k$ is the number of variables sampled for splitting at each node, and $B$ is the number of trees in the forest. We considered B sufficiently large (B=1000) for the variance and error rate to be decreased. The common choice for $k$ in the literature is to be a square root of the number of available features in the dataset [7]. As we explained in Section III.2.3, the average number of discriminant features in our study is 45, which is a large number that might lead to a high computational cost. Besides, among these features there are many variables with high correlation/redundancy, which might affect accuracy and variable importance measure[23]. Therefore, to speed up the classification process and to avoid

bias, we used 3 variables with 2 random features at each node (because $\lceil \sqrt{3} \rceil = 2$) to grow the RF's trees. The use of these numbers comes from a study by Breiman [7] that showed the RF results are insensitive to $k$, and choosing one or two random features affords near the optimum results.

Consequently, to find the most proper feature set within each training set and to evaluate the generalization quality of our model, we investigated the classification ability of all possible candidate sets of size 3 among the discriminant features of that dataset (i.e. $\binom{\text{\# of discriminant features}}{3}$), and selected the one with the best classification result. To do so, we used a nested 10-fold cross-validation procedure, such that every time data of 9 folds used as training and data of one fold used for validation. We trained RF classifier using all feature sets of size 3 and recorded their accuracy, specificity, and sensitivity over the validation folds. Finally, we selected the feature set with maximum accuracy and sensitivity as the best feature set of that training dataset. For the classification purpose, we generated another RF classifier over the selected features of each dataset and classified the participants within training and blind-testing sets as either non-OSA or moderate/severe-OSA. We repeated this process ten times and reported the average of accuracy, specificity, and sensitivity for OOB data of training set as well as blind-testing data within each dataset. The mild-OSA participants were also classified using the selected features of each training dataset and the number of participants assigning to each of the non-OSA or moderate/severe-OSA groups was reported. Figure 11-b shows the flowchart of the RF method for feature selection and classification.
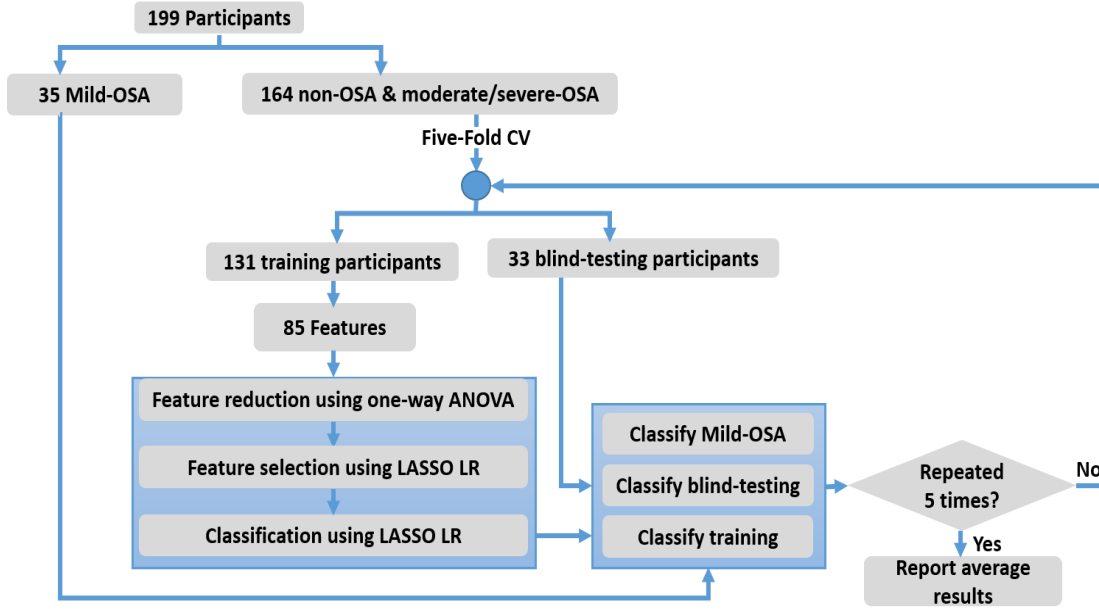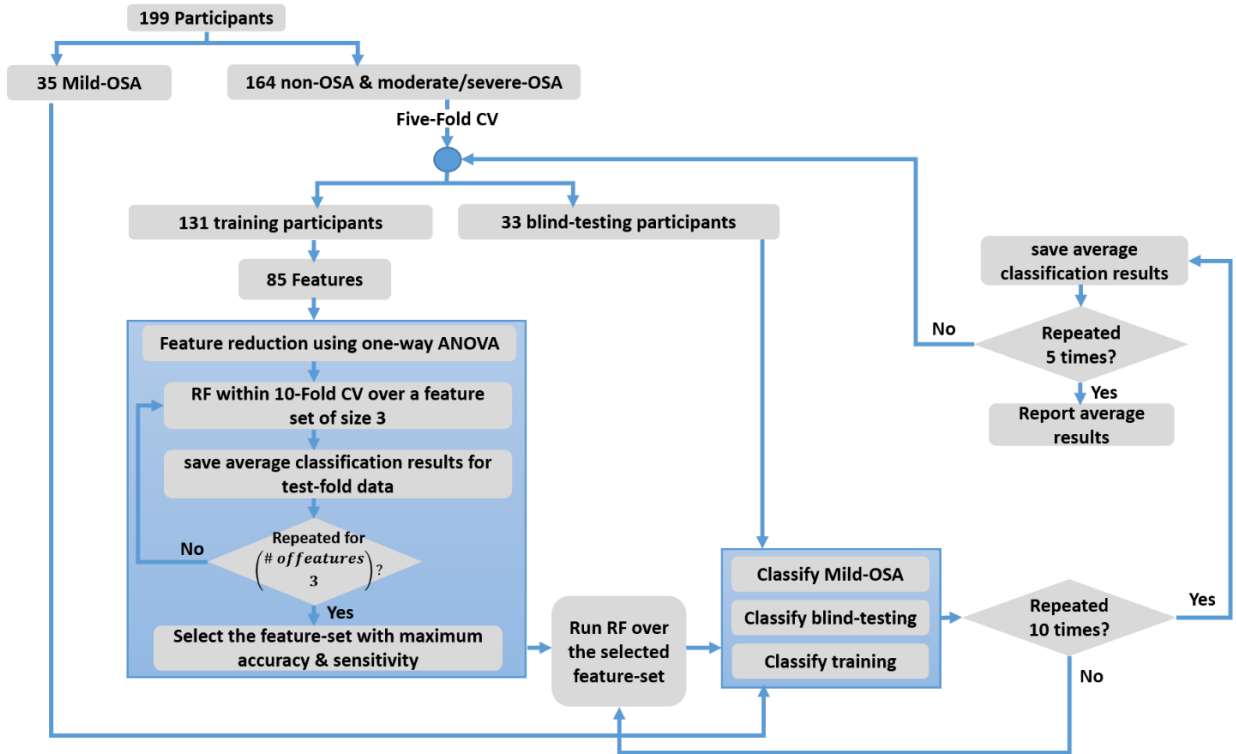
Figure 11 The flowchart of feature selection and classification methods. a) LASSO regularized LR, b) RF Procedure

### III.2.4 Statistical Analysis

Different feature selection methods choose different features, and various classification models use distinct training algorithms, therefore, they might perform differently. In order to get an indication of how similar the RF and LASSO regularized LR models are in their prediction performance and to assess their agreement, the Pearson correlation, Cohen's Kappa, and area under the receiver operating characteristics (ROC) curve (AUC) were calculated using the data of each training and blind-testing sets. The CI of correlation was used to measure how significant was the correlation between the outcomes of two models. Cohen's Kappa statistic was used to assess whether the two models are predicting the OSA severity level of the participants the same way or not. We also used this metric to compare the agreement between the predicted severities of each RF and regularized LR method to the severity levels detected by the PSG assessment. The higher the Kappa statistic, the better the agreement between the two models' output. The ROC chart obtained by plotting sensitivity vs. 1-specificity was used as a mean of comparison between the two classification models. AUC of each ROC curve is a measure of performance of the classification models and represents the probability of correct classification for a randomly chosen pair of non-OSA and moderate/severe-OSA subjects.

In addition to the comparison of the two models' performance, we also used different statistical tests to investigate the features selected by each of them. Unpaired student t-tests were used to both examine the statistical differences between selected features in each training dataset and to explore their discriminative power to separate non-OSA and moderate/severe-OSA participants. Multivariate analysis of Variance (MANOVA) tests were also used to investigate the discriminative power of the combination of the selected features, by either RF or LASSO

regularized LR, to discriminate the non-OSA and moderate/severe-OSA participants. We used the R statistical software to run the pre-mentioned tests.

## III.3 Results

The selected features of each dataset, using LASSO regularized LR and RF methods are listed in Table 7. These features were among F1 to F7 that are described in *Eq.III.5* to *Eq.III.10*. Here $PSD_{MI}$ and $PSD_{ME}$ represents the PSD of mouth inspiratory and expiratory TBS, and $PSD_{NI}$ and $PSD_{NE}$ represents the PSD of nose inspiratory and expiratory TBS, respectively.

$F1$: neck circumference (NC)

$F2: slope(10 \times log10(PSD_{MI} + PSD_{ME}(260:440\ Hz)))$        *Eq. III.5*

$F3: slope(( PSD_{ME}(250:350\ Hz)))$        *Eq. III.6*

$F4: mean(10 \times log10(PSD_{MI} + PSD_{ME}(1100:1350\ Hz)))$        *Eq. III.7*

$F5: mean(10 \times log10(PSD_{NI} + PSD_{NE}(1000:1300\ Hz)))$        *Eq. III.8*

$F6: mean(10log10(PSD_{MI} + PSD_{ME}(180:300\ Hz)))$        *Eq. III.9*

$F7: mean(PSD_{ME}(195:285\ Hz))$        *Eq. III.10*

Table 7 The features selected by the RF and the LASSO regularized LR feature selection methods in different training datasets. NC is the neck circumference; $PSD_{ME}$ is power spectrum density of mouth expiratory TBS; $PSD_{MI}$ is power spectrum density of mouth inspiratory TBS; $PSD_{NE}$ is power spectrum density of nose expiratory TBS; $PSD_{NI}$ is power spectrum density of nose inspiratory TBS. F presents a feature.

| Training set / Method | Training set 1 | Training set 2 | Training set 3 | Training set 4 | Training set 5 |
|---|---|---|---|---|---|
| Regularized-LR | F1, F2, F4 | F1, F2, F5 | F1, F2, F4 | F1, F2, F5 | F1, F3, F5 |
| RF | F1, F4, F6 | F1, F4, F7 | F1, F4, F6 | F1, F3, F7 | F1, F2, F6 |
| **F1**: neck circumference $F2: slope(10 \times log10(PSD_{MI} + PSD_{ME}(260:440\ Hz)))$ | | | | | |

$\text{F3: slope}(( PSD_{ME}(250:350 \text{ Hz})))$

$\text{F4: mean}(10 \times \log 10(PSD_{MI} + PSD_{ME}(1100:1350 \text{ Hz})))$

$\text{F5: mean}(10 \times \log 10(PSD_{NI} + PSD_{NE}(1000:1300 \text{ Hz})))$

$\text{F6: mean}(10\log 10(PSD_{MI} + PSD_{ME}(180:300 \text{ Hz})))$

$\text{F7: mean}( PSD_{ME}(195:285 \text{ Hz}))$

Figure 12 shows the features, that were selected by the RF approach in each of the five training sets and represents their importance by showing the amount of their mean-decrease-Gini index.
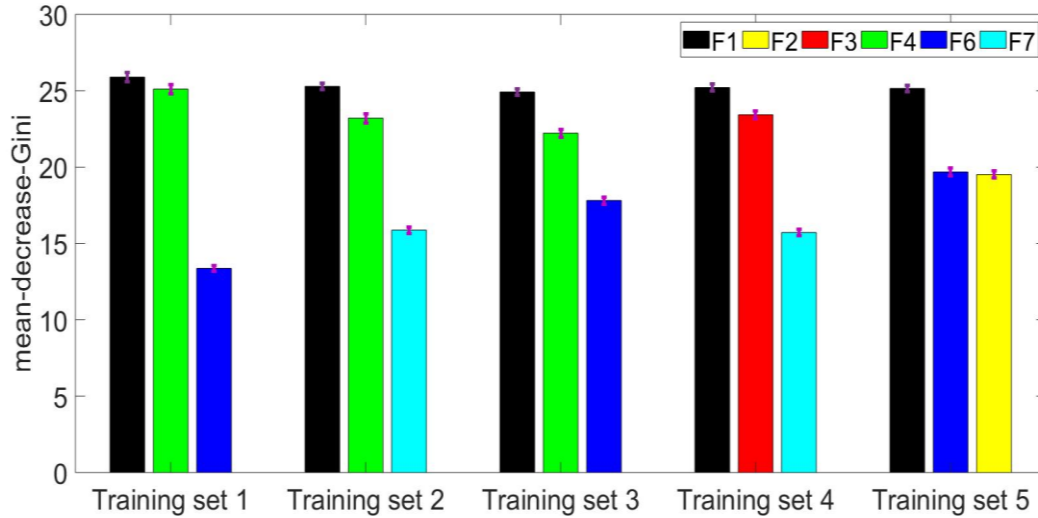


Figure 12 The features selected by applying the RF feature selection technique on data of the five training sets. The bars represent the importance of the features by showing the average amount of their mean-decrease-Gini index and their standard deviations over ten-times run of the RF model.

Our investigation depicts a significant correlation (p-value<0.05) between the AHI and the selected features of both methods within all datasets. As an example, the correlation between selected features of the third training dataset, in addition to their correlation with AHI is shown in Figure 13.

(a)                                                                                    (b)
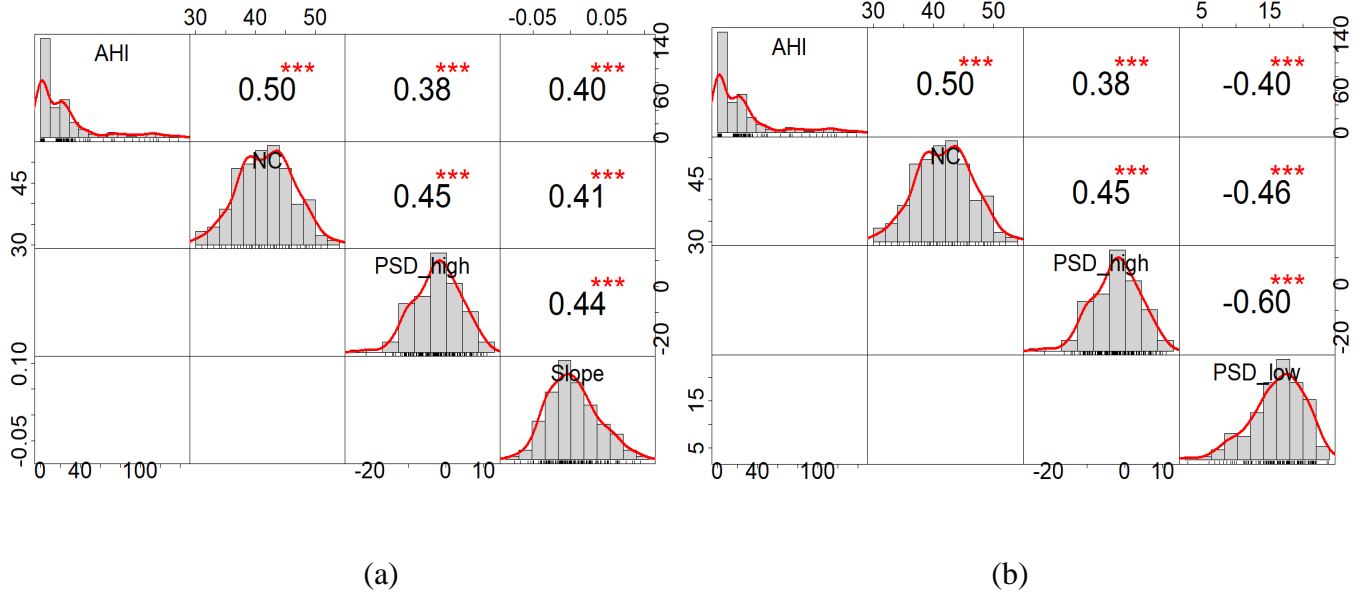
Figure 13 The values of the correlation plus their significance level as star (***: 0.001) between AHI and the selected features of the third training dataset, found using (a) LASSO regularized LR feature selection method; The selected features were neck circumference (NC), the average of the power spectrum density (PSD) calculated from summation of mouth inspiratory (MI) and mouth expiratory (ME) tracheal breathing sounds (TBS) in the high frequency range of [1100-1350 Hz], and the slope of the PSD calculated from summation of MI and ME TBS in the frequency range of [260-440 Hz], (b) RF; The selected features were NC, the average of the PSD calculated from summation of MI and ME TBS in the high frequency range of [1100-1350 Hz], and the average of the PSD calculated from summation of MI and ME TBS in the low frequency range of [180-300 Hz]. The overall distribution of features shows as histograms with a fitted curve.

Figure 14 shows the 3-dimensional scatter plot of the non-OSA and moderate/severe-OSA participants within the third training dataset, using the features selected by both feature selection methods. As shown in Figure 14-a and b, Non-OSA participants compared to moderate/severe-OSA participants have lower NC, lower power at high-frequency ranges [≥1000 Hz], lower slope around the second peak of the power spectra, and higher power at low-frequency ranges [150-300 Hz].

(a)                                                                              (b)
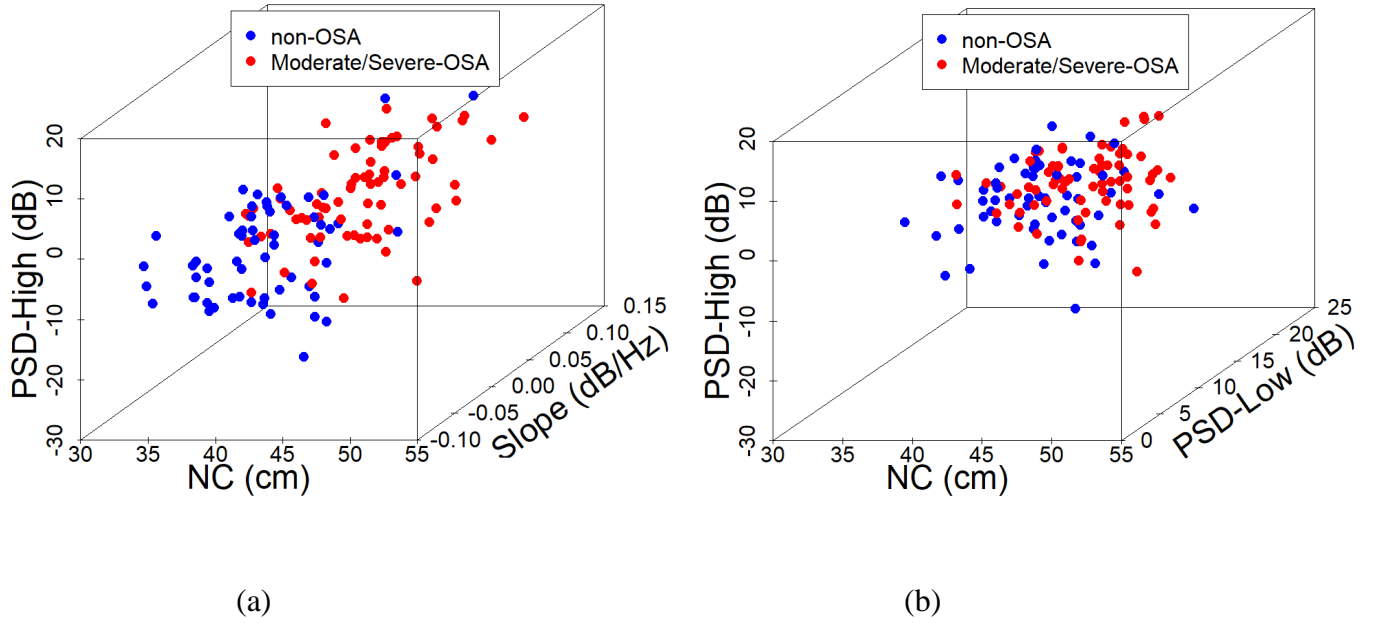
Figure 14 Three-dimensional scattering plot of the participants within the third training dataset, features found using (a) LASSO regularized LR feature selection method; The selected features were neck circumference (NC), the average of the power spectrum density (PSD) calculated from summation of mouth inspiratory (MI) and mouth expiratory (ME) tracheal breathing sounds (TBS) in the high-frequency range of [1100-1350 Hz] (PSD-High), and the slope of the PSD calculated from summation of MI and ME TBS in the frequency range of [260-440 Hz] (Slope), (b) RF; The selected features were NC, PSD-High, and the average of the PSD calculated from summation of MI and ME TBS in the low-frequency range of [180-300 Hz] (PSD-Low).

For both RF and LASSO regularized LR methods, the selected features within each dataset were shown to be statistically significantly different from each other (p-value $< 1.2 \times 10^{-35}$ and p-value $< 7.9 \times 10^{-6}$, respectively). Furthermore, the selected features of both methods depict a high discriminative power to separate the non-OSA and moderate/severe-OSA groups within each training dataset, either individually (p-value $< 4.6 \times 10^{-6}$ and p-value $< 1.3 \times 10^{-6}$, respectively) or in combination (p-value $< 2.1 \times 10^{-12}$ and p-value $< 6.5 \times 10^{-7}$, respectively).

Table 8 presents the specificity, sensitivity, and accuracy of applying the RF and LASSO regularized LR as feature selection and classification approaches over the non-OSA and

moderate/severe-OSA participants of all training and blind testing sets. The average of the RF classification results over the five datasets showed accuracy of 79.7±0.02 for OOB data of training sets and 82.1±0.06 for blind-testing data, respectively. The sensitivity (specificity) was 83.6±0.03 (75.2±0.04) for OOB data of training sets and 84.2±0.07 (79.5±0.08) for blind testing. The average of the LASSO regularized LR classification results over the five datasets showed accuracy of 82.3±0.7 and 79.3±6.1 for training and blind-testing data, respectively. The sensitivity (specificity) were 85.0±2.0 (78.4±1.0) for training and 82.2±7.2 (75.8±9.9) for blind-testing [6]. Table 9 reports the average of the AUC values associated with ROC curves of both RF and LASSO regularized LR classifiers for classification of participants within each dataset. The average of AUC values for ten times run of RF over training (blind-testing) data was 84.4±0.01 (86.0±0.005), while it was 89.9±1.7 (85.7±8.1) for the LASSO regularized LR approach.

Table 8 The classification results for participants within different training datasets and their corresponding testing datasets, using 10-times run of the RF classifier over its selected features and the LASSO regularized LR [6]. std is the standard deviation. OOB is the out-of-bag data of training sets in RF.

| Training set | Data | Random-Forest | | | Data | Regularized Logistic Regression | | |
| | | Specificity (%) ±std | Sensitivity (%) ±std | Accuracy (%) ±std | | Specificity (%) | Sensitivity (%) | Accuracy (%) |
| 1 | OOB | 82.9±0.04 | 85.1±0.02 | 84.1±0.02 | Training | 78 | 87.5 | 83.2 |
| | Testing | 72.8±0.07 | 83.0±0.06 | 78.4±0.04 | Testing | 60 | 83.3 | 72.7 |
| 2 | OOB | 75.9±0.04 | 83.5±0.03 | 80.1±0.02 | Training | 78 | 86.1 | 84.2 |
| | Testing | 73.2±0.09 | 77.8±0.05 | 75.7±0.04 | Testing | 73.3 | 72.2 | 72.7 |
| 3 | OOB | 73.5±0.04 | 86.9±0.03 | 80.8±0.02 | Training | 80 | 82 | 81.1 |
| | Testing | 85.6±0.06 | 88.9±0.08 | 87.6±0.05 | Testing | 85.7 | 77.8 | 81.3 |
| 4 | OOB | 70.4±0.08 | 82.0±0.07 | 76.8±0.05 | Training | 79.7 | 84.7 | 82.4 |
| | Testing | 86.7±0.08 | 83.6±0.07 | 85.0±0.05 | Testing | 80 | 88.8 | 84.8 |
| | OOB | 72.4±0.04 | 80.4±0.03 | 76.8±0.02 | Training | 76.2 | 87.5 | 82.4 |

| 5 | Testing | 79.1±0.1 | 87.6±0.1 | 83.8±0.1 | Testing | 80 | 88.9 | 84.8 |
|---|---|---|---|---|---|---|---|---|
| **Average** | OOB | *75.2±0.04* | *83.6±0.03* | *79.7±0.02* | Training | *78.4±1.0* | *85±2.0* | *82.3±0.7* |
|  | Testing | *79.5±0.08* | *84.2±0.07* | *82.1±0.06* | Testing | *75.8±9.9* | *82.2±7.2* | *79.3±6.1* |

Table 9 The area under the curve (AUC) values in association to their standard deviations (std) associated to receiver operating characteristics (ROC) curves of 10-fold cross-validation RF and the LASSO regularized LR classifiers [6], for classification of the non-OSA and moderate/severe-OSA participants in each training dataset.

| Training Set | Dataset | RF AUC (%)±std | LASSO regularized LR AUC (%) |
|---|---|---|---|
| **1** | Training | 89.6±0.002 | 92.1 |
|  | Blind-testing | 70.01±0.005 | 73.3 |
| **2** | Training | 84.5±0.003 | 90.0 |
|  | Blind-testing | 86.2±0.007 | 83.7 |
| **3** | Training | 85.2±0.003 | 87.3 |
|  | Blind-testing | 93.2±0.003 | 95.2 |
| **4** | Training | 81.4±0.003 | 89.4 |
|  | Blind-testing | 92.1±0.005 | 88.9 |
| **5** | Training | 81.1±0.05 | 90.5 |
|  | Blind-testing | 88.4±0.003 | 87.4 |
| **Average** | Training | 84.4±0.01 | 89.9±1.7 |
|  | Blind-testing | 86.0±0.005 | 85.7±8.07 |

Table 10 reports the Pearson correlation coefficients, their significance levels, and their CIs between the predicted OSA severity levels of the participants using RF and LASSO regularized LR methods across different datasets. Our results depict a high correlation between the outputs of

two methods within training (p-value $< 2.2 \times 10^{-16}$) and blind-testing datasets (p-value $< 2.3 \times 10^{-3}$) while neither of the CIs contains zero.

