Machine Learning-Based Obstructive Sleep Apnea Screening Using Wakefulness Tracheal Breathing Sounds and Anthropometric Information

By

Vahid Bastani Najafabadi

A Thesis submitted to the Faculty of Graduate and Postdoctoral Studies of

The University of Manitoba

In partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Electrical and Computer Engineering

University of Manitoba

Winnipeg

Abstract

Obstructive sleep apnea (OSA) is a common sleep disorder that remains underdiagnosed due to the cost, complexity, and limited accessibility of standard diagnostic methods such as polysomnography (PSG). Accessible and objective screening approaches to identify individuals in need of treatment for OSA before confirmatory sleep assessment will significantly help with the backlog of PSG sleep sessions. Previous studies have demonstrated the feasibility of OSA screening during wakefulness using tracheal breathing sounds (TBS); however, these frameworks were predominantly developed using specific recording hardware and acquisition configurations. Therefore, it remains unclear whether diagnostically relevant wakefulness TBS information can be preserved using a substantially lower-cost microphone with different electroacoustic characteristics and whether previously developed frameworks can be transferred directly to such recordings. Given that the presentation of OSA is affected by individual characteristics such as age, sex, body mass index (BMI), neck circumference, smoking history and craniofacial structure, screening frameworks should consider both physiological features and anthropometric covariates. This thesis develops a machine learning framework for wakefulness OSA screening using tracheal breathing sounds (TBS) recorded with a low-cost microphone in a practical clinical environment. Data of 247 individuals with various degrees of OSA severity were analyzed. Recorded data were segmented

into inspiration and expiration phases, followed by acoustic features extraction, feature reduction, and classification. A two-level ensemble architecture was implemented. Nine sub-classifiers were stratified by anthropometric profiles, which each sub-classifier was constructed as an ensemble of bagged decision trees, with a final prediction via probability-based voting. The proposed algorithm achieved an accuracy of 77.1%, sensitivity of 84.3%, and specificity of 59.9%. In contrast, applying the previously developed AWakeOSA framework directly to the current recordings resulted in substantially lower performance, indicating limited transferability to the present acquisition setting. These findings demonstrate that diagnostically useful wakefulness TBS information can be obtained using a low-cost microphone while suggesting that changes in recording hardware and acquisition conditions may require adaptation of previously developed classification frameworks. Future studies should validate the framework using larger and more balanced independent datasets, investigate intermediate AHI cases, and directly evaluate microphone-related effects using paired recordings under controlled acquisition conditions.

Keywords: Obstructive Sleep Apnea (OSA), Wakefulness Screening, Tracheal Breathing Sounds, Wavelet Packet Decomposition, Ensemble Machine Learning, Probability-Based Voting.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Zahra Moussavi, and co-advisor, Dr. Ahmed Elwali, for their invaluable guidance, immense patience, and academic support throughout this research. Their combined insights, expertise, and encouragement were pivotal to the completion of this thesis.

To my partner, thank you for being my rock, for your unwavering belief in me, and for standing by my side through every long night and challenging milestone of this journey. I am profoundly grateful to my family and friends, whose love, understanding, and constant encouragement kept me grounded and motivated.

I want to hold a special space in remembrance of our beloved friend, Ako, who passed away in June 2026; your presence is deeply missed, and your memory remains a source of light. Finally, my heart and thoughts are forever with the people of Iran who were tragically killed during the protests in 2026. This academic milestone was completed with their enduring struggle for justice and freedom heavy on my mind.

Dedication

Dedicated to my family and my partner for their endless love and support;

To the memory of my late friend, Ako;

And to the people of Iran who stood and fell for freedom.

Contents

List of Figures

Chapter 1: Introduction

Sleep disorders can substantially degrade sleep quality and overall quality of life, and sleep apnea is a prevalent sleep-related breathing disorder. Sleep apnea is generally categorized into central sleep apnea (CSA), obstructive sleep apnea (OSA), and mixed sleep apnea. Among these types, OSA is the most common and is the focus of this thesis. OSA is characterized by recurrent episodes of partial (hypopnea) or complete (apnea) cessation of airflow during sleep due to upper airway obstruction [1]. OSA severity is commonly quantified using the apnea-hypopnea index (AHI), which represents the number of apnea and hypopnea events per hour of sleep [2]. AHI is typically determined through polysomnography (PSG), which is an overnight sleep test conducted in a sleep clinic while multiple physiological parameters are measured simultaneously during sleep [3]. OSA can disturb sleep patterns, reduce oxygen saturation, and contribute to symptoms such as morning headache, excessive daytime sleepiness, and impaired neurocognitive function [4]. Beyond these symptoms, untreated OSA is associated with serious health consequences, including cardiovascular and cerebrovascular disease, type 2 diabetes, and increased mortality in severe OSA individuals [5–9]. Age, sex, body mass index (BMI), neck circumference, and smoking status are some risk factors associated with OSA risk and severity [10]. Given its high global prevalence and

substantial public health burden, early identification of individuals at risk of OSA remains an important clinical need.

Accurate OSA diagnosis generally requires individual monitoring of physiological signals to assess respiratory events, sleep fragmentation, and OSA severity through PSG, which is considered the gold standard of OSA diagnosis [11]. During an overnight PSG, multiple physiological signals are simultaneously recorded, including brain activity, eye movement, muscle activity, airflow, respiratory effort, and blood oxygen saturation [3,11–13]. However, PSG is expensive, resource-intensive, time-consuming, and its accessibility can be limited by the availability of specialized sleep clinics [14,15]. It may also alter natural sleep patterns because patients are required to sleep in an unfamiliar laboratory environment [16]. Home sleep apnea testing (HSAT) devices improve accessibility and reduce complexity, but they record fewer physiological signals and may underestimate OSA severity due to calculations of respiratory events based on total recording time rather than actual sleep time because brain activity recording is missing [17,18]. Questionnaires are inexpensive and easy to administer, but their reliance on self-reported information limits their ability to provide objective physiological assessment. Although questionnaires can achieve relatively high sensitivity, their specificity is generally lower and varies across clinical populations

[19]. Therefore, there is a need for screening methods that are accessible while still incorporating objective physiological information.