Table 10 The Pearson correlation coefficients and their significance level r% (p-value), in addition to the 90% confidence interval between RF and LASSO regularized LR classification methods within different training datasets and their corresponding testing data.

| Training Set | Data set | r% (p-value) | Confidence interval |
|---|---|---|---|
| Training set 1 | Training | 81.5% ($2.2 \times 10^{-16}$) | 0.76-0.86 |
| | Testing | 75.3% ($4.2 \times 10^{-7}$) | 0.59-0.86 |
| Training set 2 | Training | 81.3% ($2.2 \times 10^{-16}$) | 0.76-0.86 |
| | Testing | 51.1% ($2.3 \times 10^{-3}$) | 0.26-0.70 |
| Training set 3 | Training | 83.9% ($2.2 \times 10^{-16}$) | 0.79-0.88 |
| | Testing | 75.6% ($5.6 \times 10^{-7}$) | 0.59-0.86 |
| Training set 4 | Training | 69% ($2.2 \times 10^{-16}$) | 0.61-0.76 |
| | Testing | 76.2% ($2.7 \times 10^{-7}$) | 0.60-0.86 |
| Training set 5 | Training | 79.7% ($2.2 \times 10^{-16}$) | 0.74-0.84 |
| | Testing | 75.2% ($4.6 \times 10^{-7}$) | 0.59-0.86 |

Table 11 reports the Cohen's Kappa statistics and their significance levels between the predicted OSA severity levels of participants using RF and LASSO regularized LR methods across different datasets. This table also presents the Cohen's Kappa statistics and their significant levels for the severity levels detected by the PSG assessment and the outputs of these two methods. Our results showed a substantial agreement between the output of RF and LASSO regularized LR for training data (k>0.7) and a moderate agreement for blind-testing datasets (k>0.51). Besides, the severity levels predicted by the RF method shows a moderate to a substantial agreement with the

detected severity levels of PSG in both training (k>0.54) and blind-testing datasets (k>0.51). The same level of agreements was also observed for the severity levels predicted by the LASSO regularized LR method in both training (k>0.62) and blind-testing datasets (k>0.44).

Table 11 The Cohen's Kappa statistics and their significance level k(p-value) between the predicted severity levels of the OSA of participants using RF and LASSO regularized LR methods as well as predicted severities of these two methods to the severity levels detected by the PSG assessment, across different training datasets and their corresponding blind-testing data.

| Training Set | Data set | RF-PSG K (p-value) | LASSO-PSG K (p-value) | RF_LASSO K (p-value) |
|---|---|---|---|---|
| 1 | Training | $0.69\ (2.4 \times 10^{-15})$ | $0.66\ (4.2 \times 10^{-14})$ | $0.81\ (0)$ |
| | Testing | $0.57\ (0.001)$ | $0.44\ (0.001)$ | $0.75\ (1.5 \times 10^{-5})$ |
| 2 | Training | $0.61\ (2.2 \times 10^{-12})$ | $0.64\ (1.6 \times 10^{-13})$ | $0.81\ (0)$ |
| | Testing | $0.51\ (0.003)$ | $0.45\ (0.009)$ | $0.51\ (3.3 \times 10^{-3})$ |
| 3 | Training | $0.6\ (4.6 \times 10^{-12})$ | $0.62\ (1.2 \times 10^{-12})$ | $0.83\ (0)$ |
| | Testing | $0.75\ (2.4 \times 10^{-5})$ | $0.62\ (0.0004)$ | $0.75\ (1.9 \times 10^{-5})$ |
| 4 | Training | $0.54\ (8.7 \times 10^{-10})$ | $0.65\ (1.2 \times 10^{-13})$ | $0.69\ (2.9 \times 10^{-15})$ |
| | Testing | $0.7\ (6.2 \times 10^{-5})$ | $0.7\ (6.7 \times 10^{-5})$ | $0.76\ (1.2 \times 10^{-5})$ |
| 5 | Training | $0.54\ (8.4 \times 10^{-10})$ | $0.63\ (6 \times 10^{-13})$ | $0.78\ (0)$ |
| | Testing | $0.7\ (6.7 \times 10^{-5})$ | $0.7\ (6.7 \times 10^{-5})$ | $0.75\ (1.6 \times 10^{-5})$ |

Table 12 reports the average number of the mild-OSA participants who assigned to non-OSA and moderate/severe-OSA groups, using RF and LASSO regularized LR classification procedures. For both classification methods, on average, 12 of the mild-OSA participants were assigned to non-OSA and 23 of them to moderate/severe-OSA groups.

Table 12 The number of Mild-OSA participants classified as non-OSA or moderate/severe-OSA, using RF and LASSO regularized LR classification procedures, trained with different training datasets.

| Training Set | RF | | LASSO regularized LR | |
|---|---|---|---|---|
| | non-OSA | moderate/severe-OSA | non-OSA | moderate/severe-OSA |
| 1 | 12 | 23 | 12 | 23 |
| 2 | 11 | 24 | 11 | 24 |
| 3 | 12 | 23 | 13 | 22 |
| 4 | 11 | 24 | 11 | 24 |
| 5 | 12 | 23 | 11 | 24 |
| Average | 11.6 | 23.4 | 11.6 | 23.4 |

### III.3.1 Computational Complexities

Time complexity evaluates the asymptotic behavior of an algorithm. Big O notations are used to characterize functions according to their growth rates and to express an upper bound for them [23]. The algorithms that we utilized in this study have 3 phases of analysis: feature extraction, feature reduction and classification. As elaborated in our previous study [6], the total cost of extracting $p$ feature and applying the one-way ANOVA to select $\hat{p}$ discriminate features would be

$$16n \times O\big(Nlog(L)\big) + pO(n), \qquad \text{Eq. III.11}$$

where $n$ is the size of the participants, $N$ is the length of normalized TBS, and $L$ is the window size. In addition, the computational cost of the feature reduction and classification using the LASSO regularized LR would be [6]

$$O(\hat{p}n) + O(n). \qquad \text{Eq. III.12}$$

For the RF, the computational cost of classification is shown in *Eq. III.4*. Since in feature reduction phase we ran the RF within a 10-fold cross-validation procedure over all combinations of feature-sets of size 3, (i.e. $\binom{\hat{p}}{3}$), the total cost of the RF feature reduction would be

$$10\binom{\hat{p}}{3}O(\sqrt{3}B\tilde{n}\ log^2\tilde{n}), \qquad\qquad Eq.\ III.13$$

while $\tilde{n} = 63.3\% \times (0.9n)$ and B=1000. The RF classification has been done over the best feature set of size 3 that selected in the feature reduction phase. Consequently, the computational cost of the feature selection and classification using the RF approach would be

$$10\binom{\hat{p}}{3}O(\sqrt{3}B\tilde{n}\ log^2\tilde{n}) + O(\sqrt{3}B\tilde{n}\ log^2\tilde{n}). \qquad\qquad Eq.\ III.14$$

Comparing *Eq. III.12* to *Eq. III.14* reveals that the computational cost of the LASSO regularized LR is less than that of RF; hence, the LASSO regularized LR approach is faster.

## III.4 Discussion

When the number of extracted characteristics is large, it is difficult to build and more importantly, interpret a model based on all the features. Since the manual separation of the relevant and irrelevant features is challenging, proper feature selection procedures are necessary. In this study, we compared the RF and the LASSO regularized LR to predict the OSA severity level of the participants. These approaches have been used as feature reduction and classification methods to process the anthropometric information and spectral features of wakefulness TBS of non-OSA and moderate/severe-OSA groups.

Our classification results show that both proposed classifiers could be considered as useful tools in OSA screening with an average correct classification rate higher than 82.1% for RF and 79.3% for the LASSO regularized LR, respectively (Table 8). The Pearson correlation coefficients

showed a high correlation between the classification results of RF and LASSO regularized LR with the average correlation coefficient of 79.1 for training and 70.7 for blind-testing datasets, respectively (Table 10). The Cohen's Kappa statistic also showed a substantial agreement between the severity levels predicted by RF and LASSO regularized LR for training (>0.69) and moderate agreement for blind-testing dataset (>0.51), (Table 11). In addition, the predicted severity levels of both methods show moderate to substantial agreement with the detected severity levels of PSG assessment in training (K>0.54) and blind-testing sets (k>0.44), respectively (Table 11). We considered the AUC statistics to assess the robustness and predictive ability of these two methods. The LASSO regularized LR and RF achieved the average blind-testing AUC values of 85.7±8.07 and 86.0±0.005, respectively (Table 9). Accordingly, both approaches are efficient in diagnosing OSA using wakefulness TBS. However, the RF model outperformed the results of the LASSO regularized LR in terms of accuracy, specificity, and sensitivity of blind-testing with 3.5%, 2.4%, and 3.7% improvement, respectively. Despite better classification results for the RF model, the LASSO regularized LR is a faster technique because the computational complexities of this technique *(Eq. III.12)* is lower than that of the RF (*Eq. III.14*).

Statistical investigations over the selected features within different datasets (F1 to F7) confirmed a highly significant correlation between these features and the AHI values of participants. These features were also shown to be statistically significantly different from each other and demonstrated a high discriminative power to separate two OSA groups, either individually or in combination. Between these features, all except F5 were selected by the RF approach, while F1, F4, and F6 had the highest repetition across all datasets. In addition, according to their mean-decrease-Gini criteria, they were the most critical features in OSA prediction, respectively (Figure 12). Meantime, the LASSO regularized LR selected F1 to F5 in different

datasets, and the most repeated ones were F2, F1, F5, and F4. Among the most repeated features of both methods, two of them were common; the neck circumference and the average of the power spectra of summation of mouth inspiratory and expiratory TBS in the high-frequency range of [1100-1350 Hz]. The main difference between the two methods was that the LASSO regularized LR selected the slope around the second peak of the TBS' power spectra as its third feature, while RF chooses the average of the power spectra of TBS in the low-frequency ranges. Further investigation on the correlation between selected features of each method (Figure 13) revealed the LASSO regularized LR focusses on choosing an optimal set of variables that are not highly correlated to each other while RF considers the accuracy but does not exclude variables that are highly correlated to others. In this way, a small subset of uncorrelated features for distinguishing the two groups can be obtained using the LASSO regularized LR, but a list of influential features (important for clinical diagnosis) can be obtained using RF. These findings could explain better classification results for RF than LASSO regularized LR.

Classification of mild-OSA participants showed that on average, both classification techniques were more concern about sensitivity and assigned further participants to the moderate/severe-OSA group compared to the non-OSA group (Table 12). For the pre-screening methods, generally, high sensitivity is more preferable, because misclassification of an OSA patient as non-OSA is more troublesome than misclassification of a non-OSA individual as an OSA patient.

The last but not least point is related to the generalizability of our proposed methods. In this study, to reduce the feature extraction bias, we used a 5-fold cross-validation procedure to randomly divide our participants into five different training sets. Additional randomness was also added to the mechanism of learning of our models, either by the cross-validation in LASSO regularized LR to find the optimal value of $\lambda$ or through bootstrapping and random variable

88

selection for node splitting in RF. These randomnesses may lead to small variations in results after repetition. However, the stochastic nature of the LASSO regularized LR and RF approaches is their power, because these models could learn over time; i.e., as the number of samples increases, the predictions become more accurate.

To sum up, LASSO regularized LR is a parametric linear model and similar to all model-based methods, its prediction performance depends on whether the data follow the assumed model or not. This approach provides a restrictive set of parameters related to usually a smaller number of features than the original feature set. Using the estimated coefficients of this model, it would be possible to examine the features that are associated with the OSA severity levels and evaluate the extent of their association and importance [2], [6]. RF, contrarily, is a non-parametric approach that does not rely on any model; hence, it is more flexible than the LASSO regularized LR and can result in a better final prediction model. In RF, the importance of each feature and the relationship between risk factors and assigned OSA severity level of individuals, as well as the extent of these associations, is determined by the mean-decrease-Gini index criterion. As no assumptions are needed to be checked before the investigation of data using the RF and since this model could deal with the unbalanced and missing data, this approach seems to be more practical for the real data with lots of missing values. The main drawback of the RF is the model size, such that for the predictions we need to keep a whole forest in the memory. This may take a considerable capacity of memory and might become slow to evaluate. In this aspect, the LASSO regularized LR might be a better method because all predictions could be made using a small number of coefficients that fit in a simple LR model.

### III.4.1 Physiological interpretation of the selected features

The selected features of this study, using both approaches, are physiologically meaningful and are highly correlated with AHI. We elaborated on the physiological interpretation of these features in our previous study [6]. Briefly, for the moderate/severe OSA participants, high NC relates to the UA narrowing. Also, the higher slope around the second peak and the higher average power in the high-frequency range of their power spectra relate to the lower resonance frequency and increased stiffness of their UA tissues, respectively [6]. The only newly selected feature is the average of the power spectra in low-frequency ranges. Imaging studies have shown a narrower and thicker UA for severe OSA individuals compared to that of controls during wakefulness [24], [25]. It has been demonstrated that the narrower UA of OSA patients compared to that of controls is, at least in part, the reason for increased collapsibility or compliance of OSA subjects [26]. Studies have shown that the more compliant tubes as in severe OSA, have higher absorbance of the low-frequency sounds [27]. These points justify our observation of a lower average power at low frequencies for moderate/severe OSA group compared to that of non-OSA (p-value $< 2.7 \times 10^{-9}$).

### III.5 Conclusion

In this study, the RF and the LASSO regularized LR have been compared for the first time for the feature selection and classification of non-OSA and moderate/severe-OSA participants during wakefulness. The comparative analysis in this study may allow us to gain an insight into the performance of these two methods for the online screening of OSA. LASSO regularized LR is a fast, parametric model that is able to show the relationship between the features and OSA severity level of participants by a small number of coefficients. Contrarily, the RF is a flexible non-parametric model that identifies a list of influential features. This method measures the importance of each variable to show the extent of relationships between selected features and the OSA severity

level of individuals. Although the RF approach predicted the OSA severity level of participants in a more accurate manner, hence, have lower generalization error compared to the LASSO regularized LR, the classification results of the two methods were highly correlated. Besides, considering the computational costs and memory usage of two approaches, the LASSO regularized LR showed to be faster and have a smaller sized model than the RF. The pre-mentioned points suggest that the selection of the appropriate feature reduction and classification approach depends on the application and nature of data. Generally, in classification problems, the higher accuracy is preferred. Hence, the RF approach might be considered as the preferred approach. However, if the size of data is large and a quick real-time screening is required, the LASSO regularized LR approach might be a better choice as it can still provide relatively fast but accurate classification results.

## References

[1]    D. J. Power, "Using 'Big Data' for analytics and decision support," *J. Decis. Syst.*, vol. 23, no. 2, pp. 222–228, 2014.

[2]    G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, vol. 103. New York, NY: Springer New York, 2013.

[3]    I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *J. Mach. Learn. Res.*, vol. 3, no. 3, pp. 1157–1182, 2003.

[4]    T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY: Springer New York, 2009.

[5]    R. Tibshirani, "Regression shrinkage and selection via the lasso: a retrospective," *J. R. Stat. Soc. Ser. B (Statistical Methodol.*, vol.58, no. 1, pp. 267–288, 1996.

[6]    F. Hajipour, M. Jafari Jozani, A. Elwali, and Z. Moussavi, "Regularized logistic regression for obstructive sleep apnea screening during wakefulness using daytime tracheal breathing sounds and anthropometric information," *Med. Biol. Eng. Comput.*, vol. 57, no. 12, pp. 2641–2655, Nov. 2019.

[7]    L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[8]    B. A. Goldstein, A. M. Navar, and R. E. Carter, "Moving beyond regression techniques in cardiovascular risk prediction: Applying machine learning to address analytic challenges,"

*Eur. Heart J.*, vol. 38, no. 23, pp. 1805–1814, 2017.

[9] N. Ludwig, S. Feuerriegel, and D. Neumann, "Putting Big Data analytics to work: Feature selection for forecasting electricity prices using the LASSO and random forests," *J. Decis. Syst.*, vol. 24, no. 1, pp. 19–36, 2015.

[10] B. Zhao, "A Machine Learning Approach to Predicting Case Duration for Robot-Assisted Surgery," *J. Med. Syst.*, vol. 43, no. 2, p. 32, 2019.

[11] J. G. Park, K. Ramar, and E. J. Olson, "Updates on Definition, Consequences, and Management of Obstructive Sleep Apnea," *Mayo Clin. Proc.*, vol. 86, no. 6, pp. 549–555, Jun. 2011.

[12] C. A. Kushida *et al.*, "Practice Parameters for the Indications for Polysomnography and Related Procedures: An Update for 2005," *Sleep*, vol. 28, no. 4, pp. 499–523, 2005.

[13] S. G. Memtsoudis, M. C. Besculides, and M. Mazumdar, "A Rude Awakening — The Perioperative Sleep Apnea Epidemic," *N. Engl. J. Med.*, vol. 368, no. 25, pp. 2352–2353, Jun. 2013.

[14] American Academy of Sleep Medicine, *Hidden Health Crisis Costing America Billions: Underdiagnosing and Undertreating Obstructive Sleep Apnea Draining Healthcare System.* Mountain View, CA: Sullivan & Frost, 2016.

[15] K. N. Priftis, L. J. Hadjileontiadis, and M. L. Everard, *Breath Sounds From Basic Science to Clinical Practice*. Cham: Springer International Publishing, 2018.

[16] T. Penzel and A. Sabil, "The use of tracheal sounds for the diagnosis of sleep apnoea," *Breathe*, vol. 13, no. 2, pp. e37–e45, Jun. 2017.

[17] I. Ayappa and D. M. Rapoport, "The upper airway in sleep: physiology of the pharynx," *Sleep Med. Rev.*, vol. 7, no. 1, pp. 9–33, Feb. 2003.

[18] J. Sola-Soler, J. A. Fiz, A. Torres, and R. Jane, "Identification of Obstructive Sleep Apnea patients from tracheal breath sound analysis during wakefulness in polysomnographic studies," *Conf. Proc. ... Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2014, pp. 4232–4235, 2014.

[19] A. Elwali and Z. Moussavi, "Obstructive Sleep Apnea Screening and Airway Structure Characterization During Wakefulness Using Tracheal Breathing Sounds," *Ann. Biomed. Eng.*, vol. 45, no. 3, pp. 839–850, Mar. 2017.

[20] A. Montazeri, E. Giannouli, and Z. Moussavi, "Assessment of Obstructive Sleep Apnea and its Severity during Wakefulness," *Ann. Biomed. Eng.*, vol. 40, no. 4, pp. 916–924, Apr. 2012.

[21] A. Yadollahi and Z. M. K. Moussavi, "Acoustical flow estimation: Review and validation," *IEEE Eng. Med. Biol. Mag.*, vol. 26, no. 1, pp. 56–61, Jan. 2007.

[22] R. J. SIMES, "An improved Bonferroni procedure for multiple tests of significance," *Biometrika*, vol. 73, no. 3, pp. 751–754, 1986.

[23] G. Louppe, "Understanding Random Forests: From Theory to Practice," Doctoral

dissertation, University of Liège, 2014.

[24] Y. Finkelstein *et al.*, "Velopharyngeal Anatomy in Patients With Obstructive Sleep Apnea Versus Normal Subjects," *J. Oral Maxillofac. Surg.*, vol. 72, no. 7, pp. 1350–1372, Jul. 2014.

[25] R. J. Schwab *et al.*, "Identification of Upper Airway Anatomic Risk Factors for Obstructive Sleep Apnea with Volumetric Magnetic Resonance Imaging," *Am. J. Respir. Crit. Care Med.*, vol. 168, no. 5, pp. 522–530, Sep. 2003.

[26] R. J. Schwab, K. B. Gupta, W. B. Gefter, L. J. Metzger, E. A. Hoffman, and A. I. Pack, "Upper airway and soft tissue anatomy in normal subjects and patients with sleep-disordered breathing. Significance of the lateral pharyngeal walls.," *Am. J. Respir. Crit. Care Med.*, vol. 152, no. 5, pp. 1673–1689, 1995.

[27] F. Bechwati *et al.*, "Low frequency sound propagation in activated carbon," *J. Acoust. Soc. Am.*, vol. 132, no. 1, pp. 239–248, 2012.

Chapter IV. Acoustic Characterization of Upper Airway Variations with Respect to OSA

# IV.I Spectral and Higher Order Statistical Characteristics of Expiratory Tracheal Breathing Sounds during Wakefulness and Sleep in People with Different Levels of Obstructive Sleep Apnea

Farahnaz Hajipour and Zahra Moussavi

*Abstract* — We investigated plausible changes in spectral and higher order statistical properties of tracheal respiratory sounds from wakefulness to sleep in relation to obstructive sleep apnea (OSA). Data consisted of expiratory sounds of 30 participants suspected of OSA during wakefulness and sleep, both recorded in supine position. Participants were divided into two groups of mild and severe OSA (15 in each group) based on their apnea/hypopnea index (AHI) per hour. Three different frequency-based features of their power spectra in addition to Kurtosis and Katz fractal dimension (KFD) were estimated from each normalized expiratory sound; they were compared within and between the groups. During wakefulness, the average sounds power at low-frequency components in the severe group was less than that of the mild group. However, during sleep, the average power of high-frequency components in severe subjects was more than that of the mild group. The kurtosis value of both mild and severe OSA groups increased significantly from wakefulness to sleep using both mouth and nasal breathing sounds during wakefulness. The KFD increased significantly from wakefulness to sleep for both mild and severe OSA group using only nasal breathing sounds during wakefulness. These changes are indicative that the upper airway of severe OSA show more compliance and thickness compared to that of the mild OSA during both wakefulness and sleep and represent an increased stiffness during sleep. This implies a regional narrowing which cause both more compliance and stiffness simultaneously in different regions of the upper airway.

## IV.I.1 Introduction

Obstructive sleep apnea (OSA) syndrome is due to the partial or complete collapse of the upper airway (UA) [1], resulting from relaxation of muscles controlling the soft palate or tongue. [2]. An individual with OSA, despite persistent efforts to breathe, may experience periods of cessation of breathing (apnea) and/or more than 50% reduction in airflow (hypopnea). Such event is called apnea or hypopnea if it lasts at least 10s and is associated with a minimum of 4% drop in oxygen saturation level in blood [1]. The severity of OSA is measured by apnea/hypopnea index (AHI) per hour of sleep.

OSA is a prevalent health problem that affects both children and adults. It has been reported that about 10% of the North American population suffer from OSA [3], while it is believed there are many undiagnosed cases [4]. OSA is associated with an increased risk of cardiovascular disease, daytime sleepiness, reduced concentration and increased risk of car accidents [1]. The current gold standard for diagnosis of OSA is full-night polysomnography (PSG). Many physiological signals such as heart and muscle signals, respiratory effort, respiratory flow, and brain waves (EEG) are recorded during a PSG assessment to provide a full physiological picture of the patients' sleep apnea and sleep quality. However, PSG assessment is time-consuming, cumbersome and expensive with long waiting list around the world. Therefore, designing simpler assessment methods such as portable home monitoring devices offer an alternative way to detect sleep apnea while overcoming the drawbacks of PSG. In this study, we aimed to explore the use of breathing sounds in understanding UA differences in OSA individuals with different levels of OSA severity during both wakefulness and sleep. The outcomes of our study may lead to a better physiological understanding of the mechanism of the UA collapse during sleep using only breathing sound analysis.

95

Various features of UA anatomy and neuromuscular control contribute to OSA development [5]. Therefore, the cause of OSA may vary considerably between individuals. Some of the previous studies using MRI/CT imaging during wakefulness have shown that OSA individuals have a narrower and thicker UA compared to non-OSA individuals [6]. Also, the lateral narrowing of the UA in OSA individuals was found to be of an elliptical configuration in the anterior-posterior dimension in contrast to a more laterally open in non-OSA individuals [7], [8]. Nevertheless, it has been reported that the patency of the UA of OSA individuals during wakefulness is well maintained due to an increase in their dilator muscles activities [9], [10]. On the other hand, the UA of OSA participants during sleep has been shown to be more collapsible [11], most probably due to the changes in neuromuscular control and airway physiology at the onset of sleep [12].

Our team and a few others around the world have been using tracheal respiratory sounds to monitor and assess OSA. It is known that structural and physiological properties of the UA (such as diameter, wall thickness, and length) will affect the breathing sound generation mechanism [13]. Therefore, we hypothesize that respiratory sounds are able to convey valuable information in relation to monitoring and detecting OSA. Some previous studies focused on analyzing respiratory sound during wakefulness for OSA assessment and classification [14], [15]. In this study, however, we investigated breathing sound features during both wakefulness and sleep to determine which sound features change the most from wakefulness to sleep differently among people with different levels of OSA severity. We focused on investigating the spectral and higher order statistical characteristics of respiratory sounds during both wakefulness and sleep in relation to OSA.

## IV.I.2 Method

### IV.I.2.1 Data

We used tracheal respiratory sounds of 30 individuals with OSA. Tracheal breathing sounds during both wakefulness and sleep were recorded in the supine position with head resting on a pillow. During wakefulness, participants were instructed to breathe at their normal rate in two maneuvers: first through the nose with mouth closed and second through the mouth with a nose clip in place; five full breathing cycles in each maneuver were recorded. After wakefulness recording, participants were prepared for PSG assessment. Breathing sounds recording during sleep were conducted simultaneously with full-night PSG assessment at Misericordia Health Center (Winnipeg, Canada). The severity of OSA was determined using the PSG report and an AHI threshold of 15. Data included 15 participants with mild-OSA (11 male, AHI<15) and 15 severe-OSA (13 male, AHI>15). Demographic information of the participants is shown in Table 13. This study was approved by the Biomedical Research Ethics Board of the University of Manitoba and all the participants signed an informed consent form prior to data collection.

Table 13 Study Participants Demographic Information

| Severity | Age | AHI | BMI | Neck size | Sex (M: F) |
|----------|-----|-----|-----|-----------|------------|
| Mild-OSA | 40.7 | 2.2 | 29.6 | 40.9 | (11:4) |
| Severe-OSA | 48.7 | 37.5 | 33.6 | 45.7 | (13:2) |

The breathing sounds during both wakefulness and sleep were collected with a small microphone (Sony ECM-77B). The Microphone was inserted into a small chamber which allowed 2 mm cone-shape space with skin to ensure that it never connects the skin of participants during the recording and mounted over the suprasternal notch of their trachea. A soft neckband was

wrapped around the patient's neck to sustain the microphone and chamber in place, and to ensure that the microphone would not be misplaced during the night. The recorded breathing sounds were amplified by a Biopac (DA100C) amplifier with the band-pass filter in the range of 0.5-5000 Hz and digitized at 10240 Hz sampling rate.

## IV.I.2.2 Signal Analysis

All the recorded tracheal breathing sounds data were examined manually by listening to and observing the sounds in the time-and-frequency domain for plausible occasional swallow that may interfere with altering breath phases and to exclude noisy signals or those associated with snoring sounds. As the respiratory flow was not recorded in this study, to ensure the accuracy of the phase labels during wakefulness, each recording was always started at the inspiration. The inspiratory and expiratory phases during wakefulness and sleep were identified semi-manually using the technique elaborated in [16]. Briefly, the log of the variance of each phase is calculated and the onset of each phase is identified by an automatic algorithm. Then, based on different features from duration, shape and volume of the sound envelope each phase is labeled as inspiration or expiration. During sleep, snoring usually occurs within the inspiration phase. Therefore, to avoid plausible snoring sounds, we selected 4 noise-free expiratory sounds in supine position (determined using the PSG assessment). During wakefulness, we also analyzed 4 noise-free expiratory sounds for comparison to those during sleep. We extracted sleep data from only stage 2 of sleep because that was the most common stage in our dataset. To remove the effect of low and high-frequency noises, including heart and muscle sounds and ambient noise, each individual expiratory sound segment was passed through a 5$^{th}$-order Butterworth band-pass filter in the range of [75-2500 Hz].

Next, the same procedure as in [15] was applied to each selected expiratory sound: each expiratory phase was first normalized by its variance envelope (moving average filter of the signal with 64 sample sequence) to remove its extra fluctuations. Subsequently, they were normalized by their energy to compensate for plausible different flow rates in each breathing cycle. To capture the stationary part of the expiratory breathing sounds, we calculated the logarithm of the sound's variance using the method in [17], and the sound segments corresponding to its middle part (50% duration around the maximum) was considered for further analysis. Next, we estimated the power spectrum density (PSD) of the stationary portion of each expiratory sound signal using Welch's method in windows of 205-point (~20 ms) with 50% overlap between successive windows and averaged over the 4 expiratory phases of each participant. Next, three frequency-based features were calculated from the average PSD: 30%-freq, 50%-freq and 70%-freq, the frequency at which the PSD reaches 30%, 50% and 70% of the total power, respectively. In addition, the kurtosis and Katz fractal dimension (KFD) [18] were also calculated in every 20ms window with 50% overlap with the adjacent windows and then averaged within the stationary part of each expiratory phase. They were then averaged over the 4 expiratory phases of each participant.

The extracted features were then compared within the mild and severe groups and between the wakefulness and sleep. For comparison, we used paired and unpaired t-tests (for within and between group comparisons); a p-value of 0.05 was considered as significant.

## IV.I.3 Results

Investigating different frequency-based features showed that the PSD of mild OSA reached their 30%-freq in lower frequencies compared to that of severe OSA using both mouth and nasal breathing sounds of wakefulness; though, it was only significant ($p<0.008$) during mouth breathing maneuver (Figure 15). The same pattern was also observed during sleep ($p< 0.05$) (Figure 15).

These results imply that the PSD of the mild group at the lower frequencies are more powerful than that of the severe participants during both wakefulness and sleep.
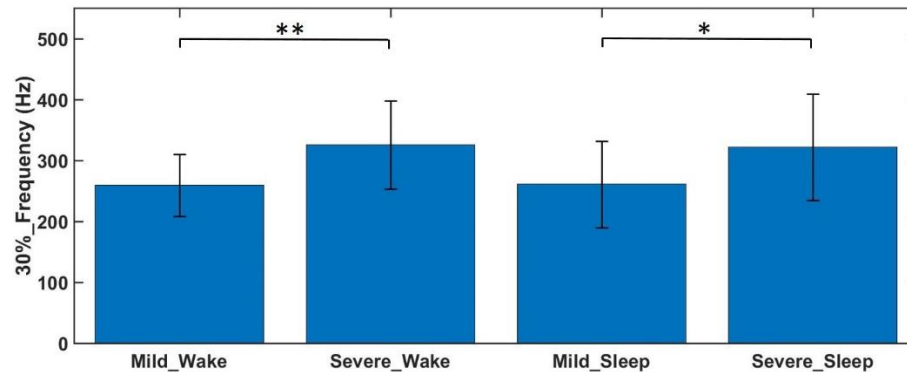


Figure 15 Mean and standard error of 30%-freq feature extracted from average expiratory breathing sounds PSD of mild and severe OSA groups during mouth breathing in wakefulness and stage 2 of sleep.

As shown in Figure 16, the average power spectra of severe OSA reached both their 50% and 70% power in significantly higher frequencies than the mild OSA ($p<0.02$, $p<0.05$, respectively). This finding implies that, the average PSD of severe participants during sleep has more power in higher frequencies compare to the average PSD of mild participants. No significant difference between mild and severe OSA groups was observed during either mouth or nasal breathing sounds of wakefulness (Figure 16).



(a)
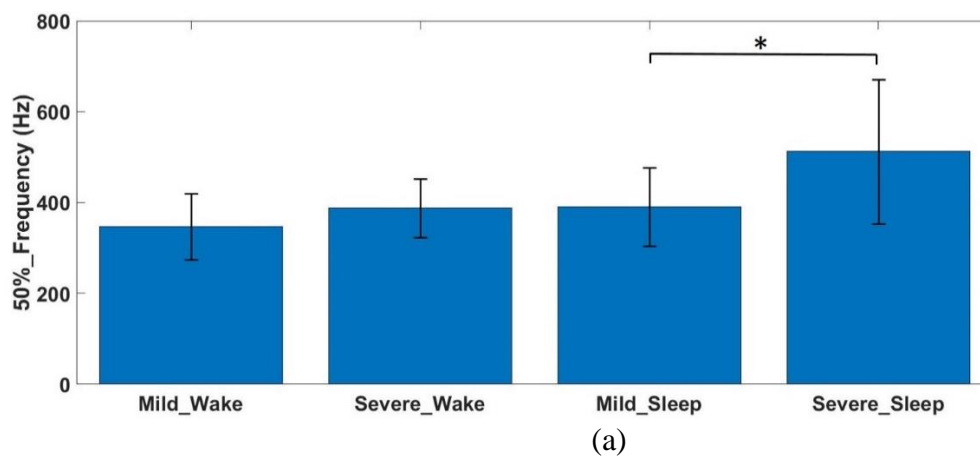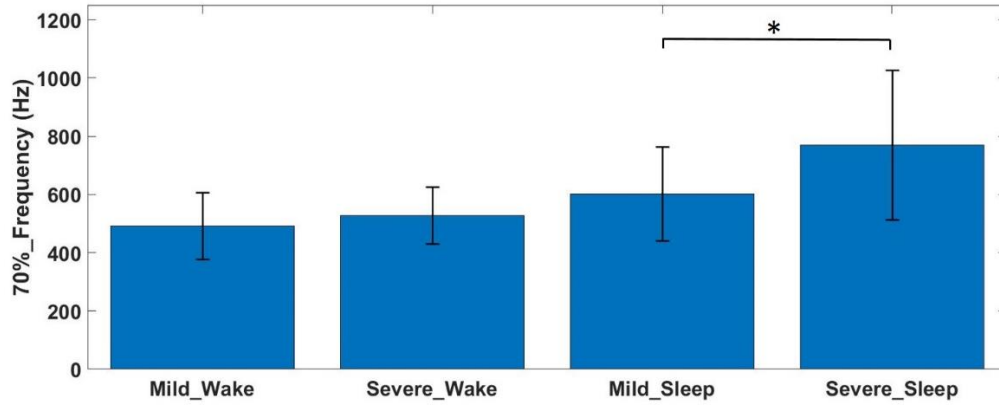
(b)

Figure 16 Mean and standard error of a) 50% -freq and b) 70%-freq features. These features were extracted from the average expiratory breathing sounds PSD of mild and severe OSA groups during mouth breathing maneuver in wakefulness and stage 2 of sleep.

Higher order statistical analysis showed that there was a significant change in the kurtosis of the breathing sounds from wakefulness to sleep in both mild ($p<0.007$) and severe ($p<0.002$) OSA groups, using the nasal breathing sounds during wakefulness (Figure 17). The same pattern was also observed when we used expiratory mouth breathing sounds instead of nasal breathing sounds during wakefulness ($p< 0.02$ and $p< 0.04$ for mild and severe OSA groups, respectively).
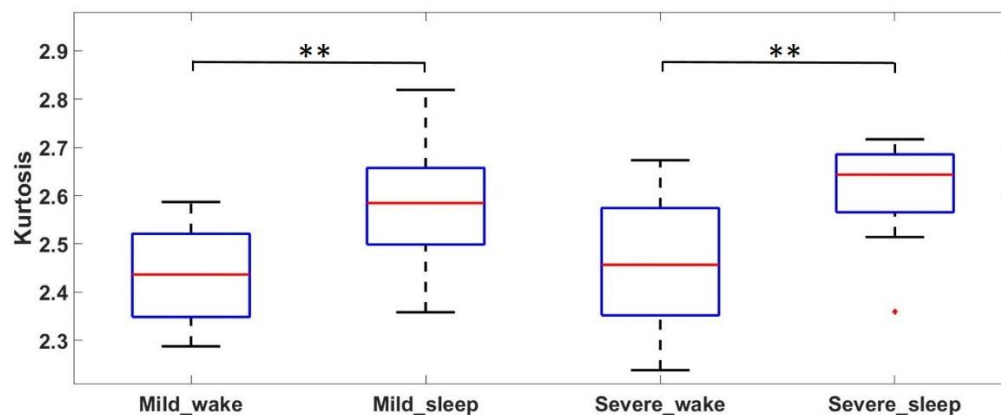


Figure 17 t-test outcome of kurtosis change from wakefulness to sleep in mild and severe OSA groups using nasal breathing sounds during wakefulness.

The KFD analysis revealed that for both mild and severe OSA groups, there was no significant change in this feature from wakefulness to sleep when using mouth-breathing sounds during wakefulness. However, the change of the KFD from wakefulness to sleep using nasal breathing sounds was significant for both mild and severe OSA group ($p< 0.05$ and $p<0.005$, respectively) (Figure 18). It should be noted that we could not be sure the breathing sounds during sleep were nasal or mouth breathing. In addition, there was no significant differences in KFD values of nasal and mouth breathing sounds during wakefulness. There was also a marginally significant ($p<0.06$) difference between KFD of mild and severe OSA during sleep.
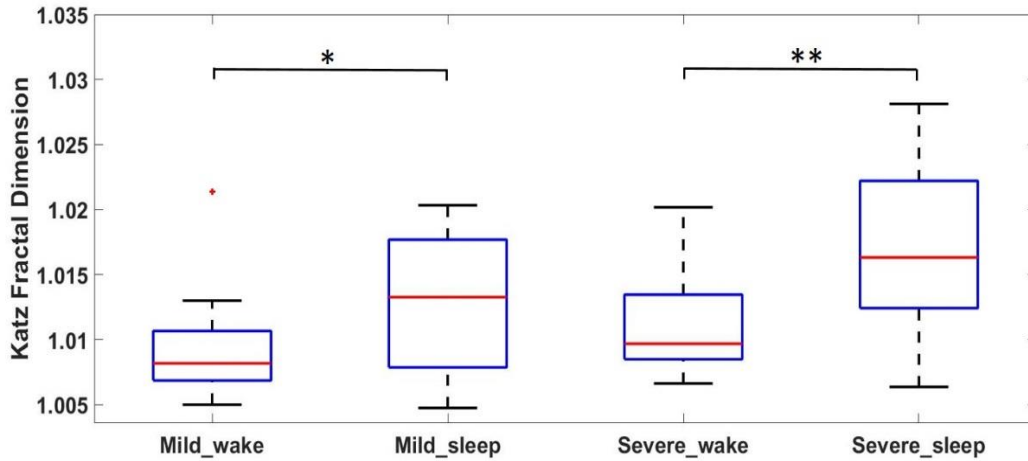


Figure 18 t-test outcome of Katz fractal dimension change from wakefulness to sleep in mild and severe OSA group using nasal breathing sounds during wakefulness.

Table 14 summarizes the score of statistical tests for the 3 frequency-based features and the kurtosis and KFD characteristics of expiratory breathing sounds during both wakefulness and sleep.

Table 14 Statistical test outcomes of the extracted features. (*) shows p-value < 0.05 that considered as significant.

| Feature | Mouth breathing sounds during wakefulness | Nasal breathing sounds during wakefulness |
|---|---|---|
| Difference of the 30%-freq during wakefulness between mild and severe OSA groups | P< .008* | P>0.1 |
| Difference of the 30%-freq during sleep between mild and severe OSA groups | P<0.05* | P<0.05* |
| Difference of the 50%-freq during wakefulness between mild and severe OSA groups | p>0.1 | p>0.1 |
| Difference of the 50%-freq during sleep between mild and severe OSA groups | P<0.02* | P<0.02* |
| Difference of the 70%-freq during wakefulness between mild and severe OSA groups | p>0.3 | p>0.3 |
| Difference of the 70%-freq during sleep between mild and severe OSA groups | P<0.05* | P<0.05* |
| Change of the kurtosis from wakefulness to sleep in mild group | P<0.02* | P<0.007* |
| Change of the kurtosis from wakefulness to sleep in severe group | P<0.04* | P<0.002* |
| Change of the Katz fractal dimension from wakefulness to sleep in mild group | p>0.3 | P<0.05* |
| Change of the Katz fractal dimension from wakefulness to sleep in severe group | p>0.2 | p<0.005* |

## IV.I.4 Discussion

Tracheal breathing sounds analysis is a non-invasive method for studying the pathophysiology of the airway [19]. They have been shown to be affected by the structural and physiological changes of the UA [13]. In this study, we hypothesized that respiratory sounds spectral and higher order statistical characteristics are able to convey valuable information in relation to OSA severity and how they change from wakefulness to sleep.

The UA anatomical structure depends on body position. Thus, during both wakefulness and sleep, we analyzed breathing sounds recorded in supine position. The AHI in supine position is also usually higher than that in lateral body positions [20].

The 30%-freq feature revealed that during wakefulness and sleep the average power at low frequencies of breathings sounds of severe OSA is less than that of the mild OSA. This pattern was expected because different imaging studies [6], [11], [21] and our previous tracheal breathing sound studies [15], [22] have shown that the UAs of severe OSA compared to mild OSA are thicker and more compliant during both wakefulness and sleep. It is known that the low frequency sounds are absorbed more by more compliant tubes [23]. Thus, we expected to see a lower average power at low frequencies during wakefulness and sleep in severe OSA individuals (Figure 15).

Based on the results of the 50%-freq and 70%-freq features, the average PSD during sleep at high frequencies was higher for severe OSA compared to that of mild OSA participants. As an increased high-frequency power represents the increase in stiffness, we speculate that this finding may suggest an increased stiffness of the UAs of severe OSA subjects during sleep. That is congruent with imaging studies results [7], [8] and the tube law [24] that have suggests more regional compliance and stiffness of the UA due to the structural regional changes of the UA shape in severe OSA. In addition, in one study based on electromyography of anesthetized rats, it was shown that simultaneous stimulation of protrudor and retractor muscles of the tongue in the case of hypercapnia and hypoxia, as in severe OSA, leads to tongue retraction and results in a narrower but stiffer pharyngeal airway [25]. Therefore, in agreement with aforementioned studies, the observed spectral pattern at low and high frequencies are representative of OSA pathophysiology during wakefulness and sleep.

The higher order statistical analysis (Kurtosis analysis) showed an increase in kurtosis from wakefulness to sleep in both mild and severe OSA groups, using both mouth and nasal breathing sounds during wakefulness. The kurtosis represents peakedness of the probability distribution of the time series signals. Thus, the higher kurtosis during sleep means the distribution of tracheal expiratory sounds is more clustered around its mean and has relatively small standard deviation. We speculate this might be representative of the higher stiffness of the UA during sleep.

Fractal dimension analysis, including KFD that was used in this study, has often been used as a measure of the complexity of biological signals [26]. Our results showed a significant change in KFD from wakefulness to sleep in both groups of mild and severe OSA. We believe this is in support of the kurtosis results that showed more peakedness during sleep. Thus, this might also be due to higher stiffness of the UA during sleep in OSA individuals.

## IV.I.5 Conclusion

In this study, we investigated the spectral and higher order statistical characteristics of expiratory sounds during both wakefulness and sleep in mild and severe OSA groups. Our results, congruent with imaging studies of the UA in OSA population, are indicative that the UAs of severe OSA are mainly characterized by having more compliance (presented by lower average power at low frequencies) and also more stiffness (presented by higher average power at high frequencies). This implies, there must be a regional narrowing to cause both more compliance and stiffness simultaneously in different regions of the UA. We also observed an increased stiffness during sleep particularly in severe OSA group. In addition, the results of higher order statistical analysis (kurtosis) and complexity measure are indicative of more stiffness as one sleeps; this was observed in both mild and severe OSA groups. Overall, the results of this study, although with a limited

sample size, are encouraging for the use of tracheal breathing sounds for examining UA structural changes due to OSA during both wakefulness and sleep.

## References

[1]     A. Malhotra and D. P. White, "Obstructive sleep apnoea," *Lancet*, vol. 360, no. 9328, pp. 237–245, Jul. 2002.

[2]     J. E. Remmers, W. J. deGroot, E. K. Sauerland, and  a M. Anch, "Pathogenesis of upper airway occlusion during sleep.," *J. Appl. Physiol.*, vol. 44, no. 6, pp. 931–938, 1978.

[3]     T. Young *et al.*, "Sleep disordered breathing and mortality: eighteen-year follow-up of the Wisconsin sleep cohort.," *Sleep*, vol. 31, no. 8, pp. 1071–1078, 2008.

[4]     N. M. Punjabi, "The Epidemiology of Adult Obstructive Sleep Apnea," *Proc. Am. Thorac. Soc.*, vol. 5, no. 2, pp. 136–143, 2008.

[5]     N. Deacon and A. Malhotra, "Potential protective mechanism of arousal in obstructive sleep apnea," *J. Thorac. Dis.*, vol. 8, no. S7, pp. S545–S546, 2016.

[6]     Y. Finkelstein *et al.*, "Velopharyngeal Anatomy in Patients With Obstructive Sleep Apnea Versus Normal Subjects," *J. Oral Maxillofac. Surg.*, vol. 72, no. 7, pp. 1350–1372, Jul. 2014.

[7]     K.-H. Liu *et al.*, "Sonographic measurement of lateral parapharyngeal wall thickness in patients with obstructive sleep apnea.," *J. Sleep*, vol. 30, no. 11, pp. 1503–8, 2007.

[8]     G. C. Barkdull, C. a Kohl, M. Patel, and T. M. Davidson, "Computed Tomography Imaging of Patients With Obstructive Sleep Apnea," *Laryngoscope*, vol. 118, no. 8, pp. 1486–1492, Aug. 2008.

[9]     J. A. Dempsey, S. C. Veasey, B. J. Morgan, and C. P. O'Donnell, "Pathophysiology of Sleep Apnea," *Physiol. Rev.*, vol. 90, no. 1, pp. 47–112, Jan. 2010.

[10]    M. Younes, "Role of respiratory control mechanisms in the pathogenesis of obstructive sleep disorders," *J. Appl. Physiol.*, vol. 105, no. 5, pp. 1389–1405, Nov. 2008.

[11]    R. J. Schwab *et al.*, "Identification of Upper Airway Anatomic Risk Factors for Obstructive Sleep Apnea with Volumetric Magnetic Resonance Imaging," *Am. J. Respir. Crit. Care Med.*, vol. 168, no. 5, pp. 522–530, Sep. 2003.

[12]    R. B. Fogel *et al.*, "The effect of sleep onset on upper airway muscle activity in patients with sleep apnoea versus controls," *J. Physiol.*, vol. 564, no. 2, pp. 549–562, Apr. 2005.

[13]    J. J. Fredberg, "Pseudo-sound generation at atherosclerotic constructions in arteries," *J. Bull. Math. Biol.*, vol. 36, no. 2, pp. 143–155, 1974.

[14]    A. Montazeri, E. Giannouli, and Z. Moussavi, "Assessment of Obstructive Sleep Apnea and its Severity during Wakefulness," *Ann. Biomed. Eng.*, vol. 40, no. 4, pp. 916–924, Apr. 2012.

[15] A. Elwali and Z. Moussavi, "Obstructive Sleep Apnea Screening and Airway Structure Characterization During Wakefulness Using Tracheal Breathing Sounds," *Ann. Biomed. Eng.*, vol. 45, no. 3, pp. 839–850, Mar. 2017.

[16] S. Huq and Z. Moussavi, "Acoustic breath-phase detection using tracheal breath sounds," *Med. Biol. Eng. Comput.*, vol. 50, no. 3, pp. 297–308, Mar. 2012.

[17] A. Yadollahi and Z. M. K. Moussavi, "Acoustical flow estimation: Review and validation," *IEEE Eng. Med. Biol. Mag.*, vol. 26, no. 1, pp. 56–61, Jan. 2007.

[18] M. J. Katz, "Fractals and the analysis of waveforms," *Comput. Biol. Med.*, vol. 18, no. 3, pp. 145–156, 1988.

[19] C. Que, C. Kolmaga, L. Durand, S. M. Kelly, and P. T. Macklem, "Phonospirometry for noninvasive measurement of ventilation: methodology and preliminary results," *J. Appl. Physiol.*, vol. 93, no. 4, pp. 1515–1526, 2002.

[20] L. M. Walter, D. U. Dassanayake, A. J. Weichard, M. J. Davey, G. M. Nixon, and R. S. Horne, "Back to sleep- or not: The impact of the supine position in pediatric OSA," *Sleep Med.*, Jul. 2017.

[21] Z. Lan, A. Itoi, M. Takashima, M. Oda, and K. Tomoda, "Difference of pharyngeal morphology and mechanical property between OSAHS patients and normal subjects," *Auris Nasus Larynx*, vol. 33, no. 4, pp. 433–439, Dec. 2006.

[22] Z. Moussavi, A. Elwali, R. Soltanzadeh, C. A. MacGregor, and B. Lithgow, "Breathing sounds characteristics correlate with structural changes of upper airway due to obstructive sleep apnea," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 5956–5959.

[23] F. Bechwati *et al.*, "Low frequency sound propagation in activated carbon," *J. Acoust. Soc. Am.*, vol. 132, no. 1, pp. 239–248, 2012.

[24] O. E. Jensen, "Flows through deformable airways," in *Centre for Mathematical Medicine School of Mathematical Sciences University of Nottingham, UK*, 2002.

[25] R. F. Fregosi and D. D. Fuller, "Respiratory-related control of extrinsic tongue muscle activity," *Respir. Physiol.*, vol. 110, no. 2–3, pp. 295–306, 1997.

[26] A. Savi, L. Nikoli, S. Budimir, and D. Janoševi, "Applications of Higuchi ' s fractal dimension in the analysis of biological signals Applications of Higuchi ' s fractal dimension in the Analysis of Biological Signals," in *elecommunications Forum (TELFOR)*, 2012, no. November, pp. 639–641.

# IV.II   Acoustic Characterization of Upper Airway Variations from Wakefulness to Sleep with Respect to Obstructive Sleep Apnea

Farahnaz Hajipour, Eleni Giannouli, and Zahra Moussavi

*Abstract* — The upper airway (UA) is in general thicker and narrower in obstructive sleep apnea (OSA) population than in normal. Additionally, the UA changes during sleep are much more in the OSA population. The UA changes can alter the tracheal breathing sounds (TBS) characteristics. Therefore, we hypothesize the TBS changes from wakefulness to sleep are significantly correlated to the OSA severity; thus, they may represent the physiological characteristics of the UA. To investigate our hypothesis, we recorded TBS of 18 mild-OSA (AHI<15) and 22 moderate/severe-OSA (AHI>15) during daytime (wakefulness) and then during sleep. The power spectral density (PSD) of the TBS was calculated and compared within the two OSA groups and between wakefulness and sleep. The average PSD of the mild-OSA group in the low-frequency range (<280 Hz) was found to be decreased significantly from wakefulness to sleep (p-value $<10^{-4}$). On the other hand, the average PSD of the moderate/severe-OSA group in the high-frequency range (>900 Hz) increased marginally significantly from wakefulness to sleep (p-value $<9 \times 10^{-3}$). Our findings show that the changes in spectral characteristics of TBS from wakefulness to sleep correlate with the severity of OSA and can represent physiological variations of UA. Therefore, TBS analysis has the potential to assist with diagnosis and clinical management decisions in OSA patients based on their OSA severity stratification; thus, obviating the need for more expensive and time-consuming sleep studies.