Machine learning has increasingly been used for OSA screening by extracting discriminative patterns from physiological and acoustic signals, including brain activity, cardiovascular activity, speech, and respiratory sounds. Previous studies have investigated several modalities, including electroencephalogram (EEG) [20], cardiovascular signals [21], speech [22], and tracheal breathing sounds (TBS) [23–25]. Although EEG-based approaches can characterize OSA-related changes during sleep, they generally require overnight acquisition and multiple body-mounted electrodes. Cardiovascular methods can provide informative physiological features but still commonly rely on physiological changes associated with sleep-time respiratory events. Speech-based approaches offer convenient wakefulness acquisition; however, their acoustic representation can depend on the selected phonetic tasks and speech characteristics [22]. In comparison, TBS analysis provides an objective respiratory-acoustic modality that can capture acoustic information reflecting airflow dynamics and upper airway characteristics [26,27]. Importantly, TBS can be acquired during wakefulness using a relatively simple and short breathing protocol, avoiding the need to wait for sleep before collecting the screening signal [26].

Analyzing TBS recorded during wakefulness is particularly relevant for practical OSA screening because anatomical and mechanical upper airway characteristics associated with OSA can remain detectable while the individual is awake [28]. This concept is supported by the acoustic airway hypothesis, which states that sounds generated by airflow through the trachea carry information about upper airway geometry [27]. Previous studies have demonstrated the feasibility of wakefulness TBS-based OSA screening, including as anthropometric subgroup-based framework that achieved 81.4% blind test accuracy, with 80.9% sensitivity and 82.1% specificity in 199 participants [23,25,29–31].

However, previous wakefulness TBS frameworks were generally developed under structured condition using specific recording hardware. For example, several studies used a Sony ECM77B microphone positioned within a fixed chamber over the suprasternal notch while participant performed controlled breathing maneuvers in the supine position [25,28,29]. The relative high cost of this microphone can limit the accessibility and scalability of such a screening approach, motivating investigation of lower-cost alternatives. However, changing the recording microphone may also change the acoustic representation of TBS because microphones can differ in their frequency response characteristics. Therefore, it remains unclear whether a low-cost microphone can preserve diagnostically relevant wakefulness TBS information and whether a framework developed using Sony

ECM77B can generalize directly to recordings acquired using a different microphone. A separate challenge arises from the recording environment, as practical clinical recordings may contain environmental noise and other acoustic components in addition to the target TBS. These challenges should therefore be distinguished from the cost of the microphone itself.

Population heterogeneity presents an additional challenge because anthropometric factors such as age, sex, BMI, neck circumference, and Mallampati score (MpS) can influence both OSA risk and TBS characteristics, while unequal representation of these phenotypes can limit the development and validation of subgroup-specific classifiers [25,32,33]. Consequently, further investigation is required to determine whether diagnostically useful wakefulness TBS can be obtained using a low-cost microphone under practical clinical recording conditions, whether previously developed TBS-based frameworks remain transferable to this acquisition setting, and whether a dedicated subgroup-based machine learning framework can provide more robust screening across a heterogeneous study population.

This thesis consists of six chapters. Chapter 2 presents the background required to understand OSA, including its pathophysiology, health consequences, risk factors, and current diagnostic methods. It also reviews machine learning approaches for OSA screening using physiological signals, with a specific focus on

wakefulness TBS studies, their recording protocols and analytical methodologies, and the limitations of previous work. Chapter 3 describes the methodology used in this thesis. It includes the study population and recording protocol, signal preprocessing, feature extraction and selection, subgroup definition, ensemble classifier development, and training strategy. Chapter 4 presents the performance of the proposed wakefulness TBS machine learning framework, including comparison with baseline models and analysis of classification outcomes. Chapter 5 discusses the main findings, clinical and technical implications, limitations, and future directions of the study. Finally, chapter 6 concludes the thesis by summarizing the main contributions and recommendations for future work.

Chapter 2: Background

Many individuals suffer from sleep disorders that degrade quality of sleep and consequently quality of life. Sleep apnea is a common sleep disorder defined as recurrent reductions or cessations of breathing during sleep. It is typically categorized as central sleep apnea (CSA), which involves impaired respiratory drive from the central nervous system during sleep [34]; obstructive sleep apnea (OSA), which results from the upper airway collapse during sleep [1]; and mixed sleep apnea, which is a combination of both central and obstructive sleep apnea [35]. OSA is the most common form of sleep apnea and is the main focus of this thesis. A recent study estimated its prevalence at 32.4% among U.S. adults aged 20 years and older in 2024 [36]. CSA is less common, accounting for approximately 5–10% of participants with sleep-related breathing disorders [37]. Mixed sleep apnea is less consistently reported; however, a study reported 5–20% of patients with untreated OSA may have mixed sleep apnea [38].

2.1 Obstructive Sleep Apnea (OSA)

2.1.1 Definition and Pathology

OSA is a sleep disorder characterized by repetitive partial or complete obstruction of the upper airway during sleep [1]. The complete blockage called apnea is a ≥ 90% drop in airflow from baseline lasting at least 10s [39], while partial

blockage called hypopnea, is defined as ≥ 30% reduction in airflow compared to baseline for at least 10s and is associated with either a ≥ 3% oxygen desaturation or an arousal [40]. The severity of OSA is commonly measured by the apnea-hypopnea index (AHI), which is number of apnea plus hypopnea events per hour of sleep [2]. Based on AHI, OSA severity is categorized into mild (5 ≤ AHI < 15), moderate (15 ≤ AHI < 30), and severe (30 ≤ AHI) [41].

The primary pathophysiology of the upper airway obstruction in OSA involves repetitive collapse of the pharyngeal airway during sleep [42]. The pharyngeal airway is a compliant segment surrounded by soft tissues and lacks rigid skeletal support along much of its length, making it susceptible to narrowing or collapse during sleep [43,44]. In healthy individuals, the pharyngeal patency is maintained by a balance of forces between the negative intraluminal pressure generated during inspiration (that promotes pharyngeal closure) and compensatory activation of the pharyngeal dilator muscles, particularly the genioglossus (that stints the airway open) [45]. The detailed anatomy of the upper airway is depicted in Figure 1.

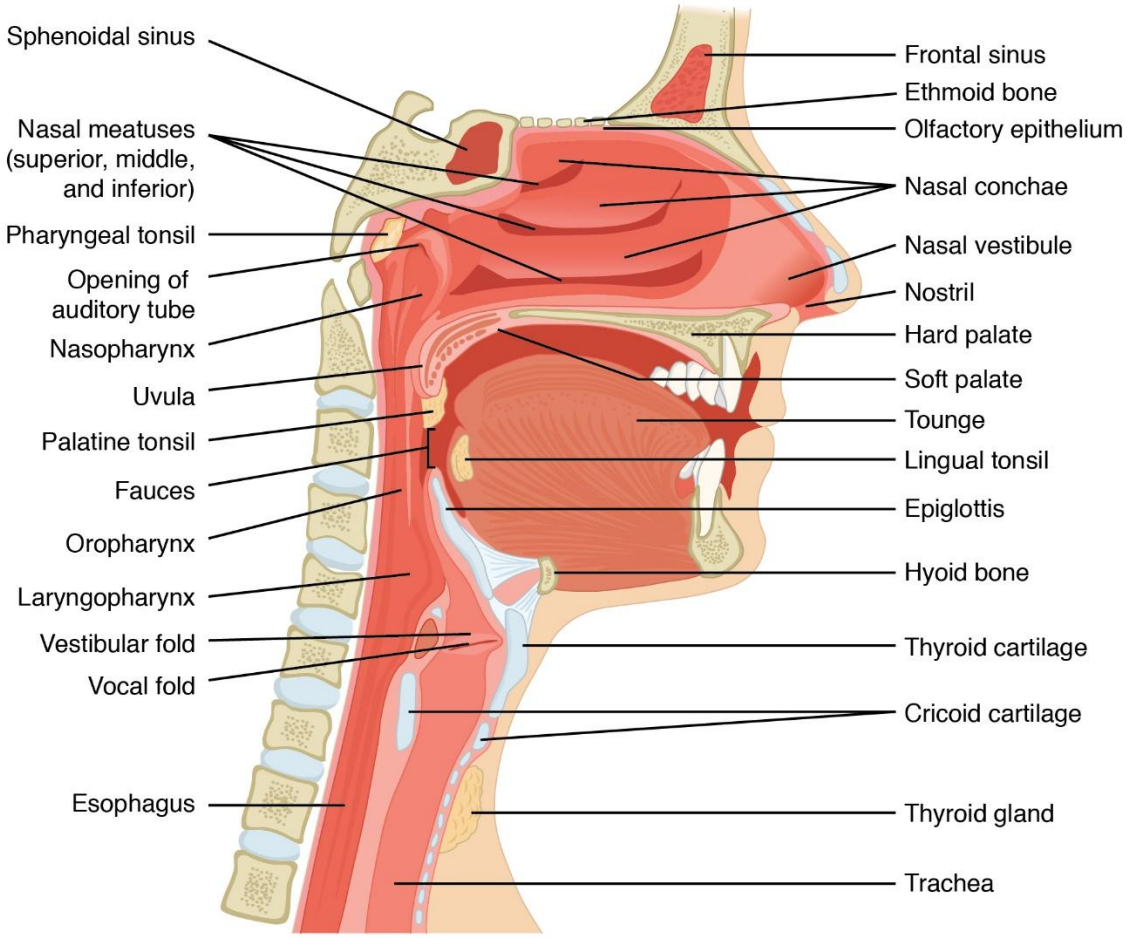

Figure 1. Anatomy of the upper airway [46]

In patients with OSA, the pharyngeal airway requires compensatory activation of the upper airway dilator muscles to maintain patency during wakefulness. During sleep, the reduction in this compensatory neuromuscular activity can become insufficient to counteract the collapsing forces acting on the airway [47]. In the presence of anatomical factors that increases pharyngeal collapsibility, such as reduced airway dimensions or increased soft tissue volume, this loss of compensatory support may result in partial or complete obstruction of the upper airway during sleep [47]. Although anatomical collapsibility is fundamental to OSA,

its pathogenesis is multifactorial and can also involve impaired upper airway muscle responsiveness, ventilatory control instability, and altered arousal response [48].