## IV.II.1    Introduction

The upper airway (UA) is a collapsible structure; it dynamically changes from wakefulness to sleep, and also between the sleep stages [1]. The patency of the UA is suggested to be dependent on the equilibrium between the dilating forces generated by the UA dilator muscles and the pressure exerted by the heterogeneous surrounding soft tissue [2]. The complete or partial collapse of the UA during sleep could lead to apnea, a cessation of airflow for $\geq$10 sec, or hypopnea, a reduction of the peak airflow by $\geq$30% from pre-event baseline if it lasts $\geq$10 sec, and is associated with a $\geq$3% oxygen desaturation [3].

Obstructive sleep apnea (OSA) is characterized by repetitive episodes of apnea and/or hypopnea during sleep [4]. OSA is a relatively common disorder that can affect the health of all age groups [5]. Between 9%-38% of the general adult population suffer from OSA [5]. As OSA is still underdiagnosed, these values are believed to underestimate the actual numbers [6]. Untreated OSA is associated with many deficits including excessive daytime sleepiness, increased risk of motor vehicle accidents, memory impairment, and stroke [2], [7]. Untreated moderate/severe OSA is also associated with increased morbidity and mortality [7].

Currently, diagnosis of OSA is based on the full-night Polysomnography (PSG) assessment, as the gold standard, and to some extent by Home Sleep Study (HSS) systems. PSG measures the Apnea-Hypopnea Index (AHI), which reports the average number of apneic events per hour of sleep [8]. However, full PSG studies are time-consuming, laborious, expensive, and usually, have a long waiting time. HSS helps with cost-effectiveness and waiting-time of PSG, yet, it remains as an overnight test that still requires considerable resources for reliable outcomes [9]. Therefore, developing an alternative technology to overcome these difficulties is momentous.

Tracheal breathing sounds (TBS) are a measure of tracheal wall vibration set into motion by the passage of turbulent airflow from the UA including trachea and pharynx [10]. Structural and physiological properties of the UA affect the resonance frequency of the UA that is detected by TBS analysis [11], [12]. It is known that physiological properties of the UA, its patency, and its resistance change dynamically, but with different degrees with respect to the sleep/wakefulness status (different sleep stages/wakefulness) and OSA severity [13]–[18]. Accordingly, our team and a few others around the world are using TBS as a quick, inexpensive, and reliable technology with comparable outcomes to that of PSG or HSS for OSA screening [19]–[23]. However, the changes of TBS from wakefulness to sleep in relationship to the severity of OSA and whether those changes have a classification power to identify the moderate and severe OSA from others have not been investigated adequately; that is the goal of this study.

In a study by Yadollahi et al. [23], a fully automatic acoustic technology was introduced to estimate AHI during sleep using pulse oximetry and TBS. The detected apnea-hypopnea events were highly correlated to those detected using PSG. Although that study reduced the number of required signals to only two, however, it was still a full overnight test. In a study by Hajipour et al. [22], contrarily, a set of TBS characteristics were used to separate OSA individuals from non-OSA individuals during wakefulness. The results showed an average area under the receiver operated characteristic curve (AUC) of 89.9±1.7%. Although the selected features in that study were correlated to AHI, it was not investigated whether those features were robust enough to reflect the changes of the UA during sleep in relation to OSA severity; that is one goal of this study.

On the other hand, according to tube law, maximum airflow ($V_{max}$) in collapsible tubes is determined by

$$V_{max} = \sqrt{A \times \left[\frac{A}{\rho(\Delta A/\Delta P)}\right]}$$

Eq. Iv.II.1

where $A$ is the cross-section area of the tube, $\rho$ is the gas density, and $(\Delta A/\Delta P)$ is the pharyngeal airway compliance; a measure of distensibility of the tube. According to the *Eq. Iv.II.1*, $V_{max}$ has a direct relationship to the cross-section area of the tube while it is inverse for the tube's compliance [24]. By this analogy, the changes in the cross-section area and compliance/collapsibility of the UA are important factors in the pathophysiology of OSA [25] and are potential factors to modify the TBS of OSA compared to that of non-OSA individuals. If TBS analysis can reveal these anatomical/physiological characteristics of the UA, it will be an excellent non-invasive, quick, and cost-effective alternative tool for OSA management and it can lead to more appropriate therapeutic decisions [26].

In our previous and relevant study [27], we recorded TBS of 30 OSA participants during sleep and in two maneuvers of mouth and nasal breathing during wakefulness. Next, we compared the spectral and higher-order statistical characteristics of their TBS during stage 2 of sleep and then during wakefulness. The results showed a significant difference between the TBS characteristics of these two OSA groups during either wakefulness or sleep. However, in that study, we did not investigate the pattern of changes from wakefulness to sleep in each of the two groups, nor investigated whether those changes have any classification power for identifying the severity of OSA. We hypothesize the changes in acoustic properties of TBS from wakefulness to sleep are highly correlated with the severity of OSA; thus, they are useful in OSA screening and classification, and may reveal the changes in physiological characteristics of the UA (including narrowing, thickness and resistance) due to OSA in a detailed but straightforward manner. Therefore, in this study using a larger database, our primary goal was to investigate TBS changes

111

from wakefulness to sleep in individuals with different levels of OSA severity and to explore whether these changes are correlated with their OSA severity. We also assessed the classification power of these acoustic features to separate the two OSA groups for screening purposes. To achieve our goal, we analyzed and compared spectral characteristics of TBS recorded during wakefulness (a combination of mouth and nasal TBS) to those during sleep. We discuss the physiological interpretation of our findings and describe their ability to show the UA characteristics changes regarding OSA.

## IV.II.2 Method

### IV.II.2.1 Participants

Sixty individuals referred for PSG assessment at Sleep Disorder Center of Misericordia Health Center (Winnipeg, Canada) participated in this study. The Biomedical Research Ethics Board of the University of Manitoba approved the study. All participants signed an informed consent before data collection. We excluded data of 20 participants from the study due to frequent noises (including vocal and blanket, audible alarms and air conditioner noises) or the need of patients for a titration that the sleep technician disconnected our acoustic device.

### IV.II.2.2 Sound Recording Procedure

The TBS during both wakefulness and sleep were collected by a miniature microphone (Sony ECM-77B) inserted into a small chamber, allowing 2 mm cone-shaped space with skin. The chamber was mounted over the suprasternal notch of the trachea with double-sided adhesive tape. To ensure the microphone would not be misplaced during the night, we used a soft neckband, which was sealed softly around the patient's neck to sustain microphone and chamber in site

(Figure 19). The sounds were band-pass filtered in the range of [0.05-5000 Hz], amplified by a Biopac (DA100C) amplifier, and digitized at 10240 Hz sampling rate.



<div align="center">(a)           (b)</div>
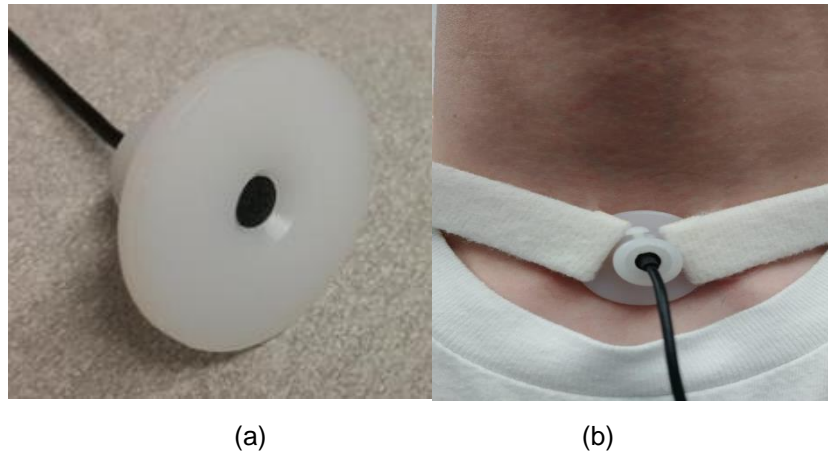
Figure 19 a) The Microphone within our custom-made chamber with 2 mm cone-shaped space with skin. b) The Microphone and chamber that sealed around the suprasternal notch of the trachea of patients using neckband.

TBS during wakefulness was recorded before PSG assessment. We recorded 5 cycles of normal TBS through the nose with mouth closed, followed by 5 cycles of normal TBS through the mouth with a nose clip in place. We choose to record 5 breaths to be a representative of steady-state pattern of breathing without acceleration/deceleration of respiratory rate. After wakefulness recording, participants were prepared for PSG assessment. TBS during sleep were recorded simultaneously with the PSG assessment; they were real-time with the snoring sounds and respiratory events as appearing in the PSG. TBS segments for analysis were extracted from stable sleep periods in stage N2, void of snoring sounds or artifacts and in the supine position (determined using the PSG score sheet). Using the AHI outcome of the PSG assessment, we grouped the participants into mild-OSA (AHI <15, N=18) and moderate/severe-OSA (AHI >15, N=22) groups. The threshold of AHI=15 has been traditionally used in many studies to identify OSA patients who might have increased cardiovascular or mortality risks and are in need of treatment [7]. Therefore, this threshold could potentially require more focus and earlier assessment for treatment.

Anthropometric information of the 40 individuals, whose data were analyzed in this study, is presented in Table 15.

## IV.II.2.3    Preprocessing and Signal Analysis

In this study, we did not record the respiratory flow of participants, however, to ensure the respiratory phases, all recording procedures during wakefulness started at the inspiration and marked by the voice of the experimenter. Using that auditory marker, the inspiratory/expiratory phases during wakefulness were separated manually. For the TBS of sleep, however, we used our semi-manual technique, elaborated in [28] to identify the inspiratory/expiratory phases of breathing. In this study, we aimed to compare TBS during wakefulness with those during sleep. As there are no snoring sounds during wakefulness, to have a fair comparison between wakefulness and sleep, we examined all the recorded TBS data by audio and visual means in the time-and-frequency domain to exclude TBS with snoring sounds and noisy signals (including artifacts, vocal noises, and swallowing).

The majority of the moderate/severe-OSA participants of this study snored most of the time in the supine position. Since snoring usually occurs in the inspiration, because the UA collapse typically occurs at the end of expiration [14], the majority of the inspiratory phases of moderate/severe-OSA group has been eliminated. Therefore, we decided to remove the inspiratory phases of breathing for all participants to have a fair comparison between mild and moderate/severe OSA individuals. Furthermore, since there is no apneic event during wakefulness for even the moderate/severe OSA individuals [24], we also excluded sleep data with respiratory apneic events. Consequently, from the recorded sounds during sleep, we selected 5 normal (free of any apneic events including flow limitation), noise- and snore-free expiratory sounds in supine position and sleep stage 2 for further analysis. Data in stage 2 of sleep was selected because that

was the most common (i.e., the highest number of individuals' data) sleep stage in our dataset. We were also interested only in the supine the position to match with position of data collection during wakefulness. From data recorded during wakefulness, we also selected 5 noise-free expiratory sounds for comparison to those extracted from data during sleep.

In this study, similar to our previous research [22], each selected sound was first passed through a $5^{th}$-order Butterworth band-pass filter in the range of [75-2500 Hz] to eliminate the effect of low- and high-frequency noises (including ambient noises, highest amplitude component of the heartbeats (which is less than 75 Hz [29]), fundamental frequency of the power line (60 Hz in Canada), and muscle sounds) while keeping the primary frequency component of the sound. Next, each filtered sound signal was normalized by its variance envelope (a smoothed version of itself using the Moving Average method with 64 sample sequence ~ 6ms) [30] to remove its extra fluctuations, and then by its standard deviation to compensate for plausible different flow rates between breathing cycles. Then, the logarithm of the variance of each TBS was calculated to acoustically estimate its respiratory flow. Afterward, the 50% duration around the maximum of the estimated flow signal was considered as the stationary portion of that sound signal [31]. Finally, using Welch's method, the power spectrum density (PSD) of the stationary portion of sound signals were estimated in segments of 20 ms with 50% overlapping windows to successive segments. Figure 20 outlines the above-mentioned pre-processing.
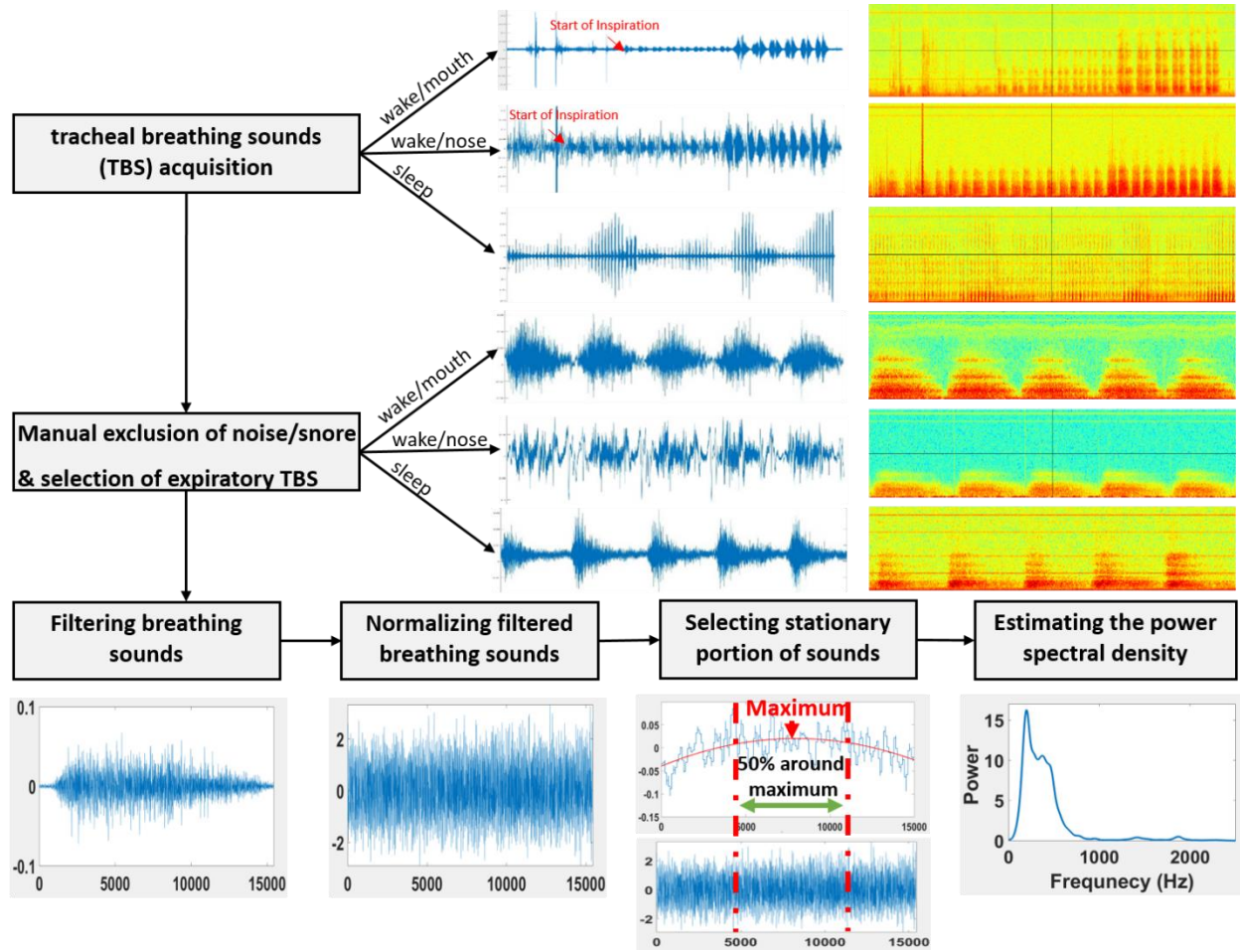
Figure 20 Pre-processing and signal analysis framework. Wakefulness TBS were recorded in two maneuvers of mouth and nasal breathing. The sleep data were recorded in >6 hours of sleep. The sleep file in this figure is a random 3 minutes of sleep out of the total sleep time of a participant.

As mentioned in Section IV.II.2.2, during wakefulness, we recorded TBS in two maneuvers of nose and mouth breathing. During sleep, however, one could be breathing through either nose or mouth, but without a video recording, we could not identify nose or mouth breathing. Thus, to have a fair comparison of wakefulness and sleep data, for each participant, we selected 3 breaths from mouth breathing and 2 breaths from nasal breathing of the wakefulness. Next, we considered the average of the estimated PSD of these signals and the average of the estimated PSD of the five TBS recorded during sleep as the representative data of wakefulness and sleep of each individual, respectively.

116

For feature extraction with the purpose of investigating TBS variations from wakefulness to sleep, within each OSA groups we averaged the PSD signals during wakefulness as well as sleep and calculated their standard error (SE). We also assessed the difference between the average PSD ($PSD_{avg}$) of TBS during wakefulness and sleep in two groups of mild (*mild-Difference*) and moderate/severe OSA (*m/s-Difference*). The regions with no overlap between the *mild-Difference* and the *m/s-Difference* were considered as characteristic regions for extracting features for further statistical investigations. These regions reflect the frequency bands that the spectral characteristics of TBS change the most from wakefulness to sleep and are introduced in Section IV.II.3. Similar to our previous studies during wakefulness [22], [32], we considered the mean of these areas as characteristic features to be selected for classification. TBS features representing changes from wakefulness to sleep were the mean of the *mild-Difference* and *m/s-Difference* in the ranges mentioned above. TBS features during wakefulness and sleep were the mean of the $PSD_{avg}$ in the pre-mentioned regions for the two OSA groups during wakefulness and sleep, respectively. To be clearer, Figure 21 shows the *mild-Difference* and *m/s-Difference* as well as the $PSD_{avg}$ during wakefulness and sleep for both mild and moderate/severe OSA groups. TBS features were extracted from the regions between the solid lines as well as the region between the dotted lines.

In addition to studying changes of TBS characteristics from wakefulness to sleep, we also studied and compared the classification ability of the TBS features reflecting changes from wakefulness to sleep (Figure 21-c) with those extracted during only wakefulness or only sleep (Figure 21-a and Figure 21-b). To perform classification, we used 10 ensembles of 2-class linear kernel support vector machine (SVM) classifiers, obtained from bootstrap samples of our data. Bootstrapping consists of repeatedly drawing samples of the same size, with replacement, from our original dataset. We trained our SVM classifier with the bootstrap samples. The remaining

data (samples that were not selected in the bootstrapping procedure) were considered as testing data and used to evaluate the performance of the SVM classifiers. We repeated this procedure 10 times and reported the average classification results.

### IV.II.2.4　Statistical Analysis

Our hypotheses of this study were the spectral characteristics of TBS change from wakefulness to sleep, and these variations are different in OSA and non-OSA populations. As we have measured the TBS characteristics over two time-points (wakefulness and sleep), also our participants have been assigned in two groups (mild and moderate/severe OSA), we applied a multivariate two way mixed analysis of covariance (mixed ANCOVA) multifactorial statistical test to compare the within-groups' changes (from wakefulness to sleep) and the between-groups' changes (between mild and moderate/severe OSA groups), while considering the neck circumference (NC) as covariate. We included the NC as covariate in ANCOVA model to ensure that our TBS features are not just measuring the differences in NC between the two OSA groups. A p-value of 0.05 was considered significant. Next, we used simple effect post-hoc analysis with p-value of 0.0064 to determine the effect of sleep/wakefulness status on OSA severity, and vice versa. The reason for choosing this p-value is explained in Section IV.II.3. Pearson's correlation was used to determine the correlations among the TBS features and AHI. Independent unpaired t-test with p-value of 0.05 was also used to compare the anthropometric information and the differences of $PSD_{avg}$ of wakefulness and sleep in mild-OSA and moderate/severe-OSA groups.

### IV.II.3　Results

Table 15 shows the average and standard deviations of the anthropometric information of the mild and moderate/severe OSA groups, as well as their basic statistical comparisons. There was

118

no significant difference between the two groups in terms of age, height and sex. The participants with moderate/severe-OSA had a significantly higher weight, AHI, body mass index (BMI) and NC values than did those of mild-OSA individuals.

Table 15 Anthropometric information's mean and standard deviations (SD) and basic statistical comparisons between mild and moderate/severe obstructive sleep apnea (OSA) groups. The data of moderate/severe-OSA group was compared to the mild-OSA group with independent sample t-test and Chi square test. (*) shows p-value < 0.05 that considered as significant. BMI refers to body mass index; NC, neck circumference; AHI, apnea-hypopnea index.

| | Mild-OSA (N=18) Mean ± SD | Moderate/severe-OSA (N=22) Mean ± SD | p-value |
|---|---|---|---|
| **Age (year)** | 41.7±14.4 | 49.5±11.2 | 0.07 |
| **Weight ($kg$)** | 87.5±16.6 | 99.5±17 | 0.03* |
| **Sex (male: female)** | 14:4 | 20:2 | 0.47 |
| **Height ($cm$)** | 173.5±7.8 | 174.8±6.5 | 0.59 |
| **BMI ($kg/m^2$)** | 29.1±5.4 | 32.6±5.4 | 0.046* |
| **NC ($cm$)** | 40.9±3.7 | 44.7±2.8 | 0.002* |
| **AHI** | 4.1±4.5 | 60.3±32.5 | $5.8\times 10^{-8}$* |

To predict OSA and its pathology by breathing sounds analysis, in this study, we focused on acoustical changes from wake to sleep. In general, in our studies we aim to have a reliable acoustic OSA prediction by having breathing sounds of a short period of time. From the sleep data, as mentioned in Methods, we decided to select Stage 2 of sleep data. On average, our participants spent 106.36 minutes before reaching Stage 2 of sleep in supine position; this average time was 114.1 and 98.6 minutes for mild-OSA and moderate/severe-OSA participants, respectively. During these times, the participants slept either on other postural positions or at stage 1 of sleep. Stage 1 data was not selected because the majority of our participants snored and/or had several

episodes of hypopneas. The average time to reach REM or stage 3 and 4 of sleep was longer than that of stage 2. It worth to mention that not all of our participants reached REM or deep sleep stages.

Figure 21-a and Figure 21-b show the $PSD_{avg}$ and SE intervals during wakefulness and sleep for mild and moderate/severe OSA participants, respectively. These figures depict clear differences in low- and high-frequency ranges between $PSD_{avg}$ of TBS recorded during wakefulness and sleep and within two OSA groups. Figure 21-c demonstrates the *mild-Difference* and *m/s-Difference*, in addition to their SE intervals. As Figure 21-c shows, among the regions with no overlap between the two OSA groups, at lower frequencies [150-280 Hz], the *mild-Difference* were significantly higher than *m/s-Difference*, (p-value <0.03, Table 16). Contrarily, at higher frequencies [950-1150 Hz] the *m/s-Difference* values were significantly higher compared to those of *mild-Difference* (p-value <0.04, Table 16).

Table 16 The average of the spectral features in association with their standard deviations (std) for the mild and moderate/severe obstructive sleep apnea (OSA) groups. $PSD_{avg}$: Average power spectra; F1: $PSD_{avg}$ in the low-frequency range of [150-280 Hz]; F2: $PSD_{avg}$ in the high-frequency range of [950-1150 Hz].

| Features (Watt) | Mild OSA ± std | Moderate/severe OSA ± std |
|---|---|---|
| Differences of F1 of wakefulness and sleep | 3.1±0.65 | 0.97±0.72 |
| Differences of F2 of wakefulness and sleep | -0.25±0.2 | -1.1±0.36 |
| F1 during wakefulness | 10.8±0.48 | 8.6±0.6 |
| F1 during sleep | 7.7±0.49 | 7.6± 0.6 |
| F2 during wakefulness | 0.55±0.13 | 0.54±0.06 |
| F2 during sleep | 0.8±0.19 | 1.7±0.36 |

(a)

(b)

(c)
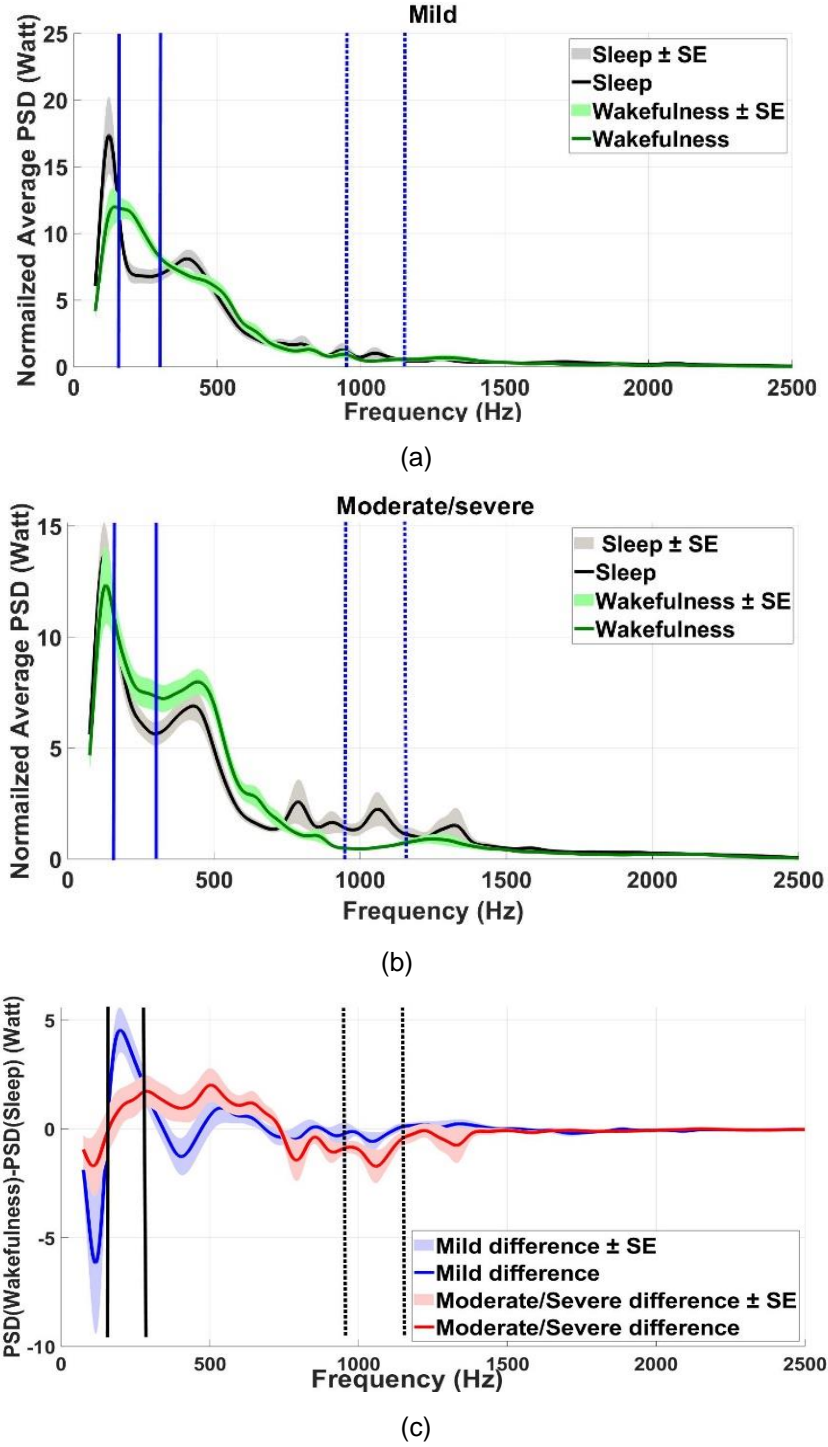
Figure 21 Average power spectra (PSD$_{avg}$) of combination of mouth and nasal tracheal breathing sounds (TBS) during wakefulness and sleep with their standard error intervals (shadows). A) Averaged among participants of mild-OSA group during wakefulness (green) and sleep (black), B) Averaged among participants of moderate/severe-OSA group during wakefulness (green) and sleep (black), C) difference of the PSD$_{avg}$ during wakefulness and sleep for mild-OSA (blue) and moderate/severe-OSA (red) groups. The areas between solid lines and dotted lines show the regions where the features were extracted.