2.1.2 Prevalence and Health Consequences

OSA can disrupt sleep continuity and disturbs the sleep pattern and can affect the individual's quality of life (e.g., morning headaches, excessive daytime sleepiness, and impaired neurocognitive function) [4,49,50]. OSA can cause repeated episodes of oxygen desaturation during sleep and sleep fragmentation [5], which can contribute to cardiovascular, metabolic, and neurocognitive complications. OSA is also associated with an increase in blood pressure and may elevate the risk of cardiovascular and cerebrovascular diseases [6,51]. In addition, severe OSA is also associated with type II diabetes 2 both clinical experiments and epidemiological populations [7]. Stroke is further associated with moderate to severe OSA, independent of confounding factors [8]. Severe OSA (AHI $\geq$ 30) has also been associated with increased all-cause and cardiovascular mortality [9].

Considering the severity of mentioned morbidities, the high global prevalence of OSA presents a critical public health concern. Approximately 936 million people around the world aged 30-69 years old are affected by OSA, including approximately 425 million with moderate to severe OSA [52]. In Canada, a recent national study estimated the prevalence of moderate to severe OSA at 28.1% among adults aged 45-85 years using STOP-Band risk assessment questionnaire, while only 1.2% had

a clinical diagnosis [53]. Furthermore, the recognized OSA prevalence increased from 0.14% in 2003 to 4.59% in 2020 in Alberta, Canada [54]. These estimations indicated that OSA is highly prevalent globally and frequently underdiagnosed, supporting the need for accessible and effective screening and diagnostic approaches.

2.1.3 Risk factors

Previous studies have investigated several significant factors that increase the risk of OSA. Being male is a major risk factor; males generally experience more severe OSA compared to females [55–57]. There are other differences in females and males in relation to OSA. Females with sleep apnea exhibit a higher AHI during rapid-eye-movement (REM) sleep, whereas males with sleep apnea show higher AHI in non-rapid-eye-movement (NREM) sleep, and higher AHI in supine position [58]. Snoring and daytime sleepiness are OSA symptoms in males, while insomnia, restless legs, depression, and nightmares are several common symptoms in females with OSA [59,60].

Age is another risk factor for OSA, as older adults are at a higher risk of developing OSA [61,62]. According to a Canadian longitudinal study on aging in 2022 [63], one in five adults aged more than 45 years is at risk of experiencing OSA. Symptomatic presentation also varies with age, as elderly with OSA typically experience less snoring [64] and daytime sleepiness [65] than younger populations.

Fatigue, memory disturbance, or nocturia are also more apparent in elderly OSA patients [66].

Increased body mass index (BMI) elevates the risk of developing OSA [67], as gaining weight can influence the pathogenesis of OSA by worsening upper airway collapsibility [68]. Longitudinal data from a cohort of 690 adults indicated that a 10% increase in body weight was found to increase the AHI by approximately 32% and resulted in a six-fold increase in the odds of developing moderate-to-severe OSA [69]. Studies suggest that weight loss can result in a reduction of AHI, restoring sleep architecture, and lowering the requirement for OSA therapies [70–72].

Smoking has also been claimed to increase the risk of OSA regardless of other risk factors, and current smokers are significantly more likely to suffer from OSA compared to former and non-smokers [73,74]. A study on 151 individuals with OSA found that current smokers had a 3.7 times higher odds of having severe OSA than non-smokers [75], and increasing the smoking level could result in an increase in OSA severity and a reduction in the oxygen saturation overnight [76]. On the other hand, some studies have found no direct relationship between OSA and smoking level. In a study on 733 of individuals with OSA, it was reported that smokers had significantly higher AHI for total sleep time; however, after adjusting for age, sex, and BMI, there was no significant correlation between smoking and OSA severity [77]. While another study detected a higher prevalence of severe OSA in male

smokers with BMI < 30, there was no significant difference amongst non-smokers and smokers [78]. A supporting analysis conducted as part of this thesis further examined smoking status in relation to AHI across demographic and anthropometric subgroups [74]. Findings of this analysis revealed that current-smokers had significantly different AHI values compared to former-smokers and non-smokers, particularly within male and high neck circumference subgroups. However, the effect of smoking status was attenuated after adjustment for age, sex, BMI, and MpS, suggesting that smoking should be interpreted alongside other OSA-related covariates. A detailed supporting statistical analysis of smoking status and AHI is provided in Appendix A.

Anatomical factors can also play an important role in developing OSA. Mallampati score (MpS), which evaluates mouth opening and computes the tongue's size in relation to the oral cavity [79], has been shown to be significantly correlated with OSA severity [80]. A study stated that for each 1-point elevation in MpS, the likelihood of having OSA increased more than 2-fold and the AHI rose by more than 5 events per hour [81]. Furthermore, a receding mandible position, specifically a small and shallow mandible, serves as an independent risk factor for the development of OSA, particularly in males [82]. Moreover, neck circumference (NC) is significantly correlated with OSA, with a higher NC significantly increasing the odds of having OSA [83–85].

2.2 Current Clinical Diagnostic Methods

The accurate diagnosis of OSA typically relies on the objective screening of physiological signals to quantify respiratory events and sleep fragmentation to compute AHI and OSA severity. This section provides an overview of the two current diagnostic methods, PSG and HSAT, followed by their diagnostic precision and existing limitations.

2.2.1 Polysomnography as the Gold Standard

The gold standard for detecting OSA involves PSG [13], which is conducted in a supervised sleep laboratory, where multiple physiological parameters are monitored during overnight sleep [86]. This includes EEG for arousal detection [12], electrooculography (EOG) for eye movement and sleep staging [87], EMG for muscle tone and limb movement [88], ECG for heart rhythm to detect arrhythmias linked to sleep apnea [89], and oronasal thermal sensors and nasal air pressure transducers for airflow [90], while pulse oximetry is used to continuously monitor oxygen saturation levels [91]. Crucially, thoracic and abdominal inductance plethysmography (RIP) belts are used in PSG to distinguish between obstructive and central respiratory events by measuring respiratory effort [92].

2.2.2 Limitations of PSG

While PSG is considered the gold standard for OSA diagnosis, it has several significant limitations, including high costs, long wait times, and often not being easily accessible to many patients [93,94]. Furthermore, sleeping in an unfamiliar environment like a sleep laboratory for conducting PSG can induce stress and anxiety in the patient and lead to the first-night effect (FNE). FNE is characterized by reduced total sleep time, prolonged sleep onset latency, and diminished sleep efficiency, thereby distorting the sleep architecture [16,95]. Conversely, some patients may sleep more comfortably than usual, a phenomenon known as the reverse first night effect (RFNE) [96]. These challenges and limitations have encouraged researchers to investigate alternative accurate diagnostic tools.

2.2.3 Home Sleep Apnea Testing (HSAT) and Its Constraints

To address the logistical and economic limitations of PSG, HSAT has emerged as a widely used alternative for diagnosing OSA. The most common HSAT systems in use are Type 3 and Type 4 devices; these are unattended monitors that recorded limited parameters, typically including respiratory effort, oxygen saturation, airflow, and heart rate [18,97]. While cost-effectiveness, high accessibility, unsupervised administration, and reduced complexity are primary advantages of HSAT, the risk of technical failure due to lack of supervision is significant [17]. Additionally, HSAT does not include standard sleep staging sensors (e.g., EEG, EOG), so the device is

unable to distinguish sleep from wakefulness. Thus, the event frequency, termed the respiratory event index (REI), is calculated based on the total recording time rather than the actual total sleep time as the denominator that can often dilute the severity of OSA compared to PSG [17]. Given these limitations, identifying at-risk individuals requires accessible, first-line screening methods that can effectively stratify patients before deploying resource-intensive diagnostic tools.

2.3 Questionnaire-Based OSA screening tools

Given the high prevalence of OSA and significant resource constraints associated with standard diagnostic tools (PSG and HSAT), efficient risk stratification is essential. Screening tools act as a critical filter to prioritize high-risk patients for confirmatory diagnosis, thereby reducing waiting times and optimizing healthcare resource allocation. These screening approaches are categorized into subjective clinical questionnaires and objective physiological signal-based systems.

2.3.1 Clinical Questionnaires

Clinical questionnaires serve as the primary first-line screening method due to their low-cost, ease of use, and immediate results. These tools typically combine subjective symptom reporting (e.g., snoring, daytime sleepiness) with objective anthropometric measurements (e.g., BMI, neck circumference) to estimate the probability of OSA. The most widely validated questionnaires include:

- **STOP-BANG:** This is currently the most popular screening tool that estimates the risk of OSA based on several specific risk factors: gender, age, BMI, neck circumference, blood pressure, observed apnea, daytime sleepiness, and snoring [98]. While STOP-Bang questionnaire can effectively diagnose OSA individuals, it has a moderate discriminative ability to identify non-OSA individuals [99]. Many of the STOP-Bang components such as older age, male sex, obesity, and increased neck circumference, are established OSA risk factors but are not specific to OSA and may also be present in individuals without OSA.

- **Epworth Sleepiness Scale (ESS):** ESS is a simple subjective study that measures the level of daytime sleepiness, particularly, by asking the participants to rate their likelihood of dozing off in eight different situations (e.g., talking to someone, watching TV) [100]. Although age is a key risk factor in developing OSA [61], ESS may not be an appropriate screening tool for older adults since it may underestimate the sleepiness severity in this population [101].

- **Berlin Questionnaire (BQ):** This is a self-administered questionnaire that flags individuals who need further OSA assessment by measuring various anthropometric and symptomatic features within three categories, involving snoring and breathing pause, daytime sleepiness, and blood pressure [102]. A

key limitation of BQ is that the scoring can be inconsistent over time in clinical patients and may reflect different aspects of sleepiness in different groups, suggesting that the same score may not mean the same thing across populations and should be interpreted alongside clinical history and other screening systems [103].

- **NoSAS:** This questionnaire assigns different weights to five factors, including neck circumference, BMI, snoring, age, and male, resulting in a score ranging from 0 to 17, with scores higher than eight indicating a high probability of having OSA [104]. However, a study stated that slim bodies such as individuals with low BMI or neck circumference may not be effectively diagnose by NoSAS [105].

2.3.2 Limitations and Motivations for Objective screening

Despite the advantages of the mentioned OSA screening questionnaires, their subjective components and limited generalizability across populations make them insufficient method for OSA diagnosis. Their reliance on self-report introduces reporting bias and feasibility issues (e.g., inconsistent interpretation of sleepiness questions), and their risk score weighting may not generalize across heterogeneous populations. Consequently, many individuals still require objective evaluation to confirm OSA presence and severity. These limitations have motivated approaches that use physiological signals and machine learning to extract objective biomarkers

associated with OSA and to improve robustness beyond risk scores based on symptoms.

2.4 Machine Learning for OSA Screening Using Physiological Modalities

Machine learning has increasingly been applied to OSA screening that uses physiological signals that reflect different manifestations and consequences of sleep-disordered breathing. These methods learn discriminative patterns from signals that reflect sleep fragmentation and arousal activity, respiratory dynamics, and cardiovascular responses. By transforming raw biosignals into informative features, machine learning frameworks can provide more objective risk probability and, in some cases, approximate severity estimation. This performance often depends on sensor burden, recording protocol, and population variability.