Based on the aforementioned observations, we considered *the $PSD_{avg}$ in the low-frequency range of [150-280 Hz] (F1)* and *the $PSD_{avg}$ in the high-frequency range of [950-1150 Hz] (F2)* as the two potential features representing the most changes from wakefulness to sleep for further analysis. Since we compared the changes of F1 and F2 from wakefulness to sleep and also between the two OSA groups, we had 8 different statistical comparisons. Therefore, to satisfy the statistical significance of 95% for the overall post-hoc tests, the significance level of the p-value of each post-hoc test was considered as $1 - (1 - 0.05)^{1/8} \cong 0.0064$ [33].

The mixed ANCOVA multifactorial test result showed a highly significant main effect of sleep/wakefulness status (being asleep or awake) on the TBS features, (p-value $<10^{-4}$, Table 17). This test also showed a significant interaction between the sleep/wakefulness status and the OSA severity level of participants (p-value $<2 \times 10^{-2}$, Table 18).

Table 17 The score of the main effect of sleep/wakefulness status over features extracted from average power spectra ($PSD_{avg}$), using tracheal breathing sounds (TBS) during sleep and combination of mouth and nasal TBS during wakefulness. The neck circumference considered as covariate variable in the Mixed ANCOVA test. (*) shows p-value < 0.05 that considered as significant. F1: the $PSD_{avg}$ in the low-frequency range of [150-280 Hz]; F2: the $PSD_{avg}$ in the high-frequency range of [950-1150 Hz].

| | **Main effect (p-value)** |
|---|---|
| Sleep/wakefulness status on TBS features | $F(2,36) = 11.7, (< 10^{-4})$ * |
| Sleep/wakefulness status on F1 | $F(1,37) = 18.6, (< 10^{-4})$ * |
| Sleep/wakefulness status on F2 | $F(1,37) = 11.9, (< 10^{-3})$ * |

Table 18 The score of the interaction effect of sleep/wakefulness status and OSA severity level over features extracted from average power spectra ($PSD_{avg}$), using tracheal breathing sounds (TBS) during sleep and combination of mouth and nasal TBS during wakefulness. The neck circumference considered as covariate variable in the Mixed ANCOVA test. (*) shows p-value <

0.05 that considered as significant. F1: the $PSD_{avg}$ in the low-frequency range of [150-280 Hz]; F2: the$PSD_{avg}$ in the high-frequency range of [950-1150 Hz].

| | Interaction effect (p-value) |
|---|---|
| Sleep/wakefulness status and OSA severity level | $F(2,36) = 4.7, (< 2 \times 10^{-2})$ * |
| Sleep/wakefulness status and OSA severity level for F1 | $F(1,37) = 7.9, (< 8 \times 10^{-3})$ * |
| Sleep/wakefulness status and OSA severity level for F2 | $F(1,37) = 0.2, (< 0.6)$ |

Our results on the investigation of F1 depicts both a highly significant main effect of sleep/wakefulness status (p-value $<10^{-4}$, Table 17) and a significant interaction effect of sleep/wakefulness status and OSA severity level on this feature (p-value $<8 \times 10^{-3}$,

Table 18). The simple effect post-hoc analysis showed a strongly significant decrease in F1 from wakefulness to sleep within the mild group (p-value $<3 \times 10^{-5}$, Figure 21-a, Table 16 and Table 19). A decrease was also observed in moderate/severe-OSA groups, but it was not significant (p-value =0.2, Figure 21-b, Table 16 and Table 19).

Table 19 The p-values of the simple effect post hoc analysis for features extracted from average power spectra ($PSD_{avg}$), using breathing sounds during sleep and combination of mouth and nasal breathing sounds during wakefulness. The neck circumference considered as covariate variable in the Mixed ANCOVA test. (*) shows p-value $< 0.0064$ that considered as significant. F1: the $PSD_{avg}$ in the low-frequency range of [150-280 Hz]; F2: the $PSD_{avg}$ in the high-frequency range of [950-1150 Hz].

| | The simple effect post-hoc analysis p-value |
|---|---|
| Change of F1 from wakefulness to sleep in mild group | $< 3 \times 10^{-5}$* |
| Change of F1 from wakefulness to sleep in moderate/severe group | $> 0.0064$ |
| Change of F2 from wakefulness to sleep in mild group | $> 0.0064$ |
| Change of F2 from wakefulness to sleep in moderate/severe group | $< 9 \times 10^{-3}$ |

| | |
|---|---|
| F1 during wakefulness between mild and moderate/severe groups | < 0.07 |
| F1 during sleep between mild and moderate/severe groups | > 0.0064 |
| F2 during wakefulness between mild and moderate/severe groups | > 0.0064 |
| F2 during sleep between mild and moderate/severe groups | > 0.0064 |

Our results on the investigation of F2 depicts a significant main effect of sleep/wakefulness status (p-value$<10^{-3}$, Table 17) but no significant interaction effect of sleep/wakefulness status and OSA severity level was observed(p-value =0.6, Table 18) . The simple effect post-hoc analysis showed a marginally significant increase in this feature from wakefulness to sleep in moderate/severe OSA group (p-value $< 9\times 10^{-3}$, Figure 21-b, Table 16 and Table 19). Nevertheless, no significant change was observed for the mild group (p-value =0.2, Figure 21-a, Table 16 and Table 19).

Table 20 shows the Pearson correlation coefficients between the AHI and the spectral TBS features. AHI was correlated with F1 during wakefulness (-0.42, p-value <0.01), with F2 during sleep (0.36, p-value <0.05) and with the change of F2 from wakefulness to sleep (-0.39, p-value <0.05).

Table 20 Pearson Correlation between the AHI and features extracted from the average power spectra ($PSD_{avg}$), using breathing sounds during sleep and combination of mouth and nasal breathing sounds during wakefulness. r is the correlation coefficient. (*) shows p-value < 0.05 that considered as significant. F1: the $PSD_{avg}$ in the low-frequency range of [150-280 Hz]; F2: the $PSD_{avg}$ in the high-frequency range of [950-1150 Hz].

| Features | r (p-value) |
|---|---|
| F1 during wakefulness | $-0.42(< 0.01)$ * |
| F1 during Sleep | $-0.29(> 0.05)$ |
| F2 during wakefulness | $-0.09(> 0.05)$ |
| F2 during sleep | $0.36(< 0.05)$ * |

| | |
|---|---|
| Change of F1 from wakefulness to sleep | $-0.12 (> 0.05)$ |
| Change of F2 from wakefulness to sleep | $-0.39 (< 0.05)$ * |

Table 21 reports the percentage of the participants that their F1 and F2 decreased or increased from wakefulness to sleep within the two OSA groups and for the total participants of this study. Overall, F1 decreased from wakefulness to sleep in 77.5% of participants, and F2 increased from wakefulness to sleep in 67.5% of participants.

Table 21 Number and Percentage (%) of mild-OSA, moderate/severe-OSA, and overall participants that their spectral characteristics decreased/increased from wakefulness to sleep. F1: the $PSD_{avg}$ in the low-frequency range of [150-280 Hz]; F2: the $PSD_{avg}$ in the high-frequency range of [950-1150 Hz].

| | Decrease in F1 (%) | Increase in F2 (%) |
|---|---|---|
| **Mild-OSA (n=18)** | 16 (89.9%) | 12 (66.7%) |
| **Moderate/severe-OSA (n=22)** | 15 (68.2%) | 15 (68.2%) |
| **Overall (n=40)** | 31 (77.5%) | 27 (67.5%) |

Table 22 reports the classification results of the SVM classifiers using TBS features' changes from wakefulness to sleep, TBS features during wakefulness, and TBS features during sleep. The SVM classifier using the F1 and F2 changes from wakefulness to sleep (i.e. F1$_{wakefulness}$-F1$_{sleep}$, and F2$_{wakefulness}$-F2$_{sleep}$) resulted in 87.5$\pm$4.5% accuracy with the sensitivity (specificity) of 87.5$\pm$6.9% (87.5$\pm$13%). On the other hand, the SVM classifier using F1 and F2 extracted from wakefulness resulted in 78.8$\pm$4.2% accuracy with the sensitivity (specificity) of 85.7$\pm$1.7% (70.8$\pm$6.3%), and using F1 and F2 extracted from sleep data resulted in 70.1$\pm$6.5% accuracy with the sensitivity (specificity) of 75.7$\pm$16.6% (63.3$\pm$8.5%).

Table 22 The classification results of the SVM classifier over the tracheal breathing sounds (TBS) features using TBS during sleep, and the combination of mouth and nasal TBS during wakefulness. From wakefulness to sleep features are the difference between F1 during wakefulness and sleep as well as the difference between F2 during wakefulness and sleep. In the wakefulness feature are F1 and F2 during wakefulness. In the sleep feature are F1 and F2 during sleep. F1: the $PSD_{avg}$ in the low-frequency range of [150-280 Hz]; F2: the $PSD_{avg}$ in the high-frequency range of [950-1150 Hz].

|  | Specificity (%) | Sensitivity (%) | Accuracy (%) |
|---|---|---|---|
| **From wakefulness to sleep** | $87.5\pm13$ | $87.5\pm6.9$ | $87.5\pm4.5$ |
| **Wakefulness** | $70.8\pm6.3$ | $85.7\pm1.7$ | $78.8\pm4.2$ |
| **Sleep** | $63.3\pm8.5$ | $75.7\pm16.6$ | $70.1\pm6.5$ |

## IV.II.4    Discussion

TBS are affected by structural and physiological properties of the UA and contain rich information of the airway structure [11]. We hypothesize the TBS changes from wakefulness to sleep are highly correlated to the severity of OSA; thus, they may reveal UA structural and physiological characteristics and their variations due to OSA; thus, they may be used as a screening/diagnostic tool to identify the severity of OSA. While there have been many investigations to examine TBS concerning OSA including our team's previous studies [22], [23], [32], [34], [35], to the best of our knowledge, this study is the first to explore the potential role of TBS analysis in assessing UA changes from wakefulness to sleep with respect to OSA and its severity.

The findings of our TBS investigations manifested a significant main effect of sleep/wakefulness status on the TBS spectral characteristics and a significant interaction between the sleep/wakefulness conditions and the OSA severity level of participants. These results imply that the TBS spectral features change significantly, when one sleeps, and these changes are clearly

different within mild and moderate/severe OSA groups. In addition, our results have shown that these spectral features are correlated to the OSA severity level of the participants (Table 20).

One of this study's objectives was to investigate whether TBS characteristics have the potential to reveal the physiological and structural changes of UA regarding OSA and its severity. Imaging studies during wakefulness have reported individuals with OSA compared to non-OSA individuals have thicker velum, thicker pharyngeal wall, a reduced pharyngeal cross-sectional area, and a narrower pharyngeal lumen [15], [16]. According to the Hagen-Poiseuille's law, a narrower UA implies more resistance to airflow. Higher resistance absorbs more energy, indicating lower average PSD at low frequencies (F1) for the generated TBS of the moderate/severe-OSA group compared to that of the mild-OSA group during wakefulness. This phenomenon can be seen clearly by comparing the wakefulness $PSD_{avg}$ of mild and moderate/severe OSA groups in Figure 21-a and Figure 21-b. The predominant energy absorbency at low frequencies is related to the higher net parallel impedance of the UA at lower frequencies. According to the cascade T circuits modeling of UA proposed in the study by Harper et al. [36], the net impedance of the UA is due to the UA walls tissues' capacitance effect in parallel to the UA resistance. These findings might also conclude a thicker UA for moderate/severe OSA group compared to the mild-OSA group during wakefulness. This inference is based on findings of an acoustic study that shows the low-frequency sound absorbency of materials is directly related to the materials' thickness [37]. Therefore, the lower the F1 of moderate/severe OSA group during wakefulness (Table 16), the higher their UA sound absorbency, and therefore the thicker their UA.

On the other hand, studies that investigated UA during sleep have been shown that the sleep would lead to a decrease in the airway caliber and an increase in the UA resistance [12]. As we mentioned earlier, higher resistance absorbs more energy at lower frequencies. Therefore, we

expect a reduction in the F1 of both mild and moderate/severe OSA groups from wakefulness to sleep, as seen in Figure 21-a and Figure 21-b. As Figure 21-c shows, the intensity of change from wakefulness to sleep is different in the two OSA groups. In moderate/severe-OSA group, the changes of the low-frequency PSD from wakefulness to sleep was much less pronounced compared to that of the mild-OSA group. We speculate this might be related to the narrower UA of moderate/severe OSA groups compared to that of mild group during wakefulness [15], [16].

Moreover, it has been shown that the UA collapse during sleep does not happen in the entire pharynx, but it is non-homogeneous and may narrow regionally at multiple anatomic locations with various degrees [16]. Tortuosity and flow resistivity are reasons for a drop of high-frequency sound absorbency [37], indicating an increase in the energy. Therefore, we expect to see an increase in high-frequency power (F2) of moderate/severe participants from wakefulness to sleep, as seen in Figure 21-b. Another reason for observing a higher F2 during sleep and a marginally significant increase of F2 from wakefulness to sleep for moderate/severe OSA group compared to those of mild-OSA is the increased stiffness of their UA [38]. Stiffness effect is reflected at higher frequencies as it increases the wave velocity [39]. Interestingly this higher stiffness of the UA of the moderate/severe-OSA group shows itself more during sleep than wakefulness (Figure 21-b); it is in keeping with the rational expectation that as the airway caliber decreases during sleep, its stiffness is increased [12].

It worth mentioning that the findings of the present study are congruent with our previous study results using different datasets [22]. In that study we used TBS data of many more participants than this current study; data were recorded only during wakefulness, and we considered 80% of data for training and 20% of data as blind testing for evaluation. Using the training data, we considered the non-overlapped regions between the averaged PSD of the mild and moderate/severe

OSA participants as potential areas for feature extraction. To eliminate the bias, we repeated this procedure 10 times and selected the common non-overlapped regions between various training sets. The non-overlapped regions of that study were common with the selected regions of this current study. Thus, the findings of that study, congruent with results of this current study, indicate a lower average power at lower frequencies and higher average power at higher frequencies for moderate/severe OSA individuals during wakefulness compared to those of non-OSA.

TBS power spectra do have variability from person to person. We speculate PSD analysis of the TBS and how its pattern changes (low versus high frequencies) from wakefulness to sleep may be representative of the shape of the individual's velopharyngeal narrowing. For example, in a study by Finkelstein et al. [15], it was shown that the majority (85%) of the OSA participants had a predominantly velopharyngeal narrowing, while that narrowing was observed only in 12.5% of the control group. Narrowing of the pharyngeal area results in higher compliance leading to more collapsibility in the OSA population. In [15], it was also found that the shape of the narrowed airway in OSA subjects was predominantly bottle-shaped due to the increase in dilator-muscles activities that work on maintaining airway patency during wakefulness. That finding suggests regional stiffening and more compliance in the UA of the OSA group. Accordingly, the UA of OSA individuals may show both more stiffness and compliance compared to non-OSA individuals. Higher compliance is represented by a decrease in F1 (low-frequencies decreased power) from wakefulness to sleep that was observed in 89.9% of mild and 68.2% of moderate/severe-OSA (Table 21, Figure 21). Higher stiffness in the UA due to narrowing is representative of an increase in F2 (high-frequency increased power) from wakefulness to sleep that was observed in 66.7% of mild and 68.2% of moderate/severe OSA (Table 21, Figure 21). To prove this directly though, we will need simultaneous recording of TBS with UA imaging and pressure measurements.

In the classification point of view, we used an ensemble of linear kernel SVM classifiers to classify the participants as either mild or moderate/severe OSA (Table 22). The classification results of these classifiers over $F1_{wakefulness}$-$F1_{sleep}$, and $F2_{wakefulness}$-$F2_{sleep}$ resulted in the average accuracy of $87.5\pm4.5\%$. This classification result was higher than the average accuracy associated to the SVM classifiers over F1 and F2 during only wakefulness ($78.8\pm4.2\%$) and only sleep ($70.1\pm6.5\%$). These findings indicate the superiority of investigating the TBS changes from wakefulness to sleep rather than only sleep or wakefulness for OSA screening and differentiating between the two OSA groups. According to the mentioned points, one reason for the better classification results of the features reflecting changes from wake to sleep is the increased stiffness and thickness of the UA of moderate/severe-OSA group from wakefulness to sleep compared to those of mild group; such increased stiffness and thickness manifest themselves in the acoustical properties of TBS more significantly. The selected features of this study are following the features of our previous study during wakefulness [22]; therefore, confirm the robustness of those characteristics for OSA screening during wakefulness. In addition, it should be note that the sleep data in our study requires only a few breaths during sleep. Therefore, even with a short period of sleep, our proposed TBS analysis during both wakefulness and sleep can be useful for differentiating the two OSA severity groups accurately and reliably; thus, reducing the need for the entire night sleep study.

The results of this study also showed a significantly higher weight, BMI, and NC for the moderate/severe-OSA group compared to those of the mild-OSA group (Table 15). These observations are in accordance with the studies that suggest OSA might be a partial indication of the effect of general processes like increase in BMI on the UA [15]. These findings support our

team's previous study that used the anthropometric information as characteristic features for OSA screening during wakefulness, and achieved the test accuracy of more than 76% [32].

Lastly, in this study, we used data of the subjects who were referred for PSG assessment in a hospital. Therefore, it may be thought that the findings presented here may not be reproducible at other hospitals/institutions or homes with different level of ambient noise compared to the sleep labs. However, as we used a firm preprocessing step to eliminate the effect of noises and artifacts, the different noise level would not be a limitation factor; hence, the result of this study could be generalizable in other sleep lab environments and home sleep studies.

## IV.II.4.1    Limitation of the study

The main limitation of this study is the lack of a direct measure of the UA collapsibility of the study participants. Therefore, we compared our results with the general finding of the imaging studies. Another limitation of this study was related to investigating the TBS in only stage 2 of sleep. It was because not all of our study participants had enough data in other sleep stages that were also in supine position. Furthermore, we only had high-quality sounds recorded for a few hours and not the entire night data. In future studies, we will investigate TBS in various sleep stages in relation to OSA severity. It would also be interesting to analyze the transition period of wakefulness to sleep (stage 1) as that may show the dynamic of the UA changes better; that is a goal of our future studies. Another limitation is related to the identification of mouth-nose breathing during sleep. It would be beneficial to record the respiratory flow of individuals using a nasal cannula to help in automatic mouth-nose breathing identification. It is true that nasal cannula does not register mouth breathing and it may look like an apnea episode; however, it is possible to distinguish mouth breathing from apnea mouth-breathing by sound analysis. The other limitation

of the study is the limited number of participants. With more participants, it would be desirable to investigate the potential of TBS analysis to differentiate severe and moderate groups of OSA.

## IV.II.5    Conclusion

In this study, we investigated the application of spectral characteristics of TBS to reveal the pathophysiology of the UA and their change due to OSA. We also studied the relationship of the changes of these characteristics in correlation with the severity of OSA. Our results show significant differences in spectral characteristics of TBS between the mild- and moderate/severe-OSA groups during wakefulness, sleep, and from wakefulness to sleep; congruently indicative of changes in UA thickness and regional collapsibility. Consequently, spectral characteristics of the high sampling rate TBS during wakefulness, sleep and their changes from wakefulness to sleep have potential to reveal the pathophysiology of the UA in relation to OSA. The findings of this study are especially useful to find the TBS characteristics that indirectly and non-invasively reveal the structural changes of the UA in relation to OSA. They are also beneficial to enhance the current OSA diagnosis methods to stratify the severity of OSA patients in a non-laborious, non-time-consuming, and less expensive manner using a short period of sleep instead of full overnight sleep study. From sleep data, we only considered a few normal breathing sounds at stage 2 of sleep. Since Stage 2 is among the first sleep stages to reach when one sleeps, the proposed technology does not have to be run overnight; it can be during any short nap during daytime as well. Running a short-time sleep study during daytime will reduce healthcare cost significantly by reducing the need to expensive full overnight PSG study. Moreover, having a short sleep study during daytime is much more convenient for people, which is particularly important for dementia population that are usually reluctant to sleep the night in an unfamiliar environment away from their spouse/caregiver. Overall, the proposed technology will assist the sleep clinicians in the

appropriate therapeutic decisions and focusing on the investment of resources to optimize compliance to treatment, particularly in the moderate/severe-OSA group. Further studies in larger sample size are needed to assess the efficacy of adding our proposed TBS analysis in diagnostic and therapeutic tools of sleep study centers.

References

[1]     L. E. Bilston and S. C. Gandevia, "Biomechanical properties of the human upper airway and their effect on its behavior during breathing and in obstructive sleep apnea," *J. Appl. Physiol.*, vol. 116, no. 3, pp. 314–324, Feb. 2013.

[2]     A. Malhotra and D. P. White, "Obstructive sleep apnoea," *Lancet*, vol. 360, no. 9328, pp. 237–245, 2002.

[3]     B. RB, A. CL, and H. SM, *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.5*. 2018.

[4]     J. G. Park, K. Ramar, and E. J. Olson, "Updates on Definition, Consequences, and Management of Obstructive Sleep Apnea," *Mayo Clin. Proc.*, vol. 86, no. 6, pp. 549–555, Jun. 2011.

[5]     C. V. Senaratna *et al.*, "Prevalence of obstructive sleep apnea in the general population: A systematic review," *Sleep Med. Rev.*, vol. 34, no. October, pp. 70–81, Aug. 2017.

[6]     S. G. Memtsoudis, M. C. Besculides, and M. Mazumdar, "A Rude Awakening — The Perioperative Sleep Apnea Epidemic," *N. Engl. J. Med.*, vol. 368, no. 25, pp. 2352–2353, Jun. 2013.

[7]     Terry Young, M. Palta, Jerome Dempsey, P. E. Peppard, F. J. Nieto, and K. Mae Hla, "Burden of Sleep Apnea: Rationale, Design, and Major Findings of the Wisconsin Sleep Cohort Study," *WMJ Off. Publ. State Med. Soc. Wisconsin*, vol. 108, no. 5, pp. 246–249, Nov. 2010.

[8]     C. A. Kushida *et al.*, "Practice Parameters for the Indications for Polysomnography and Related Procedures: An Update for 2005," *Sleep*, vol. 28, no. 4, pp. 499–523, 2005.

[9]     P. A. Deutsch, M. S. Simmons, and J. M. Wallace, "Cost-effectiveness of split-night polysomnography and home studies in the evaluation of obstructive sleep apnea syndrome," *J. Clin. Sleep Med.*, vol. 2, no. 2, pp. 145–153, 2006.

[10]    K. N. Priftis, L. J. Hadjileontiadis, and M. L. Everard, *Breath Sounds From Basic Science to Clinical Practice*. Cham: Springer International Publishing, 2018.

[11]    T. Penzel and A. Sabil, "The use of tracheal sounds for the diagnosis of sleep apnoea," *Breathe*, vol. 13, no. 2, pp. e37–e45, Jun. 2017.

[12]    I. Ayappa and D. M. Rapoport, "The upper airway in sleep: physiology of the pharynx,"

*Sleep Med. Rev.*, vol. 7, no. 1, pp. 9–33, Feb. 2003.

[13] Z. Lan, A. Itoi, M. Takashima, M. Oda, and K. Tomoda, "Difference of pharyngeal morphology and mechanical property between OSAHS patients and normal subjects," *Auris Nasus Larynx*, vol. 33, no. 4, pp. 433–439, Dec. 2006.