2.4.1 EEG-Based OSA Assessment Methods

EEG provides direct information on cortical activity and is commonly used for sleep staging and identification of arousals, making it a potentially informative signal for OSA assessment. A study extracted spectral, graph, and microstate features from 19-channel EEG recordings of 38 participants and reported 88.3% accuracy, 92% sensitivity, and 84% specificity using adaptive boost (AdaBoost) classifier [20]. Another study used a single C4–A1 EEG channel and db8 DWT-based statistical

features with an Extra Trees classifier to distinguish apnea from non-apnea events, reporting an average accuracy of 86% [106]. A further study reported 91% accuracy, 88% sensitivity, and 93% specificity using K-Nearest Neighbour (KNN) classifier, trained on time-domain, spectral, and nonlinear features extracted from 44-channel EEG recordings of 25 individuals [107]. Despite the efficacy of OSA screening using EEG signals, lengthy acquisition procedure, electrode placement and signal quality control, and susceptibility to artifacts (e.g., eye movements, breathing, and muscle activity) are some of the challenges and limitations which can affect generalizability and reproducibility across populations.

2.4.2 Cardiovascular Signal-Based Methods

OSA is also associated with cardiovascular dysfunction through interacting mechanisms that include intermittent hypoxemia, recurrent arousals, increased sympathetic activity, and intrathoracic pressure fluctuations [108]. These physiological responses can produce characteristic changes in cardiac rhythm and autonomic regulation, motivating the use of electrocardiographic signals for automated OSA assessment [108]. A study utilized a one-dimensional convolutional neural network (1D-CNN) and bidirectional gated recurrent units (Bi-GRU) to detect OSA, trained on a single-lead ECG recording of 70 individuals, achieving 95.19% accuracy, 93.45% sensitivity, and 96.12% specificity [109]. By applying XGBoost to single-lead ECG derived heart rate variability (HRV) features of 18 participants,

researchers achieved recall of 96%, precision of 95% and an F1-score of 95.5% in classifying apnea from non-apnea states [110]. Another study evaluated nine notch-filtered ECG features across 19 machine leaning and four deep learning models, that achieved an 86.25% accuracy in identifying OSA events across 70 recordings [111]. A primary limitation of these studies is their reliance of ECG data derived from public repositories or clinical PSG experiments. Consequently, their high event detection performance relies on cardiovascular responses occurring during sleep. In addition, several studies used relatively small cohorts or public datasets, requiring further subject-independent and external validation before clinical generalization.

2.4.3 Speech-Based Methods

Previously, the characteristic differences of speech among healthy individuals and patients with OSA were established [112]. A study analyzed wakefulness speech recordings from 93 participants using a Gaussian Mixture Models (GMM) framework which acoustic features including vocal tract length and linear prediction coefficients. The system achieved 79% sensitivity and 83% specificity in males and sensitivity of 84% and specificity of 86% in females [113]. Another study applied nonlinear analysis to various voice samples and utilized support vector machine (SVM) and KNN classifiers, found that consonants (96% accuracy for SVM) are more distinctive than vowels (83.5% accuracy for KNN) in OSA detection [22]. Another study also developed a multimodal fusion framework that integrates self-

supervised learning (SSL) speech representations with traditional acoustic features to estimate OSA severity from vowel phonations, obtained balanced accuracies between 74% to 80% for various AHI thresholds [114]. Despite the accessibility of wakefulness speech recording, several challenges remain. Many earlier studies involved relatively limited or demographically unbalanced cohorts, and speech features can depend on language, phonetic content, sex, and other speaker characteristics, potentially limiting generalization across populations [115].

2.4.4 Tracheal Breathing Sound Methods

Several studies have investigated the efficacy of tracheal breathing sounds (TBS) in OSA monitoring and detection. A study involving 1,852 patients by analyzing spectrograms generated from 60s segments to train CNNs for apnea/hypopnea event detection and sleep/wake classification. The trained framework achieved 97% sensitivity and 90% specificity [116]. Another study employed a CNN based on the EfficientNetV2B1 architecture to analyze Mel spectrograms derived from 10s tracheal audio segments [117]. This study attained an accuracy exceeding 90%, an F1-score of 91%, and an area under the curve of 0.96. Furthermore, research that integrated tracheal sounds with motion and oximetry data across 233 participants demonstrated a sensitivity of 90% and a specificity of 85.9% at an AHI threshold of 15 [118]. While these results highlight the potential of using tracheal sounds as a physiological source for OSA detection,

these studies relied on recordings obtained during overnight clinical sleep. However, these approaches primarily derive their diagnostic information from acoustic and physiological changes occurring during sleep and therefore require sleep-state recordings, even when full PSG may not ultimately be necessary. Consequently, they do not address whether diagnostically useful OSA-related upper-airway information can be obtained during wakefulness, which could enable shorter screening protocols without requiring the occurrence of overnight respiratory events.

2.4.5 Synthesis and Motivation

The current OSA screening methods have shown a trade-off physiological information content, sensor burden, recording duration, and practical accessibility. EEG and cardiovascular methods demonstrate high accuracy but are limited supervised, overnight clinical environments. These methods require multi-channel sensor configurations and specialized expertise, which limits their scalability for large populations. On the other hand, while speech analysis offers a more convenient alternative, its performance is often compromised by linguistic nuances, phonetic variability, and gender anatomical differences leading to inconsistent generalizability across diverse cohorts. TBSs are particularly relevant to OSA assessment because their acoustic characteristics are influenced by respiratory airflow and airway geometry, providing a non-invasive source of information related to upper airway structure and function. However, these approaches obtain diagnostic

information during sleep and generally rely on the occurrence of sleep-disordered breathing events. A complementary approach is therefore to investigate whether OSA-related physiological information can be detected during wakefulness, enabling shorter screening protocols that do not require observation of overnight obstructive events.