[14] A. S. Jordan *et al.*, "Airway Dilator Muscle Activity and Lung Volume During Stable Breathing in Obstructive Sleep Apnea," *Sleep*, vol. 32, no. 3, pp. 361–368, Mar. 2009.

[15] Y. Finkelstein *et al.*, "Velopharyngeal Anatomy in Patients With Obstructive Sleep Apnea Versus Normal Subjects," *J. Oral Maxillofac. Surg.*, vol. 72, no. 7, pp. 1350–1372, Jul. 2014.

[16] R. J. Schwab *et al.*, "Identification of Upper Airway Anatomic Risk Factors for Obstructive Sleep Apnea with Volumetric Magnetic Resonance Imaging," *Am. J. Respir. Crit. Care Med.*, vol. 168, no. 5, pp. 522–530, Sep. 2003.

[17] R. B. Fogel *et al.*, "The effect of sleep onset on upper airway muscle activity in patients with sleep apnoea versus controls," *J. Physiol.*, vol. 564, no. 2, pp. 549–562, Apr. 2005.

[18] Z. Wu, W. Chen, M. C. K. Khoo, S. L. Davidson Ward, and K. S. Nayak, "Evaluation of upper airway collapsibility using real-time MRI," *J. Magn. Reson. Imaging*, vol. 44, no. 1, pp. 158–167, 2016.

[19] H. Nakano, T. Furukawa, and T. Tanigawa, "Tracheal Sound Analysis Using a Deep Neural Network to Detect Sleep Apnea," 2019.

[20] R. Beck, G. Rosenhouse, M. Mahagnah, R. M. Chow, D. W. Cugell, and N. Gavriely, "Measurements and Theory of Normal Tracheal Breath Sounds," *Ann. Biomed. Eng.*, vol. 33, no. 10, pp. 1344–1351, Oct. 2005.

[21] A. Kulkas *et al.*, "New tracheal sound feature for apnoea analysis," *Med. Biol. Eng. Comput.*, vol. 47, no. 4, pp. 405–412, 2009.

[22] F. Hajipour, M. Jafari Jozani, A. Elwali, and Z. Moussavi, "Regularized logistic regression for obstructive sleep apnea screening during wakefulness using daytime tracheal breathing sounds and anthropometric information," *Med. Biol. Eng. Comput.*, vol. 57, no. 12, pp. 2641–2655, Nov. 2019.

[23] A. Yadollahi, E. Giannouli, and Z. Moussavi, "Sleep apnea monitoring and diagnosis based on pulse oximetery and tracheal sound signals," *Med. Biol. Eng. Comput.*, vol. 48, no. 11, pp. 1087–1097, Nov. 2010.

[24] M. Younes, "Role of respiratory control mechanisms in the pathogenesis of obstructive sleep disorders," *J. Appl. Physiol.*, vol. 105, no. 5, pp. 1389–1405, Nov. 2008.

[25] C. M. Ryan and T. D. Bradley, "Pathogenesis of obstructive sleep apnea," *J. Appl. Physiol.*, vol. 99, no. 6, pp. 2440–2450, 2005.

[26] A. N. Rama, S. H. Tekwani, and C. A. Kushida, "Sites of Obstruction in Obstructive Sleep Apnea," *Chest*, vol. 122, no. 4, pp. 1139–1147, Oct. 2002.

[27] F. Hajipour and Z. Moussavi, "Spectral and Higher Order Statistical Characteristics of

Expiratory Tracheal Breathing Sounds During Wakefulness and Sleep in People with Different Levels of Obstructive Sleep Apnea," *J. Med. Biol. Eng.*, vol. 39, no. 2, pp. 244–250, Apr. 2019.

[28] S. Huq and Z. Moussavi, "Acoustic breath-phase detection using tracheal breath sounds," *Med. Biol. Eng. Comput.*, vol. 50, no. 3, pp. 297–308, Mar. 2012.

[29] F. Arvin, S. Doraisamy, and E. Safar Khorasani, "Frequency shifting approach towards textual transcription of heartbeat sounds," *Biol. Proced. Online*, vol. 13, no. 1, pp. 1–7, 2011.

[30] N. Gavriely and D. W. Cugell, *Breath sounds methodology*. cRc PrEss, 1995.

[31] A. Yadollahi and Z. M. K. Moussavi, "Acoustical flow estimation: Review and validation," *IEEE Eng. Med. Biol. Mag.*, vol. 26, no. 1, pp. 56–61, Jan. 2007.

[32] A. Elwali and Z. Moussavi, "Obstructive Sleep Apnea Screening and Airway Structure Characterization During Wakefulness Using Tracheal Breathing Sounds," *Ann. Biomed. Eng.*, vol. 45, no. 3, pp. 839–850, Mar. 2017.

[33] R. J. SIMES, "An improved Bonferroni procedure for multiple tests of significance," *Biometrika*, vol. 73, no. 3, pp. 751–754, 1986.

[34] C. Kalkbrenner, M. Eichenlaub, S. Rüdiger, C. Kropf-Sanchen, W. Rottbauer, and R. Brucher, "Apnea and heart rate detection from tracheal body sounds for the diagnosis of sleep-related breathing disorders," *Med. Biol. Eng. Comput.*, vol. 56, no. 4, pp. 671–681, 2018.

[35] A. Kulkas *et al.*, "Detection of compressed tracheal sound patterns with large amplitude variation during sleep," *Med. Biol. Eng. Comput.*, vol. 46, no. 4, pp. 315–321, 2008.

[36] P. Harper, S. S. Kraman, H. Pasterkamp, and G. R. Wodicka, "An acoustic model of the respiratory tract," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 5, pp. 543–550, May 2001.

[37] H. S. Seddeq, "Factors Influencing Acoustic Performance of Sound Absorptive Materials," *Aust. J. Basic Appl. Sci.*, vol. 3, no. 4, pp. 4610–4617, 2009.

[38] Veldi, Vasar, Vain, Hion, and Kull, "Computerized endopharyngeal myotonometry (CEM): A new method to evaluate the tissue tone of the soft palate in patients with obstructive sleep apnoea syndrome," *J. Sleep Res.*, vol. 9, no. 3, pp. 279–284, Sep. 2000.

[39] O. E. Jensen, "Flows through deformable airways," in *Centre for Mathematical Medicine School of Mathematical Sciences University of Nottingham, UK*, 2002, pp. 27217–222.

# Chapter V.    Summary and Concluding Remarks

## V.1  Summary of Findings

Obstructive sleep apnea (OSA) is a common and serious respiratory disorder. A timely diagnosis of OSA is essential, especially for severe cases in need of quick treatment. Polysomnography (PSG) assessment is the gold standard for OSA diagnosis; however, due to its time- and resource-consuming nature is not suitable as an early-stage and/or quick diagnostic tool. The current quick screening tools are questionnaires. Yet, due to their poor specificity (high false positive) they have a high referral rate for the PSG study, which defeats the purpose of quick screening and could even impose a substantial impact on health care costs. Thus, a fast, reliable, and objective assessment tool for OSA diagnosis is required. Throughout the series of studies in this dissertation, tracheal breathing sounds (TBS) of individuals with different levels of OSA severity during wakefulness (daytime when the subjects are fully awake) and a short period of sleep have shown their ability to predict the likelihood of OSA with reasonable accuracy, and to show the upper airway (UA) structural variations due to OSA.

We started our studies by recording a few minutes of TBS data during the daytime while the participants were awake. After respiratory phase identification (inspiration/expiration), we conducted various processes to remove unwanted noises from collected data. Then we extracted characteristic features that could potentially differentiate individuals with different levels of OSA severity.

Following, we used the LASSO Regularized LR as a parametric feature selection and classification approach for predicting the likelihood severity of OSA disorder using TBS data and anthropometric information of individuals. We showed this model has an unbiased and high generalization performance for OSA screening. Our results depicted high AUCs for different

LASSO Regularized LR classifiers over different training sets; indicating the robustness of those classifiers with respect to our selected features. The features were chosen by this regression modeling were shown to be highly correlated with the apnea-hypopnea index (AHI) and statistically significantly different between the non-OSA(AHI<5) and moderate/severe-OSA (AHI≥15) groups.

Through the analysis, we showed the extent of the importance of the selected features for predicting the correct OSA severity level using their estimated coefficient in the LASSO Regularized LR model. Our investigations on the physiological interpretation of the selected features have shown that moderate/severe OSA individuals, compared to non-OSA individuals, exhibit higher NC, higher slope around the second formant, and higher average power at high-frequency ranges. These findings are congruent with the fact that OSA individuals, in general, have narrower UA with lower resonance frequency and stiffer tissues, compared to non-OSA individuals. This information can be used clinically for therapeutic and diagnostic purposes. The physiological interpretations and reasons are discussed in previous chapters.

One of the key findings of this study was that the LASSO Regularized LR model presented in this work does not need respiratory phase identification, as the best features were extracted from the combination of inspiratory and expiratory TBS. Eliminating this part from the preprocessing procedure will speed up the data analysis significantly; hence, providing a faster online screening tool for OSA. Concerning the speed of the process, we showed that the overall feature extraction, feature selection, and classification using the LASSO Regularized LR approach is linear in terms of the number of extracted features and number of individuals; hence, the procedure is very fast. Consequently, we showed that the LASSO regularized LR is simple, quick, and computationally

more effective than other previous methods for wakefulness OSA screening; thus, it is suitable for real-time applications.

In our first objective, we assumed a linear relationship between the selected features and the predicted OSA severity level of participants. However, it is possible that our linearity assumption might not be valid all the time. Therefore, after developing the first parametric TBS analysis algorithm with successful outcomes for finding the association between the features and AHI as well as OSA prediction and classification, we continued the research by investigating the application of Random Forest (RF). RF is a *flexible* non-parametric model, without any assumption of a particular functional form of the relationship between the selected features and severity level of participants for predicting the likelihood of OSA during wakefulness. Accordingly, unlike the LASSO regularized LR that fits a stochastic model to draw conclusions, no specific model is fitted to the data in an RF model. The approach is to find an algorithm that operates on TBS and anthropometric features to generate the OSA severity level. One limitation of this approach is that a very large number of participants, compared to that typically needed in the LASSO Regularized LR approach, is required to obtain an accurate estimate of the function describing the relationship between selected features and OSA severity level of individuals. Our classification results showed a higher generalization performance for OSA screening of RF model compared to the LASSO Regularized LR model with 3.5%, 2.4%, and 3.7% improvement in accuracy, specificity and sensitivity of blind testing, respectively. The high AUC statistics of the RF model revealed the robustness of this approach and confirmed its efficiency in wakefulness OSA diagnosis.

It is worth mentioning that the findings of our studies showed the superiority of classification testing results over the OOB results using the RF model. While using LASSO regularized LR, the training performance was better than the testing performance. RF uses the bagging procedure to

138

enhance its generalization error (test set accuracy). In this approach, the out-of-bag estimates are used to give approximately optimal estimates of generalization errors for bagged predictors. Essentially, the out-of-bag estimate for the generalization error is the error rate of the out-of-bag classifier on the training set. In a study by Breiman [1], it has been shown that the out-of-bag estimate is as accurate as using a test set of the same size as the training set and that these estimates are unbiased. Also, RF fits long trees, which essentially overfits training data. Accordingly, training errors will be very small for RF, and reporting it does not give any insight. Therefore, instead of training error, we reported the classification results for the OOB data of training sets. On the other hand, for the LASSO regularized LR, we reported the training vs. blind testing classification results. As the training error tends to be smaller than that of blind testing, it is in line with the theory to have a smaller prediction error (1-accuracy) for training (17.7%) compared to that of blind-testing (20.7%).

The findings of our studies also showed that the RF model could provide the importance of the selected features and the extent of their relationship to the assigned OSA severity level of individuals. In terms of physiological interpretation of the selected features, our findings have shown similar results to those of the LASSO regularized LR, except that when using the RF model, the moderate/severe OSA individuals compared to non-OSA individuals exhibited lower average power at low-frequency ranges. This finding is congruent with the fact that OSA individuals, in general, have narrower and more compliant UA compared to non-OSA individuals that facilitate more power absorption at low-frequency ranges of power spectra.

Regarding comparing the RF and LASSO regularized LR models together, the statistical findings showed a substantial agreement between the severity levels predicted by the two models. The statistical tests also demonstrated that the selected features of the RF approach, like features

139

chosen by the LASSO regularized LR, have both a highly significant correlation with AHI and a high discriminative power to separate two OSA groups. Our further investigations on the two mentioned models showed a small subset of uncorrelated features for distinguishing the OSA groups could be obtained using the LASSO regularized LR, but a list of influential features (important for clinical diagnosis) can be obtained using RF. Considering the computational costs and memory usage of two approaches, we showed that RF is slower and have larger sized model than LASSO regularized LR. Overall, we concluded that the selection of the appropriate feature reduction and classification approach depends on the application and nature of data. Generally, due to the high accuracy of the RF model, its model-independent nature, and its ability to deal with unbalanced and missing values, this approach might be preferred for OSA screening. Though, if the size of data is large and a quick real-time screening is required, the LASSO regularized LR approach might be a better choice as it can still provide relatively fast and accurate classification results.

In the analysis of our wakefulness TBS data, we only used non-OSA and moderate/severe-OSA individuals, and data of mild-OSA subjects ($5 \leq$AHI$<15$) were not used for feature extraction and feature reduction. It was due to the point that despite the AHI is the most commonly used metric for OSA severity that has been used in the literature as well as in our study, however, the clinical diagnosis of OSA is not only based on AHI but a combination of AHI with the OSA clinical symptoms (frequency of arousal, daytime fatigue, etc.). Therefore, many individuals with similar AHI have different clinical signs and, consequently, different OSA severity levels. As TBS features are affected by the pathophysiology of OSA, they are indeed good representatives of the OSA severity. However, the UA pathology of two individuals with similar AHI (e.g., AHI of 14 and 16) may not be that much different to affect their TBS differently enough to be detected by

sound analysis. For this reason, to avoid the misclassification in the borderline ranges with artificially crisp nature, we allowed a small gap ($5 \leq AHI < 15$) in continuous AHI values of the non-OSA and moderate/severe OSA groups to form the two groups. Our findings on the assessment of mild-OSA subjects showed that while there was not much difference between the AHI values of participants of this group, their TBS features were different. This finding confirms the shortcoming of using TBS for individuals with similar (close range) AHI. This shortcoming could be due to many confounding variables such as BMI that impact the TBS and make it challenging to come with a good classification for individuals with similar AHI. Thus, ideally, we should have the OSA groups matched in such confounding characteristics. However, that requires a large dataset.

In addition to 2-class OSA classification, we also investigated 3 severity group classification using RF modeling on TBS data and anthropometric information of individuals (Appendix B). In the dataset of that study, the size of data in 3 classes was unbalanced and the majority of the participants were from the non/mild OSA group. Consequently, the prediction error between the classes was highly unbalanced; the non/mild-OSA group ($AHI < 10$) showed a low prediction error, while moderate-OSA ($15 < AHI < 25$) and severe-OSA ($AHI > 30$) groups showed a high prediction error. We suggest the implementation of assigning different weights to different groups for the RF model, besides the use of higher-order statistics for feature extraction to improve the discrimination and multi-class OSA classification.

The premise of up to this stage of this dissertation was based on the fact that OSA is associated with chronic physiologic and anatomic changes that persist during wakefulness [2]–[5]. Thus, we analyzed TBS data during wakefulness. However, it is known that the physiological properties of the UA change dynamically, but this change is dependent on the sleep/wakefulness status and severity of OSA [6]–[8]. Therefore, we decided to investigate TBS properties during both

wakefulness and a short period of sleep and analyze the pattern of their changes from wakefulness to sleep, to assess their power in the improvement of the OSA screening/classification, and their ability to revealing the UA anatomy and physiology in a detailed but straightforward manner. Based on recommendations of sleep physicians and several sleep studies, we used the AHI threshold of 15 (mild-OSA (AHI <15) and moderate/severe-OSA (AHI >15)) because individuals with AHI>15 have an increased chance for cardiovascular or mortality risk and are in need of treatment [9]. Accordingly, we continued our work by recording TBS data during sleep. After respiratory phase identification, we selected five normal (free of any apneic events, including flow limitation), noise- and snore-free expiratory sounds in the supine position, and sleep stage 2 for further analysis. Then we extracted characteristic features that could potentially differentiate individuals with different levels of OSA severity.

Our analysis of spectral characteristics of TBS revealed that during wakefulness, the low-frequency average power of TBS for the moderate/severe OSA group was significantly lesser than that of the mild-OSA group. However, there was an inverse pattern for their high-frequency average power during sleep. For assessing the pattern of changes in TBS properties from wakefulness to sleep, we considered the neck circumference (NC) as a covariate to ensure that our TBS features were not just measuring the differences in breathing sounds due to NC between the two OSA groups. The findings manifested a significant main effect of sleep/wakefulness status on the TBS spectral characteristics and a significant interaction between the sleep/wakefulness conditions and the OSA severity level of participants. These findings imply that the TBS properties change significantly when one sleeps, and these changes are different within mild and moderate/severe OSA groups. Our detailed study on the average power spectra at low- and high-frequency ranges demonstrate a significant decrease in the low-frequency average power of the

mild group and a marginally significant increase in high-frequency average power of the moderate/severe group from wakefulness to sleep. The results also show a significant correlation between OSA severity level and spectral features, including low-frequency power during wakefulness, high-frequency power during sleep, and the change of high-frequency power from wakefulness to sleep.

The above findings indicate TBS properties during wakefulness and sleep are different, and more importantly the TBS changes from wakefulness to sleep could reveal the pathophysiology of the UA and their change due to OSA. They represent the higher resistance, more compliance, and thickness for UA of moderate/severe OSA compared to mild OSA during wakefulness. They also express a higher reduction in airway caliber of moderate/severe OSA group as well as non-homogeneity and increased resistance and stiffness of their UA compared to those of mild OSA during sleep. Furthermore, how the spectral pattern of TBS changes from wakefulness to sleep is representative of a regional stiffening and compliance in UA of OSA individuals; thus, may be representative of the shape of their velopharyngeal narrowing.

From classification point of view, an ensemble of linear kernel SVM classifiers indicates the superiority of investigating the TBS changes from wakefulness to sleep with the accuracy of $87.5\pm4.5\%$ for OSA screening and differentiating between the two OSA groups rather than only sleep or wakefulness, with accuracies of $70.1\pm6.5\%$ and $78.8\pm4.2\%$, respectively. We explained that it would likely be due to the increased stiffness and thickness of the UA of moderate/severe-OSA group from wakefulness to sleep compared to those of mild group.

It worth noting that for the last objective of this dissertation, we used the SVM classifier instead of LASSO Regularized LR and RF. It is because the RF model requires a high number of participants to provide accurate estimation; however, for the last part of our study, we had only 40

participants. Also, the number of extracted features in our final study was quite low (only two), and we did not want to have feature selection; therefore, we have not used the LASSO regularized LR model.

Figure 22 shows a summary diagram of the experimental methods for all the objectives that have been conducted in this dissertation. Overall, the findings of these studies showed the TBS characteristics can indirectly and non-invasively reveal the structural changes of the UA in relation to OSA and are beneficial to enhance the current OSA diagnostic methods to stratify the severity of OSA patients in a non-laborious, non-time-consuming, and less expensive manner using a short period of sleep instead of a full overnight sleep study, which can be during any short nap during daytime as well. Running a short-time sleep study during the daytime will reduce healthcare costs significantly by reducing the need for expensive full overnight PSG study.
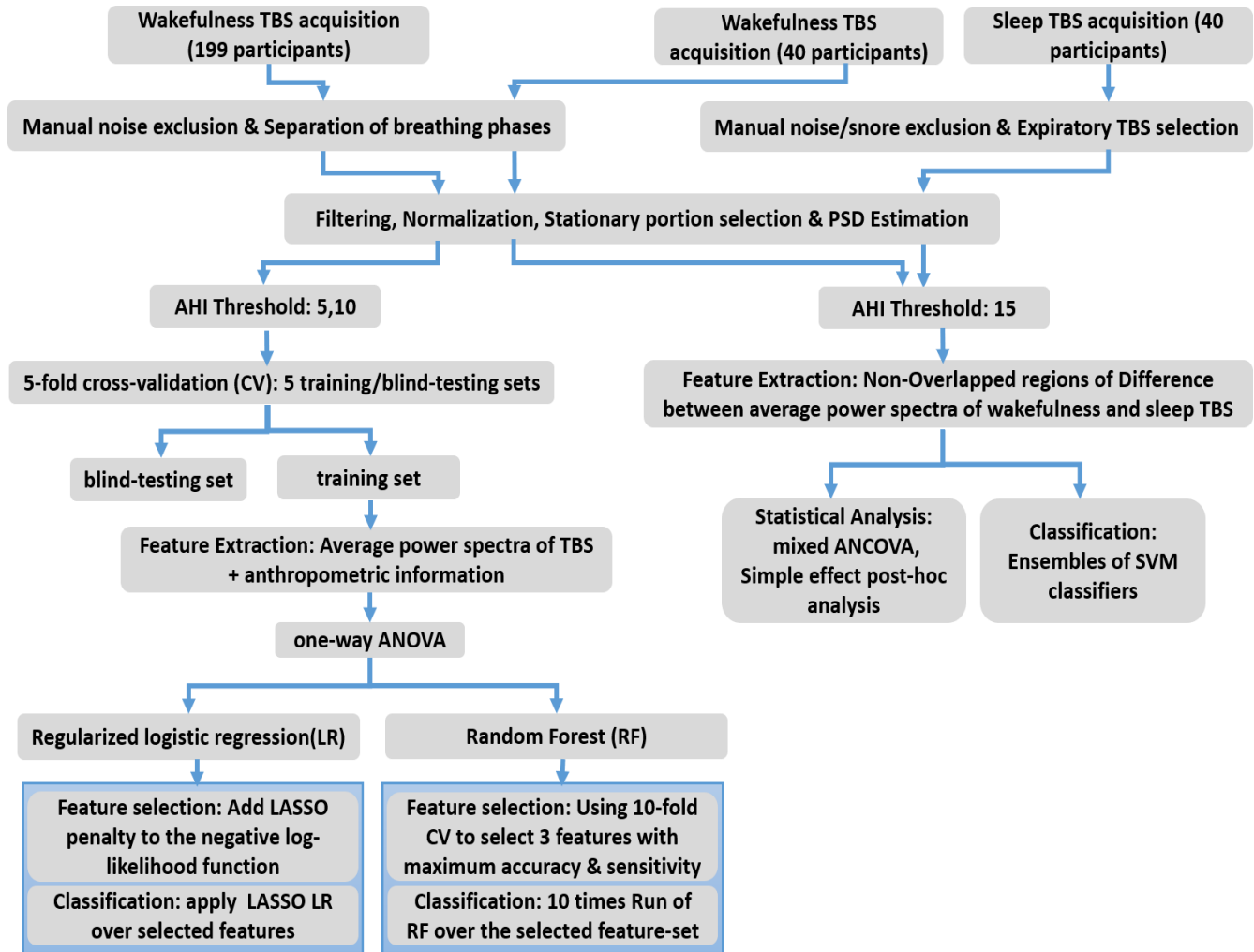
Figure 22 a summary diagram of the experimental methods for all the objectives performed in this dissertation

## V.2  Limitation of the study

In the studies explained in Chapter I and Chapter III, we had 199 participants. Although it is a large dataset studied on wakefulness TBS analysis of OSA, however, it is not enough to overcome the heterogeneity of the participants. Also, for studies explained in Chapter IV, we had 40 participants. Due to the small sample size of our studies, we didn't have the OSA groups matched in anthropometric variables that might have a confounding effect on TBS. Another limitation was

the imbalance of the severity groups in our data such that we didn't have enough moderate and severe OSA participants. Thus, in our future study, we have to increase our dataset to include an equal number of participants considering all confounding variables to overcome the above limitations. Another limitation of our studies was the lack of access to imaging machines to have a direct measure of the UA collapsibility of the study participants. Therefore, for interpretation purposes, we compared our results with the general finding of the imaging studies.

Another limitation of our studies during sleep was investigating the TBS in only stage 2 of sleep. It was due to both recording of TBS for a few hours of sleep, not the entire night, and not having enough data for all participants in other sleep stages that were also in the supine position. Additional limitation of our sleep studies was the no identification of mouth-nose breathing during sleep due to the recording of breathing flow using a full-facemask that covered both mouth and nose.

## V.3  Future Work Recommendations

Although many studies have investigated the acoustical analysis of breathing sounds and its clinical application, especially for OSA screening and diagnosis during both wakefulness and sleep, it still requires further research and quality improvements such as the followings.

- The lack of standardization for many study factors including participants of the study (required population size and the anthropometric information that matched between individuals), preprocessing steps, and training and blind testing data consideration lead to huge differences between the published results and methods in the literature. Therefore, it is necessary to have a guideline for standardization of all stages of breathing sounds analysis.

- In our studies, one of the major challenges in TBS analysis during sleep was the extraction of breathing sound segments from stable sleep periods in stage N2 and in the supine position, void of snoring sounds or noises (including artifacts, vocal noises and swallowing). It was also challenging to extract noise-free breathing sounds form the recorded sounds during wakefulness. Automating this part of preprocessing would significantly speed up the analysis process.

- Accurate detection and classification of individuals with different levels of OSA severity are essential for diagnosis cost management and appropriate therapy selection. It will particularly assist the sleep clinicians in optimizing compliance to treatment in the moderate and severe OSA groups. In the studies mentioned in this dissertation, our data was imbalanced in the number of participants within different severity groups. Therefore, having a reasonably large balanced number of participants, matched in confounding variables in various OSA groups, in addition to investigations for extracting more discriminative features and optimizing classifiers, would be of interest to improve the potential of TBS analysis to differentiate all OSA severity groups especially moderate and severe groups of OSA.