2.5 Wakefulness OSA Screening Using Tracheal Breathing Sounds

TBS has found to be a promising physiological source for identifying respiratory events and persistent anatomical and functional traits of the upper airway during wakefulness. This approach is founded on the acoustic airway hypothesis, which states that the sound of air moving through the trachea carries a unique signature of the airway's geometry, even while awake [27].

2.5.1 Rationale for Wakefulness-Based OSA Screening

TBS have been found to vary systematically with velopharyngeal narrowing, indicating a strong relationship between upper airway geometry and acoustic response [27]. Several studies have demonstrated the anatomical differences in the upper airway within OSA individuals, including reduced pharyngeal cross-sectional area [119], increased velopharyngeal compliance and collapsibility [120], and increased thickness of the lateral pharyngeal walls [121], compared with healthy individuals. These differences were found to produce alterations in TBS spectra,

with velopharyngeal and oropharyngeal constrictions affecting resonance peaks in different ways, which might be localizable via TBS analysis [122]. Therefore, recording TBS for only a few minutes during wakefulness presents a compelling alternative to resource-intensive modalities requiring supervised overnight sleep studies and specialized clinical expertise.

2.5.2 Data Acquisition and Protocol

In previous studies, short and structured breathing protocols were used to record respiratory sounds before or during PSG assessment. Most studies recorded participants in the supine position with the head resting on a pillow [23,25,28–33,123], while one study also included an upright sitting position to evaluate the effect of body posture on breathing sound characteristics [32]. The recordings were performed while subjects were awake, often shortly before the overnight PSG, assessment, allowing the extracted acoustic features to be compared with AHI derived from the subsequent PSG.

A major similarity across these studies is the use of a microphone placed over the trachea. Most studies used a Sony omnidirectional electret condenser microphone (ECM77B) embedded in a small chamber and positioned over the suprasternal notch, leaving approximately 2 mm spacing between the microphone and the skin [23,25,28–33,123]. This microphone has nominal sensitivity of -52 dB ± 2 dB and broad frequency response from 40 Hz to 20 kHz. Other studies used

different recording devices, including a unidirectional electret condenser microphone [124]. Despite difference in hardware, the main goal was consistent to capture acoustic changes associated with upper airway airflow during wakefulness.

The breathing protocols were also relatively consistent across studies. Participants were usually instructed to complete five deep breathing cycles through the nose with the mouth closed, followed by five deep breathing cycles through the mouth while wearing a nose clip [23,25,28–33,123]. Some protocols also included a brief breath-hold period, which was used to estimate the background noise and calculate the signal to noise ratio (SNR). Overall, prior data acquisition protocols used short and repeatable recordings intended to capture upper airway acoustic behavior during wakefulness without requiring an overnight acoustic recording.

2.5.3 Feature Extraction and representations

Feature extraction for TBS analysis typically involves raw acoustic signals transformation into multidimensional representations that capture the structural and dynamical changes of the upper airway. Based on current literature, these features include spectral, statistical, and nonlinear representations, while anthropometric information is often incorporated as a complementary source of information [26].

The power spectral density (PSD), which is often estimated via the Welch method, serves as a commonly used representation in most studies [25,28,29,31–

33,123]. Features such as spectral entropy, centroid, and energy distribution across specific frequency bands (e.g., 100-300 Hz, 1000 – 1700 Hz) are extracted to characterize OSA-related differences in TBS spectra [25,29,33]. Beyond power spectra, higher-order statistics (HOS), particularly bispectral analysis, are frequently employed to characterize nonlinear interactions, phase coupling and non-Gaussian properties of breathing sounds [25,29,33,123]. Moreover, Mel-Frequency Cepstral Coefficients (MFCC) [23] have been used to provide compact representation of the TBS spectral envelope, while autoregressive (AR) modeling [124] has been used to characterize spectral envelope and vocal tract resonance (formants), where local maxima provide proxies for airway geometry.

To characterize the complexity and temporal dynamics of TBS during wakefulness, several studies utilized nonlinear features. These include:

- Fractal dimension: Katz and Higuchi fractal dimensions are used to measure the complexity and self-similarity of the tracheal sound signal [25,29,32,33,123].
- Hurst Exponent: This characterizes long-range dependence or persistence in the time series and provides information about the temporal organization of the TBS signal [25,29,33,123].
- Entropy and Higher-Order Moments: Spectral entropy is used to characterize the distribution and complexity of spectral energy, while

skewness and kurtosis describe characteristics of the acoustic-signal distribution [25,29,30,33,123].

An important approach in wakefulness screening is the fusion of acoustic features with clinical and anthropometric data [25,29,30]. Because TBS characteristics can be influenced by anatomical and anthropometric factors, many models incorporate age, sex, BMI, neck circumference, and MpS.

Given the high dimensionality of these feature sets, feature reduction is a critical step. Common methodologies include One-Way Analysis of Variance (ANOVA) [30,32] or unpaired t-tests [25,28,29,31,33,123] to exclude non-significant features. For example, the AWakeOSA framework [25] further evaluated feature robustness and redundancy after significance testing, reducing approximately 250 initial acoustic features to around 15 candidate features within each anthropometric subset. More advanced regularized approaches, such as Least Absolute Shrinkage and Selection Operator (LASSO) [30,31] and model-based feature ranking using Random Forest (RF) [30] rankings, which can reduce feature dimensionality and model complexity.

2.5.4 Machine Learning Pipelines

The machine learning models used for awake OSA screening have evolved from traditional statistical models to more complex ensemble and deep learning

architectures. Foundational studies frequently utilize Linear Discriminative Analysis (LDA) [23,32,124], Quadratic Discriminative Analysis (QDA) [23,32], and Logistic Regression [23,31]. These relatively simple classifiers can provide interpretable decision rules and have also been evaluated in combination with anthropometric features [23,32]. To capture the non-linear dynamics of respiratory acoustics, researchers employ Support Vector Machine (SVM) with linear [23,28,33,123] and RBF kernels [23], Random Forests [23,25,29], and Classification Trees [30]. Other ensemble techniques, including Gradient Boosting Machines (GBM) and RUSBoost have also been evaluated for OSA classification [23].