- As known, the power spectra of TBS are variable from person to person. The power spectral variability between OSA individuals might be explained by the level and place of obstruction in the UA. In our studies (Chapter IV), we have shown that the decrease in low-frequency power and the increase in high-frequency power from wakefulness to sleep are representative of the increased compliance and stiffness of UA, respectively. A regional bottle shape narrowing represents both more compliance and more stiffness of the UA simultaneously. Therefore, we speculate analysis of low- and high-frequency power

spectral pattern changes from wakefulness to sleep may be representative of the shape of the individual's velopharyngeal narrowing. Therefore, to investigate this hypothesis, one has to simultaneously record TBS with some imaging modality in a large population.

- In all studies within this dissertation, we recorded TBS from one location, which was over suprasternal notch of the trachea, and the multi-site recording of TBS was not investigated. However, multi-site recordings by placing microphones in various areas including in the ear, over different locations on the neck, and the nose, might help in extracting more information from the TBS to detect the site of obstructions. It might be especially useful when there is more than one area of obstruction in the UA of individuals. This field deserves further investigation.

- In the studies mentioned in Chapter IV, we only investigated the TBS in stage 2 of sleep. It was because stage 2 was the most common sleep stage in our dataset. However, it has been shown that the UA structure [10], and therefore the breathing sounds characteristics changes depending on the sleep stages [11]. Therefore, it worth investigating TBS in various sleep stages regarding OSA and its severity level. It might be especially beneficial to investigate the TBS characteristics in the transition period of wakefulness to sleep (stage 1) as that may show the dynamic of the UA changes better.

## References

[1]    L. Breiman, "out-of-bag estimation, Technical Report," pp. 1–13, 1996.

[2]    G. C. Barkdull, C. a Kohl, M. Patel, and T. M. Davidson, "Computed Tomography Imaging of Patients With Obstructive Sleep Apnea," *Laryngoscope*, vol. 118, no. 8, pp. 1486–1492, Aug. 2008.

[3]    K.-H. Liu *et al.*, "Sonographic measurement of lateral parapharyngeal wall thickness in patients with obstructive sleep apnea.," *J. Sleep*, vol. 30, no. 11, pp. 1503–8, 2007.

[4]    R. J. Schwab *et al.*, "Identification of Upper Airway Anatomic Risk Factors for Obstructive Sleep Apnea with Volumetric Magnetic Resonance Imaging," *Am. J. Respir.*

*Crit. Care Med.*, vol. 168, no. 5, pp. 522–530, Sep. 2003.

[5]     Y. Finkelstein *et al.*, "Velopharyngeal Anatomy in Patients With Obstructive Sleep Apnea Versus Normal Subjects," *J. Oral Maxillofac. Surg.*, vol. 72, no. 7, pp. 1350–1372, Jul. 2014.

[6]     F. J. Trudo, W. B. Gefter, K. C. Welch, K. B. Gupta, G. Maislin, and R. J. Schwab, "State-related Changes in Upper Airway Caliber and Surrounding Soft-Tissue Structures in Normal Subjects," *Am. J. Respir. Crit. Care Med.*, vol. 158, no. 4, pp. 1259–1270, Oct. 1998.

[7]     Z. Lan, A. Itoi, M. Takashima, M. Oda, and K. Tomoda, "Difference of pharyngeal morphology and mechanical property between OSAHS patients and normal subjects," *Auris Nasus Larynx*, vol. 33, no. 4, pp. 433–439, Dec. 2006.

[8]     A. S. Jordan *et al.*, "Airway Dilator Muscle Activity and Lung Volume During Stable Breathing in Obstructive Sleep Apnea," *Sleep*, vol. 32, no. 3, pp. 361–368, Mar. 2009.

[9]     Terry Young, M. Palta, Jerome Dempsey, P. E. Peppard, F. J. Nieto, and K. Mae Hla, "Burden of Sleep Apnea: Rationale, Design, and Major Findings of the Wisconsin Sleep Cohort Study," *WMJ Off. Publ. State Med. Soc. Wisconsin*, vol. 108, no. 5, pp. 246–249, Nov. 2010.

[10]    L. E. Bilston and S. C. Gandevia, "Biomechanical properties of the human upper airway and their effect on its behavior during breathing and in obstructive sleep apnea," *J. Appl. Physiol.*, vol. 116, no. 3, pp. 314–324, Feb. 2013.

[11]    A. Azarbarzin and Z. Moussavi, "Intra-subject variability of snoring sounds in relation to body position, sleep stage, and blood oxygen level," *Med. Biol. Eng. Comput.*, vol. 51, no. 4, pp. 429–439, 2013.

# Appendix A. Anatomy and Physiology of the Respiratory System

## A.1  Anatomy and Physiology of the respiratory system

The principal function of the respiratory system is to deliver oxygen to the blood for metabolism and to expel the metabolic by-products, of which carbon dioxide is the main constituent, from the body. The respiratory system is divided into two primary components:

- Upper respiratory tract: consists of the nose and mouth, the pharynx, and the larynx.

- Lower respiratory tract: consists of the trachea, the lungs, and the bronchial tree segments (including bronchi, bronchioles, and alveoli).

Inspiration is the process that causes air to enter the lungs, and expiration is the process that causes air to leave the lungs. During inspiration, air enters via nose and mouth as a result of the contraction of the diaphragm and the external intercostal muscles. The air is then passing through the shared pharynx (both food and air pass through the pharynx) and then it passes though the specialized (only air passing) larynx that contains vocal cords. The epiglottis covers the larynx and prevents food/drinks from entering airways. The air then enters the lower airways through the trachea. The trachea branches into two smaller tubes called bronchi, which extend laterally into the hilum of the left and right lung, respectively. The bronchi are further subdivided into bronchioles. These, in turn, branch off to smaller parts and finally divide into two to eleven alveolar ducts. There are five to six alveolar sacs associated with each alveolar duct and lead to a grape-like cluster called alveolus. The alveolus is the smallest anatomical unit of the lung and the site of gas exchange between the lung and the bloodstream.

The lungs are pyramid-shaped, paired organs lying in the thoracic cavity—each lung's base resting on the diaphragm and is anchored by the mediastinum. The lungs are composed of

smaller units called lobes, which are separated from each other by fissures. The right lung is consisting of three lobes: the superior, middle, and inferior lobes. The left lung is composed of two lobes: the superior and inferior lobes. The lungs and lobes are enclosed by the pleural sac, which consists of two continuous membranes: the visceral pleura that surrounds the lungs, and the parietal pleura that connects to the thoracic wall, the mediastinum, and the diaphragm. The potential space between the visceral and parietal layers is called the pleural cavity and contains the pleural fluid, which lubricates the pleural surfaces and allows a sliding movement of lungs relative to the chest wall and diaphragm during respiration. For an intact respiratory system, the lungs are elastic and want to be smaller. Contrarily, the thorax is elastic and wants to be bigger.

To provide the inspiratory and expiratory cycles of breathing, a pressure gradient is required to move air in and out of the lungs. At the end of expiration, the alveolar pressure ($P_{ALV}$) is 0 cm $H_2O$, and the intra-plural pressure ($P_{PI}$) is -5 cm $H_2O$; hence, the trans-pulmonary pressure ($P_{ALV}$ - $P_{PI}$), that represents a pressure gradient, is 5 cm $H_2O$. Inspiration is an active process that initiated by the diaphragm and supported by the external intercostal muscles. When the diaphragm contracts, it moves downward and vertically enlarges the thoracic cavity. When the external intercostal muscles contract, they move the ribs and sternum upward, hence, horizontally increase the thoracic cavity. Thus, the respiratory muscle contraction increases thoracic volume to stretch the lungs. The increased thoracic volume is resulting in an increase in the trans-pulmonary pressure to encounter the elastic recoil of the lungs. This increased gradient pressure moves the air inside the lungs. At the end of the inspiration, $P_{ALV}$ is -1 cm $H_2O$, $P_{PI}$ is -8 cm $H_2O$, and trans-pulmonary pressure is calculated as -1 cm $H_2O$ – (-8 cm $H_2O$) = +7 cm $H_2O$. During forceful inhalation (more than 35 breaths per minute) or in approaching respiratory failure, accessory muscles, including the

sternocleidomastoid, scalene group, and pectoralis minor, are also involved in sustaining the respiratory rate.

Exhalation, contrarily, is a passive process and results from the elastic recoil of the lungs, rib cage, and diaphragm. During expiration, the contracted muscles relax, the thoracic cavity restores to pre-inspiratory volume, and the air exhaled until the pressures in the chest and the atmosphere reach equilibrium. The auxiliary muscles for active expiration include internal intercostal, abdominals, and Quadratus Lumborum. During this process, the air is forcefully exhaled out (e.g., blowing out a candle).

## A.1.1 Lung Volume

Lung volume denotes the volume of air in the lung at a specific point of the breathing cycle. The maximum volume of gas that the lungs can accommodate is the total lung capacity (TLC). TLC is calculated by the summation of two primary lung volumes: vital capacity (VC) and residual volume (RV). VC is the total amount of air exhaled after a maximal inhalation, and RV is the amount of air that remained in the lung after maximum exhalation. The functional residual capacity (FRC) is the amount of gas remaining in the lung at the end of a normal quiet expiration. It is calculated by adding up the RV and the expiratory reserve volume (ERV), a volume that can be exhaled by maximum effort.

Studies have been shown that several factors, including obesity, position, and sleep state, can change the lung volumes [1], [2]. It has been observed that in healthy normal subjects, the FRC reduced from wakefulness to sleep [1]. This reduction starts almost immediately following sleep onset [1], [3].

## A.2  Potential Factors Affecting Upper Airway Patency

The focus of this dissertation is to evaluate acoustical upper airway (UA) changes due to obstructive sleep apnea (OSA) and screening OSA acoustically. Obstructive apneas and hypopneas occur due to the recurrent episodes of complete and/or partial collapse of the UA during sleep. A detailed description of the UA anatomy in healthy and OSA individuals and an elaboration on anatomic factors contributing to UA collapse are provided in Chapter I. The critical contributors to the pathogenesis of OSA are the sleep-related narrowing and increased compliance/collapsibility of the upper airway [4]. In this section we elaborate on a few conditions that influence UA narrowing and collapsibility.

### A.2.1 Obesity

Obesity is certainly the most common predisposing factor for the development of OSA [5]. It has been estimated that in 41% to 58% of adults with OSA, the disorder is directly attributed to obesity [6]. Obesity could be either centripetal with fat preferentially distributed to the abdominal viscera, upper body, and neck, or peripheral, in which fat is preferentially distributed to the subcutaneous tissues of the hips and thighs. The centripetal pattern is more closely linked to OSA than the other one [4]. Two mechanisms could justify why obesity may worsen OSA by affecting UA anatomy/collapsibility. First, through fat deposition around the collapsible segment of UA (i.e., hard palate to the base of the epiglottis) [7], which resulted in a smaller lumen and increased collapsibility of UA [5]. Second, through reductions in FRC of the lung, which leads to a more collapsible pharyngeal airway [7].

153

## A.2.2 Gender

OSA is more widespread in men than in women [8]. It has been mentioned that the pattern of fat distribution and differences in sex hormones are two critical factors leading to this difference in sex distribution [4]. Using the MRI technique, it has been shown that obese men have increased centripetal fat distribution, a higher proportion of fat deposition, and larger overall soft tissue mass on their UA compared to the women; hence, they are more susceptible to the development of OSA [8].

On the other hand, it has been identified that in postmenopausal women compared to premenopausal women, the prevalence of OSA is higher [9]. It could be either due to the change in the fat distribution pattern that takes on a more centripetal (associated with a greater tendency to develop OSA) or due to the low progesterone/estrogen or high testosterone hormonal level [4], [9]. It has been reported that progesterone is a respiratory stimulant hormone and may stabilize the respiratory control system and maintain UA patency; hence, protect against OSA [4]. Contrarily, testosterone contributes to fat deposition in the neck and is associated with UA collapsibility in OSA patients.

## A.2.3 Age

The prevalence and severity of OSA increase substantially with aging, from age 35 to about 60 years old [10]. It has been shown that aging leads to several anatomical changes that could predispose UA to collapse during sleep. These changes include variations in UA bony structure, an increase in soft palate length (more considerable change in women than men), an increase in the pharyngeal airway length in women, an increase in the parapharyngeal fat pads size, and a decrease in pharyngeal lumen size [11], [12].

## A.2.4 Lung Volume

With the change in lung volume, not only the size of intrathoracic airways change but also the UA size varies [13]. Increases in lung volume seem to promote pharyngeal patency and make it more resistant to collapse through increased UA dilator muscle activity or increased longitudinal tracheal traction [14]. With rising lung volume, the tension on the pharyngeal walls increases, thereby rendering them stiffer and less likely to collapse. The decrease in lung volume that occurs during sleep can lead to essential reductions in longitudinal traction on the airway, resulting in a more collapsible pharynx [7]. The degree of dependence of UA cross-sectional area on lung volume and the tendency of the pharynx to collapse is highest at low lung volume (below FRC) in OSA patients compared to control subjects [15].

As the lung volume increases during inspiration and decreases during the expiration, the same change is expected for the UA dimension during respiratory cycles. Generally, the UA size rises slightly at the beginning of inspiration due to the activation of its dilator muscles and remains constant for the rest of inhalation. At the primitive stage of expiration, the UA expands to its maximum due to the positive intraluminal pressure but reaches its minimum at the final stage; hence is more susceptible to obstruction [16]. As Figure 23 shows, the change in UA size occurs in healthy subjects too, but the amount of variability is small [17]. Imaging studies have been confirmed a relationship between the UA narrowing and the lateral pharyngeal wall as it is relatively sustained during inspiration but thinned in initial expiration and thickened toward the end of expiration [18].
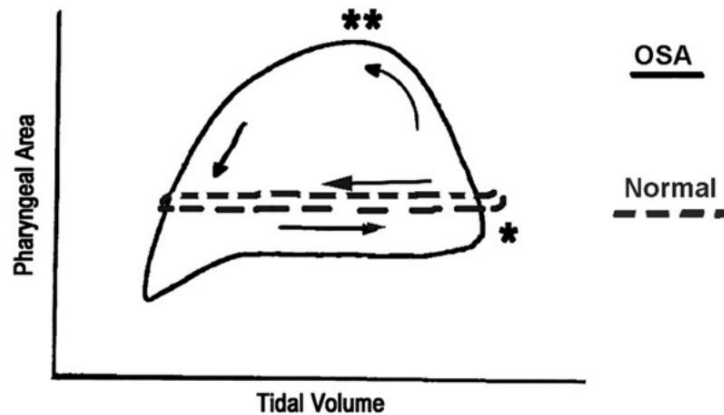
Figure 23 pattern of change in the upper airway of normal and apneic individuals during inspiratory and expiratory breathing phases. Reprinted from [17]. Copyright © 2007, with permission from Elsevier

## A.2.5 Racial Factors

Epidemiological studies suggest ethnicity, through genetics and environmental influences, differentially affect factors contributing to the pathogenesis of OSA [19]. For example, it has been shown that Asians compared to the Caucasians have shorter maxillae and mandibles and smaller anteroposterior facial dimensions; hence, shrinking of the bony cage is a more critical factor in OSA development of than obesity in this racial group [4], [19]. Contrarily, it has been demonstrated that African-Americans compared to the Caucasians have increased tongue area and increased soft palate length; therefore, soft tissue factors play a more significant role in susceptibility to OSA in this ethnic group [20].

## Reference

[1]  R. D. Ballard, C. W. Clover, and D. P. White, "Influence of non-REM sleep on inspiratory muscle activity and lung volume in asthmatic patients," *Am. Rev. Respir. Dis.*, vol. 147, no. 4, pp. 880–886, 1993.

[2]  S. A. Joosten *et al.*, "Evaluation of the role of lung volume and airway size and shape in supine-predominant obstructive sleep apnoea patients," *Respirology*, vol. 20, no. 5, pp. 819–827, 2015.

[3]  P. Koo, E. J. Gartman, J. M. Sethi, E. Kawar, and F. D. McCool, "Change in end-expiratory

lung volume during sleep in patients at risk for obstructive sleep apnea," *J. Clin. Sleep Med.*, vol. 13, no. 8, pp. 941–947, 2017.

[4]     C. M. Ryan and T. D. Bradley, "Pathogenesis of obstructive sleep apnea," *J. Appl. Physiol.*, vol. 99, no. 6, pp. 2440–2450, 2005.

[5]     A. Romero-Corral, S. M. Caples, F. Lopez-Jimenez, and V. K. Somers, "Interactions Between Obesity and Obstructive Sleep Apnea," *Chest*, vol. 137, no. 3, pp. 711–719, Mar. 2010.

[6]     A. Malhotra and D. P. White, "Obstructive sleep apnoea," *Lancet*, vol. 360, no. 9328, pp. 237–245, 2002.

[7]     D. P. White and M. K. Younes, "Obstructive Sleep Apnea," in *International Encyclopedia of Public Health*, vol. 2, no. October, Elsevier, 2012, pp. 308–314.

[8]     A. T. Whittle, I. L. Mortimore, N. J. Douglas, I. Marshall, P. K. Wraith, and R. J. Sellar, "Neck soft tissue and fat distribution: Comparison between normal men and women by magnetic resonance imaging," *Thorax*, vol. 54, no. 4, pp. 323–328, 1999.

[9]     E. O. BIXLER *et al.*, "Prevalence of Sleep-disordered Breathing in Women," *Am. J. Respir. Crit. Care Med.*, vol. 163, no. 3, pp. 608–613, 2013.

[10]    T. Sleep and H. Health, "Predictors of Sleep-Disordered Breathing in Community-Dwelling Adults," *Heal. San Fr.*, vol. 162, no. 40, pp. 893–900, 2002.

[11]    A. Malhotra *et al.*, "Aging influences on pharyngeal anatomy and physiology: the predisposition to pharyngeal collapse," *Am. J. Med.*, vol. 119, no. 1, pp. 72.e9-72.e14, 2006.

[12]    S. E. Martin, R. Mathur, I. Marshall, and N. J. Douglas, "The effect of age, sex, obesity and posture on upper airway size," *Eur. Respir. J.*, vol. 10, no. 9, pp. 2087–2090, 1997.

[13]    V. Hoffstein, N. Zamel, and E. A. Phillipson, "Lung Volume Dependence of Pharyngeal Cross-Sectional Area in Patients with Obstructive Sleep Apnea," *Am. Rev. Respir. Dis.*, vol. 130, no. 2, pp. 175–178, 2018.

[14]    J. M. Williams, M. B. Chb, and C. D. H. Frca, "Obstructive sleep apnoea," vol. 3, no. 3, pp. 75–78, 2003.

[15]    T. D. Bradley *et al.*, "Pharyngeal size in snorers, nonsnorers, and patients with obstructive sleep apnea.," *N. Engl. J. Med.*, vol. 315, no. 21, pp. 1327–1331, 1986.

[16]    B. T. Woodson, "Expiratory pharyngeal airway obstruction during sleep: A multiple element model," *Laryngoscope*, vol. 113, no. 9, pp. 1450–1459, 2003.

[17]    B. T. Woodson and R. Franco, "Physiology of Sleep Disordered Breathing," *Otolaryngol. Clin. North Am.*, vol. 40, no. 4, pp. 691–711, 2007.

[18]    A. I. Pack, *Sleep apnea: Pathogenesis, diagnosis and treatment.* CRC Press, 2016.

[19]    K. Sutherland, R. W. W. Lee, and P. A. Cistulli, "Obesity and craniofacial structure as risk factors for obstructive sleep apnoea: Impact of ethnicity," *Respirology*, vol. 17, no. 2, pp. 213–222, 2012.

[20]  S. Redline, B. Cakirer, M. G. Hans, G. Graham, J. Aylor, and P. V. Tishler, "The relationship between craniofacial morphology and obstructive sleep apnea in whites and in African-Americans," *Am. J. Respir. Crit. Care Med.*, vol. 163, no. 4, pp. 947–950, 2001.

# Appendix B. Multi-class Classification of OSA Severity Using Wakefulness Tracheal Breathing Sounds and Random Forest Algorithm

## B.1 Introduction

Obstructive sleep apnea (OSA) syndrome is a serious respiratory disorder. OSA is characterized by recurrent episodes of complete or partial upper airway obstruction during sleep [1]. One of the most important measures of OSA severity is the number of apneic episodes per hour of sleep, termed the apnea/hypopnea index (AHI), obtained using polysomnography (PSG) assessment. Individuals with OSA syndrome can have different severity levels, including non-OSA (AHI<5), mild-OSA ($5 \leq$ AHI<15), moderate-OSA ($15 \leq$ AHI<30), and severe-OSA (AHI$\geq$30) [2].

Accurate detection and classification of individuals with different OSA severity levels are essential for diagnosis cost management and appropriate therapy selection with optimizing compliance to treatment. In one recent study by Nakano et al., the image analysis of the all-night tracheal breathing sound (TBS) spectrogram of participants using a convolutional deep neural network (CNN) was investigated for OSA multi-class classification [3]. As their method requires the full night study assessment, therefore, it is labor-intensive, expensive, and time-consuming. Also, even with reasonable accuracy, estimating the spectrogram of the all-night TBS data besides the use of CNN image analysis makes their method very slow, hence, non-applicable for fast and on-line OSA screening and classification. Therefore, in this Appendix, we expanded our investigations on the use of TBS recorded during wakefulness for multi-class OSA classification, using the Random-Forest (RF) approach (explained in Chapter III). As moderate and severe OSA individuals might have increased cardiovascular or mortality risks and are in need of treatment [4],

therefore, in this study, we combined the data of non-OSA and mild-OSA groups as one group and studied the classification ability of RF to differentiate the non/mild-OSA, moderate-OSA, and severe-OSA individuals.

## B.2  Method

This study's dataset is adopted from our previous research [5], elaborated in Chapter III, and consists of observations taken from 199 individuals suspected of OSA, referred to Misericordia Health Centre (Winnipeg, MB, Canada) for a nocturnal PSG assessment. The Biomedical Research Ethics Board of the University of Manitoba approved the study. All participants signed an informed consent form prior to participating in the study.

When the participants were awake and lay in the supine position with their head resting on a pillow, they were requested to breathe five times through their mouth, with a nose clip in place, followed by five breaths through their nose, with mouth closed. The participants' TBS were recorded using a miniature Sony ECM-77B microphone embedded in a 2mm diameter chamber. The chamber was implanted within a soft neckband and placed over the suprasternal notch of the participants' trachea. The recorded signals were band-pass filtered in the frequency range of [0.05-5000 Hz], amplified using a Biopac (DA100C) amplifier, and digitized using a sampling rate of 10240 Hz. After the wakefulness TBS recording, participants proceeded to an overnight PSG assessment.

### B.2.1  Signal analysis and Feature Extraction

This study's preprocessing phase is similar to our previous works [5], [6]. Briefly, first, we manually separated the inspiratory and expiratory phases of breathing by considering the experimenter's marked voice at the start of inspiration. Then we investigated the signals in the time-and-frequency domain and eliminated the noisy signals. Next, we band-pass filtered the

remaining signals in the range of [75-2500 Hz] and normalized them by their variance envelope and energy. Subsequently, we found the stationary portion of the TBS signals as the segments corresponding to the 50% duration around the maximum of their logarithm of variance [7]. Finally, we estimated the power spectrum density (PSD-using Welch's method) in windows of 20 ms with 50% overlap between adjacent windows for the stationary part of each respiratory phase (inspiration/ expiration) in each breath of mouth and nasal maneuvers and also for the summation and subtraction of the respiratory phases. We calculated the PSD signals on linear and logarithmic scales and averaged them over the five respiratory phases of the participant's data. Overall, we obtain 16 PSD signals for each participant.

Since there may not be much difference in breathing sounds characteristics of individuals with similar AHI values (say AHI of 14 and AHI of 16), we used AHI range groupings with gaps in AHI between them for feature extraction. Thus, for the feature selection part, we only used data of three groups with 178 participants: non/mild-OSA (AHI<10, N=100, 43 males), moderate-OSA (15<AHI<25, N=38, 27 males), and severe-OSA (AHI>30, N=40, 30 males). Data of the mild-OSA (10≤AHI≤15, N=9, 7 males) and moderate/severe-OSA (25≤AHI≤30, N=12, 9 males) groups were dealt with separately to prevent the misclassification in the crisp-nature borderline ranges (elaborated thoroughly in Sections II.2.5 and V.1 of this dissertation).

To avoid sampling bias, we used the 5-fold cross-validation routine to randomly split this set of 178 participants into five non-overlapping groups (folds), each consisting of approximately 36 participants (20% of non/mild-OSA, 20% of moderate/OSA, and 20% of severe-OSA participants). Each time, TBS data related to individuals of one-fold were considered as blind-testing data, and data of the remaining 4 folds (142 participants) were considered as training data and used for feature extraction. In each training dataset and for each of the pre-mentioned 16 PSD

signals, we calculated the average power spectra and their standard error (SE) for the non/mild-OSA, moderate-OSA, and severe-OSA groups. The averages of these power signals in regions with no overlap between the SE's of non/mild-OSA, moderate-OSA, and severe-OSA groups were considered as features for further investigations. We extracted 13 TBS features. Combining these features to individuals' anthropometric information, including sex, age, height, weight, body mass index (BMI), neck circumference (NC), and Mallampati score, led to a total number of 20 features per participant.

## B.2.2 Feature Selection and Classification

After feature extraction, feature selection algorithms were used to remove redundant features and those that do not add any information to our model. In Chapter III, we have explained the RF feature selection and classification model. We have shown that the RF model is a non-parametric approach that does not rely on any model. Thus, this approach is more flexible than the parametric approaches, such as LASSO regularized Logistic Regression [5], and can result in a better final prediction model [6].

Our strategy for 3-class classification is the same as one we explained in Chapter III for 2-class OSA classification [6]. In summary, to find the most proper feature set within each training set we, investigated the classification ability of all possible candidate sets of size 3 among the extracted features of that dataset (i.e. $\binom{\text{\# of features}}{3}$) and selected the one with the best classification result. For the classification purpose, we generated another RF classifier over the selected features of each dataset and classified the participants within the training and blind-testing sets as non/mild-OSA, moderate-OSA, and severe-OSA. We repeated this process ten times and reported the average accuracy for out-of-bag (OOB) data of the training set and blind-testing data within each dataset. The mild-OSA and moderate/severe-OSA participants were also classified using the

selected features of each training dataset, and the percentage of participants assigning to each of the non/mild-OSA, moderate-OSA, and severe-OSA groups was reported. Figure 24 shows the flowchart of this method for feature selection and classification.
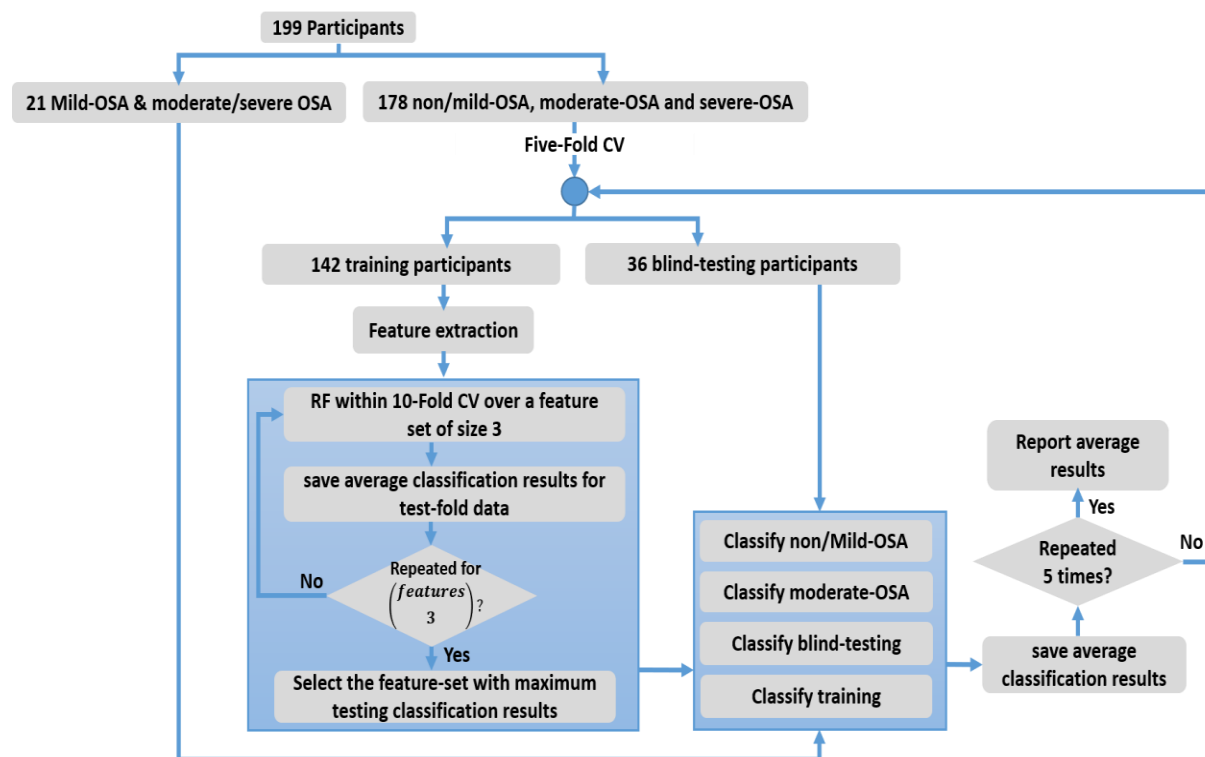


Figure 24 The flowchart of feature selection and classification of non/mild-OSA, moderate-OSA and severe-OSA groups, using the RF approach.

## B.3  Results

The anthropometric information of the participants is reported in Table 23.

Table 23 Anthropometric information's mean and their corresponding standard deviations (std) for the non/mild-OSA, mild-OSA, moderate-OSA, moderate/severe-OSA, and severe-OSA groups. AHI is apnea/hypopnea index, NC is neck circumference, BMI is body mass index, and MP is Mallampati-score.

|  | Non/mild (n=100) | Mild (n=9) | Moderate (n=38) | Moderate/severe (n=12) | Severe (n=40) |
|---|---|---|---|---|---|
| **AHI±std** | 2.8±3.05 | 12.3±1.6 | 19.6±2.7 | 27.5±1.4 | 69.5±33.3 |

| | | | | | |
|---|---|---|---|---|---|
| **NC ($cm$) $\pm$std** | 39.4$\pm$4.8 | 44.3$\pm$6.4 | 42.7$\pm$3.7 | 44.3$\pm$2.1 | 45.3$\pm$3.6 |
| **Age (year) $\pm$std** | 48.4$\pm$12.6 | 50.2$\pm$14.2 | 56.1$\pm$11.4 | 50.5$\pm$10.5 | 49.0$\pm$11.1 |
| **BMI ($kg/m^2$) $\pm$std** | 31.6$\pm$7.0 | 33.7$\pm$10.0 | 34.1$\pm$6.2 | 32.9$\pm$7.1 | 39.7$\pm$8.6 |
| **Sex (Male: Female)** | 43: 57 | 7: 2 | 27: 11 | 9: 3 | 30:10 |
| **Height ($cm$) $\pm$std** | 168.5$\pm$9.6 | 169.4$\pm$10.4 | 168.9$\pm$9.9 | 173.7$\pm$11.7 | 196.9$\pm$10.1 |
| **Weight ($kg$) $\pm$std** | 89.7$\pm$20.9 | 97.9$\pm$22.1 | 97.8$\pm$18.2 | 106.0$\pm$18.0 | 114.4$\pm$26.6 |
| **MP (I-II-III-IV)** | 55-23-13-8 | 4-4-4-1 | 15-13-5-5 | 2-4-3-3 | 5-13-14-8 |

Figure 25 shows the average power spectra calculated from the subtraction of inspiratory and expiratory mouth TBS signals for all participants in non/mild-OSA, moderate-OSA, and severe-OSA groups (178 participants) as well as their corresponding SE in a linear scale.



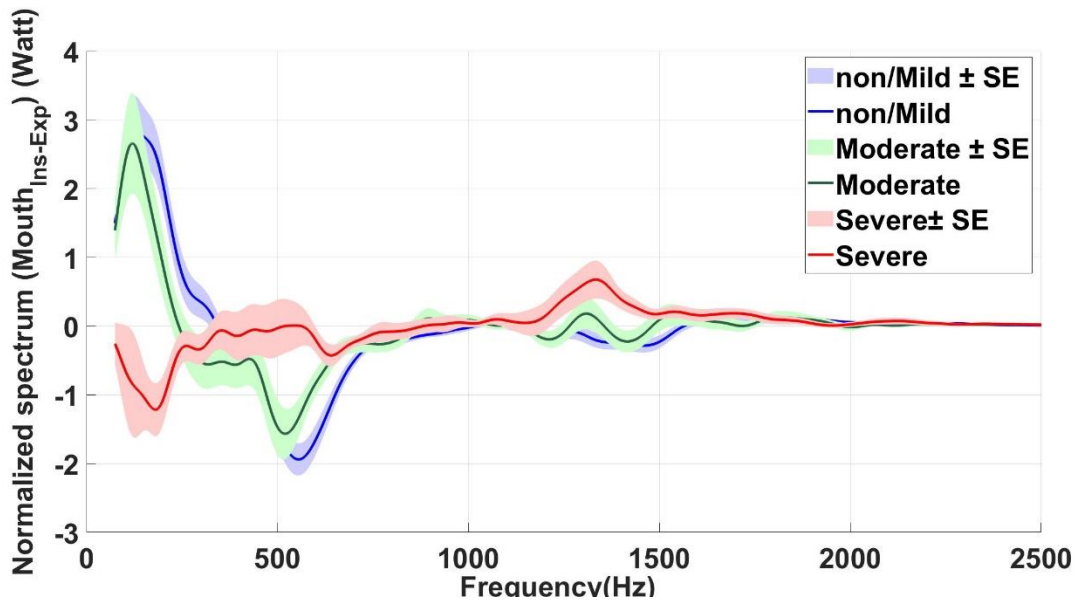Figure 25 Average power spectra of the subtraction of mouth inspiratory and expiratory breathing sounds with their standard errors (SE; shadows) in linear scale for non/mild-OSA (blue), moderate-OSA (green) and severe-OSA (red) groups.

The selected features of each dataset, using the RF method, were among F1 to F5 that are described in Table 24. Table 24 also presents the importance of each feature within each training

set by showing the amount of their mean-decrease-Gini index criterion [8]. This criterion is defined as the decrease in node impurity metric (i.e., Gini-index), weighted by the proportion of data reaching a given node, between the parent and child nodes that the specific feature is splitting, averaged across all trees in the forest.

Table 24 The features selected by the RF feature selection method in different training datasets, for 3-class OSA classification. The values represent the importance of the features by showing the average amount of their mean-decrease-Gini index and their standard deviations (std) over a ten-times run of the RF model. NC is the neck circumference; BMI is body mass index; $PSD_{ME}$ is power spectrum density of mouth expiratory TBS; $PSD_{MI}$ is power spectrum density of mouth inspiratory TBS; $PSD_{NE}$ is power spectrum density of nose expiratory TBS. F presents a feature.

| Training set | F1 $Gini \pm$std | F2 $Gini \pm$std | F3 $Gini \pm$std | F4 $Gini \pm$std | F5 $Gini \pm$std | F6 $Gini \pm$std |
|---|---|---|---|---|---|---|
| 1 | 0.06±0.002 | - | 0.06±0.002 | 0.02±0.002 | - | - |
| 2 | 0.07 ±0.002 | - | 0.03±0.002 | 0.03 ±0.002 | - | - |
| 3 | 0.04 ±0.002 | - | 0.05 ±0.002 | 0.09 ±0.002 | - | - |
| 4 | - | 0.07 ±0.002 | - | - | 0.04 ±0.002 | 0.06 ±0.002 |
| 5 | 0.05 ±0.002 | - | - | 0.03 ±0.002 | - | 0.04 ±0.002 |
| **F1**: Body mass index<br>**F2**: Neck circumference<br>**F3**: $mean(PSD_{NE}(380:430\ Hz))$<br>**F4**: $mean(PSD_{MI} - PSD_{ME}(560:600\ Hz))$<br>**F5**: $mean(PSD_{MI} - PSD_{ME}(160:240\ Hz))$<br>**F6**: $mean(PSD_{ME}(1100:1210\ Hz))$ | | | | | | |

Our findings show a significant correlation (p-value<0.05) between the AHI and the selected features of each dataset. As an example, Figure 26 depicts the correlation between selected features of the third training dataset, in addition to their correlation with AHI.
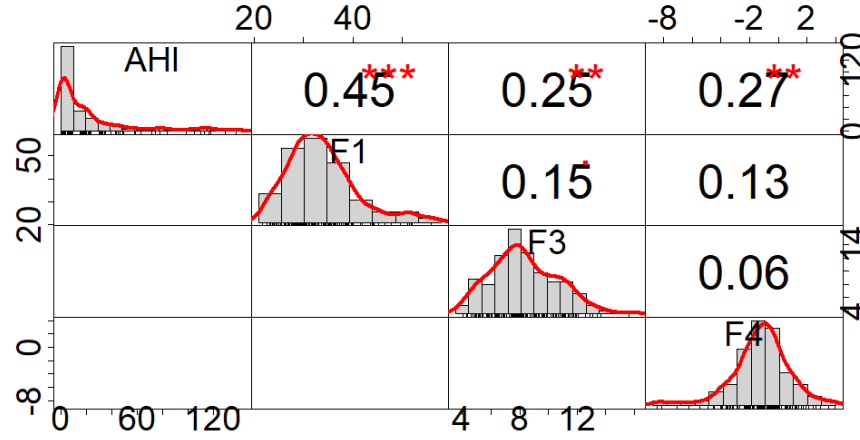
Figure 26 Visualization of the correlation between AHI and the selected features of the RF model on one of the training datasets. On top, the values of the correlation between features plus their significance level as star (***: p<0.001, **: p<0.05). F1 is BMI, F3 is the average of the power spectrum of nose expiratory breathing sounds in the frequency range of (380-430 Hz), and F4 is the average of the power spectrum calculated from the subtraction of mouth inspiratory and expiratory breathing sounds in the frequency range of (560-600 Hz).

Figure 27 shows the 3-dimensional scatter plot of the non/mild-OSA, moderate-OSA, and severe-OSA participants for the selected features of the third training dataset.
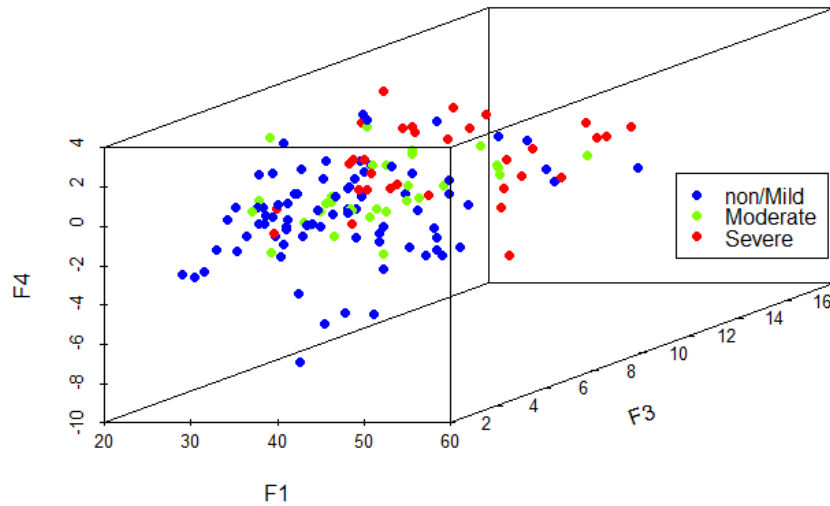


Figure 27 The 3-dimensional scatter plot of the non/mild-OSA, moderate-OSA, and severe-OSA participants for the selected features of the RF model on the data of third training dataset. The selected features were F1 as BMI, F3 as the average of the power spectrum of nose expiratory breathing sounds in the frequency range of (380-430 Hz), and F4 as the average of the power spectrum calculated from the subtraction of mouth inspiratory and expiratory breathing sounds in the frequency range of (560-600 Hz).

166

Our statistical investigation, using the unpaired t-test, have shown that the selected features of each dataset were statistically significantly different from each other. All the p-values were found to be highly significant ($p<2.2\times10^{-16}$).

Table 25 presents the average accuracy for OOB data of training and data of blind-testing sets of non/mild-OSA, moderate-OSA, and severe OSA groups, in addition to the overall accuracy of the RF classifier, within each training set. The average of the RF classification results over the five datasets showed an overall accuracy of 65.3±2.3% for OOB data of training sets and 61.5 ±7.7% for blind-testing data, respectively. The average accuracy of non/mild-OSA, moderate-OSA, and severe-OSA were 80.02±1.0%, 39.8±6.03%, and 52.5±6.1% for OOB data of training sets and 72.3±4.6%, 52.1±14.5%, and 47.7 ±12.3% for blind testing sets, respectively.

Table 25 The classification results for non/mild-OSA, moderate-OSA, and severe OSA participants within different training datasets and their corresponding testing sets, using 10-times run of the RF classifier over its selected features. std is the standard deviation. OOB is the out-of-bag data of training sets in RF.

| Training set | Data | Accuracy Non/mild (%) ±std | Accuracy moderate (%) ±std | Accuracy Severe (%) ±std | Accuracy Overall (%) ±std |
|---|---|---|---|---|---|
| 1 | OOB | 80.8±0.01 | 46.5±0.01 | 46.8±0.04 | 65.9±0.01 |
| | Testing | 76.1±0.01 | 33.9±0.04 | 66.7±0.00 | 64.3±0.02 |
| 2 | OOB | 78.4±0.01 | 33.9±0.02 | 45.9±0.02 | 61.7±0.01 |
| | Testing | 68.7±0.01 | 74.4±0.02 | 52.3±0.04 | 64.9±0.02 |
| 3 | OOB | 80.6±0.01 | 44.2±0.01 | 54.8± 0.04 | 67.0±0.01 |
| | Testing | 66.1±0.02 | 52.2±0.06 | 38.6±0.03 | 57.4±0.01 |
| 4 | OOB | 79.8±0.01 | 41.2±0.03 | 60.5±0.02 | 67.3±0.01 |
| | Testing | 75.5±0.02 | 50.0±0.03 | 36.4±0.02 | 55.4±0.02 |
| | OOB | 80.5±0.01 | 33.1±0.02 | 54.7±0.02 | 64.5±0.01 |

| 5 | Testing | 75.1±0.1 | 50.0±0.1 | 44.6±0.1 | 65.7±0.01 |
|---|---|---|---|---|---|
| | OOB | 80.02±1.0 | 39.8±6.03 | 52.5±6.1 | 65.3±2.3 |
| **Average** | Testing | 72.3±4.6 | 52.1±14.5 | 47.7 ±12.3 | 61.5 ±7.7 |

Table 26 reports the average percentage of the mild-OSA and moderate/severe-OSA participants assigned to non/mild-OSA, moderate-OSA, and severe-OSA groups, using the RF classification procedure. On average, for the mild-OSA group, 62.2±11 participants were assigned to non/mild-OSA, 21.1±7 individuals were assigned to moderate-OSA, and 16.7±9 individuals were assigned to severe-OSA groups. These values were 50.8±15.8, 16.7±14.2, and 32.5±2 for the moderate/severe-OSA group, respectively.

Table 26 The average percentage of Mild-OSA (10≤AHI≤15) and moderate/severe-OSA (25≤AHI≤30) participants classified as non/mild-OSA, moderate-OSA, or severe-OSA, using RF classification procedures, trained with different training datasets.

| Training set | Assigned group / Real group | Non/mild-OSA (%) | moderate -OSA (%) | severe-OSA (%) |
|---|---|---|---|---|
| 1 | Mild | 66.7 ±0.0 | 22.2 ±0.0 | 11.1 ±0.0 |
| | moderate/severe | 62.5±0.04 | 8.3±0.0 | 29.2±0.04 |
| 2 | Mild | 65.28±0.4 | 11.7±0.4 | 23.02±0.04 |
| | moderate/severe | 58.3±0.0 | 8.3±0.0 | 33.4±0.0 |
| 3 | Mild | 74.4±0.06 | 22.8±0.01 | 2.8±0.04 |
| | moderate/severe | 58.3±0.0 | 8.3±0.0 | 33.4±0.0 |
| 4 | Mild | 44.4±0.0 | 33.4 ±0.0 | 22.2±0.0 |
| | moderate/severe | 24.1±0.02 | 41.7 ±0.0 | 34.2±0.02 |
| 5 | Mild | 66.7 ±0.0 | 11.1±0.0 | 22.2 ±0.0 |
| | moderate/severe | 50.0 ±0.0 | 16.7±0.0 | 33.3 ±0.0 |
| Average | Mild | 62.2±11 | 21.1±7 | 16.7±9 |
| | moderate/severe | 50.8±15.8 | 16.7±14.2 | 32.5±2 |

## B.4 Discussion

In this project, we used the RF algorithm for analyzing TBS recorded during wakefulness with the purpose of multi-class OSA classification and severity prediction by a few minutes of breathing sound recording and data analysis. We used the wakefulness TBS data to extract features from the sounds' power spectra. Despite the fast speed of our analysis, explained in Section III.3.1 of this dissertation, our results show that combining the individuals' anthropometric information with their spectral TBS characteristics might not be enough for multi-class OSA classification during wakefulness. It suggests the need to use higher-order statistics of the TBS signals, besides the current features, to extract more discriminative characteristics for multi-class OSA classification. The power spectrum describes the distribution of energy (variance) with frequency; thus, it is associated with the signal's second-order statistics. However, it is known that only the Gaussian probability distributions are totally specified by their statistics up to order two. In general, the real-world signals contain more information than those accessible through second-order techniques; therefore, higher-order techniques are required. The bispectrum is the third-order analog of the power spectrum and describes the distribution of skewness with frequency. Besides, the bispectrum-based signal processing allows extracting the phase information. Consequently, higher-order statistics might help extract novel information features useful for discrimination and multi-class OSA classification.

Another reason for the poor multi-class OSA classification results of this study might be related to using the AHI as the true label of OSA participants. Despite AHI is one of the most important OSA severity measures and is the commonly used indicator for OSA diagnosis, many individuals with similar AHI have different clinical symptoms, therefore, different OSA severity levels. Also, it has been shown that the UA of individuals with the majority of apnea relative to hypopnea events

169

are more collapsible than the UA of individuals with the majority of hypopnea relative to apnea events [9]. However, AHI considers similar weight for both apnea and hypopnea events. Different UA collapsibility levels imply different structural and physiological properties of the UA, which might have a different effect on TBS [10]–[12].

In addition, the results presented in Table 25 show that prediction error between classes is highly unbalanced, i.e., non/mild-OSA group has a low prediction error, while moderate-OSA and severe-OSA groups have a high prediction error. Such unbalanced prediction error usually occurs when the size of one group (non/mild-OSA) is much larger than another (moderate and severe-OSA), which was the case in this study. It is because the RF classifier mainly using data from the larger class to train the model. In theory, it has been explained that in the RF model, error balancing can be done by assigning different weights to different classes according to their size. However, since in our studies, we used the *randomForest package of R Version 3.6.3,* where the weighting is not yet implemented, we considered adding the weighting ability to our RF model as our future goal. It should be added that we repeated the whole feature selection and classification procedures on an equal number of participants in each OSA group. Surprisingly, the results were only slightly better. This analysis strengthens the idea that both the current AHI definition and the spectral features extracted from the regions of the non-overlapped area between the three OSA classes are not enough for multi-class classification during wakefulness, especially for separating the moderate and severe OSA groups from each other.

It is worth also noting that OSA is a heterogeneous disorder, and various factors, including anatomical abnormalities, impaired UA dilator muscle activity, instability of ventilatory control (high loop gain), and low arousal threshold, are major contributors to OSA development [13], [14]. The non-anatomical contributors make OSA prediction challenging during wakefulness.

To sum up, we believe that for improving the multi-class OSA classification during wakefulness one has to consider a developed measure of OSA severity instead of AHI, which considers adjusted weights for apnea/hypopnea events and accounts for daytime symptoms of OSA as well. Besides, a balanced dataset with an approximately equal number of participants between various OSA groups that also matched in terms of anthropometric information should be considered.

## Reference

[1]     A. Malhotra and D. P. White, "Obstructive sleep apnoea," *Lancet*, vol. 360, no. 9328, pp. 237–245, Jul. 2002.

[2]     American Academy of Sleep Medicine, *Hidden Health Crisis Costing America Billions: Underdiagnosing and Undertreating Obstructive Sleep Apnea Draining Healthcare System*. Mountain View, CA: Sullivan & Frost, 2016.

[3]     H. Nakano, T. Furukawa, and T. Tanigawa, "Tracheal sound analysis using a deep neural network to detect sleep apnea," *J. Clin. Sleep Med.*, vol. 15, no. 8, pp. 1125–1133, 2019.

[4]     Terry Young, M. Palta, Jerome Dempsey, P. E. Peppard, F. J. Nieto, and K. Mae Hla, "Burden of Sleep Apnea: Rationale, Design, and Major Findings of the Wisconsin Sleep Cohort Study," *WMJ Off. Publ. State Med. Soc. Wisconsin*, vol. 108, no. 5, pp. 246–249, Nov. 2010.

[5]     F. Hajipour, M. Jafari Jozani, A. Elwali, and Z. Moussavi, "Regularized logistic regression for obstructive sleep apnea screening during wakefulness using daytime tracheal breathing sounds and anthropometric information," *Med. Biol. Eng. Comput.*, vol. 57, no. 12, pp. 2641–2655, Nov. 2019.

[6]     F. Hajipour, M. J. Jozani, and Z. Moussavi, "A comparison of regularized logistic regression and random forest machine learning models for daytime diagnosis of obstructive sleep apnea," *Med. Biol. Eng. Comput.*, vol. 58, no. 10, pp. 2517–2529, Oct. 2020.

[7]     A. Yadollahi and Z. M. K. Moussavi, "Acoustical flow estimation: Review and validation," *IEEE Eng. Med. Biol. Mag.*, vol. 26, no. 1, pp. 56–61, Jan. 2007.

[8]     L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[9]     I. C. Gleadhill, A. R. Schwartz, N. Schubert, R. A. Wise, S. Permutt, and P. L. Smith, "Upper airway collapsibility in snorers and in patients with obstructive hypopnea and apnea," *Am. Rev. Respir. Dis.*, vol. 143, no. 6, pp. 1300–1303, 1991.

[10]    T. Penzel and A. Sabil, "The use of tracheal sounds for the diagnosis of sleep apnoea," *Breathe*, vol. 13, no. 2, pp. e37–e45, Jun. 2017.

[11]    I. Ayappa and D. M. Rapoport, "The upper airway in sleep: physiology of the pharynx,"

*Sleep Med. Rev.*, vol. 7, no. 1, pp. 9–33, Feb. 2003.

[12]  I. Sanchez and H. Pasterkamp, "Tracheal sound spectra depend on body height.," *Am. Rev. Respir. Dis.*, vol. 148, no. 4 Pt 1, pp. 1083–1087, 1993.

[13]  D. J. Eckert, A. Malhotra, and A. S. Jordan, "Mechanisms of Apnea," *Progresas Cardiovasc. Dis.*, vol. 51, no. 4, pp. 313–323, Jan. 2009.

[14]  S. Javaheri *et al.*, "Sleep Apnea: Types, Mechanisms, and Clinical Cardiovascular Consequences," *J. Am. Coll. Cardiol.*, vol. 69, no. 7, pp. 841–858, 2017.