Rigorous validation is critical to ensure that these diagnostic tools generalize across diverse populations. Methods such as stratified K-fold [33,123] and leave-one-out (or leave-two-out) [28,30] have been used to maximize the utility of smaller clinical datasets. Some studies also reserved a dedicated blind test set to evaluate model performance on subjects not used during model development [25,29–31]. Some studies also balanced or stratified subject groups with respect to OSA severity and relevant demographic or anthropometric factors, such as age and BMI, to reduce the likelihood that classification performance was driven primarily by population differences rather than OSA-related characteristics [23,33,123].

2.5.5 Prior Work on Wakefulness TBS for OSA Detection

The implementation of awake OSA screening using TBS has yielded a range of diagnostic outcomes, establishing it as a viable physiological source for non-invasive detection. Early studies mainly focused on classifying OSA severity groups using traditional statistical classifiers. For example, LDA and QDA classifiers achieved more than 91% accuracy (92% sensitivity and 85% specificity) for separating severe OSA from non-OSA subjects, and more than 83% accuracy (95% sensitivity and 81% specificity) for classifying non-OSA/mild OSA from moderate/severe OSA [32]. Another study using formant features estimated by linear predictive coding, combined with anthropometric information, reported up to 86.4% accuracy, 88.9% sensitivity, and 84.6% specificity for mild/moderate versus severe OSA groups classification [124]. These findings supported the feasibility of using respiratory acoustics recorded during wakefulness to capture OSA-related changes in the upper airway.

Subsequent studies shifted toward binary OSA screening using TBS-derived features either independently or in combination with anthropometric information. Using a linear SVM to distinguish subjects with AHI ≥ 10 from those with AHI ≤ 5, a study reported 83.9% testing accuracy, 82.6% sensitivity, and 85.2% specificity [28]. Another study incorporated anthropometric subgrouping and a voting procedure across subgroup-specific classifier, achieving 83.6% accuracy, 74.5%

sensitivity, and 90.1% specificity [33]. Regularized logistic regression achieved 79.3% accuracy, 82.2% sensitivity, and 75.8% specificity [31], while a subsequent Random Forest achieved 82.1% accuracy, 84.2% sensitivity, and 79.5% specificity [30]. More recent work extended wakefulness TBS analysis to severe classification using MFCC features and multiple machine learning models, with an SVM achieving 88.2% accuracy, 83.3% sensitivity, and 90.9% specificity for non-OSA versus severe OSA classification [23]. These studies suggest that anthropometric information can provide complementary information to TBS features and may improve the classification performance is some wakefulness OSA screening frameworks.

The work most closely related to the present thesis proposed the AWakeOSA decision-making procedure, which used anthropometric subsets, subgroup-specific spectral and bispectral TBS features, Random Forest classifiers, and a weighted voting schema to obtain the final OSA classification [25]. The training population was divided into available subgroups based on age, sex, BMI, neck circumference, and MpS, and separate OSA-sensitive sound features and classifiers were developed within these subgroups [25]. TBS were recorded during wakefulness using a Sony ECM77B microphone placed over the suprasternal notch [25]. In that study, 199 participants were divided into OSA (AHI < 15; n = 109) and non-OSA (AHI ≥ 15; n = 90) [25], and the final model achieved 81.4% accuracy, 80.9% sensitivity, and

82.1% specificity on a blind test set [25]. This study provides the primary methodological foundation for the present thesis, particularly its use of anthropometric subgroup-specific classifiers and fusion of their decisions for wakefulness OSA screening. Table 1 summarises the mentioned studies on wakefulness OSA screening using TBS.

Table 1. Summary of previous studies on wakefulness OSA screening using tracheal breathing sounds.

Ref.	Study aim	features	Model	Main Result
[25]	Decision making framework using anthropometric subset	Spectral, bispectral TBS features	Random Forest	81.4% accuracy, 80.9% sensitivity, 82.1% specificity
[29]	Prediction of PSG parameters	Spectral, bispectral TBS features	Random forest	Demonstrated PSG-parameter prediction
[32]	Severity Assessment across non-OSA, mild, moderate, and severe OSA groups	PSD, fractal dimensions, kurtosis	LDA, QDA	>91% accuracy for severe OSA vs. non-OSA
[30]	Comparison of LR and RF-based models	PSD, Anthropometric features	LR, RF, classification trees	82.1% accuracy, 84.2% sensitivity, 79.5% specificity
[123]	Identify robust OSA-related TBS features	Spectral, bispectral, nonlinear features	Linear SVM	Identified features with lower sensitivity to selected factors
[33]	Binary OSA screening using anthropometric subgroups	spectral, bispectral features with anthropometrics	Linear SVM with voting	83.6% accuracy, 74.5% sensitivity, 90.1% specificity
[31]	Screen OSA using daytime TBS and anthropometrics	PSD, anthropometric features	Lasso logistic regression	79.3% accuracy, 82.2% sensitivity, 75.8% specificity
[23]	Classify OSA severity using MFCC-based TBS features	MFCC-based sub-frame features	Multiple ML models	88.2% accuracy, 83.3% sensitivity, 90.9% specificity
[28]	Screen OSA and airway structure using wakeful TBS	Average spectral features	Linear SVM	83.9% accuracy, 82.6% sensitivity, 85.2% specificity
[124]	Classify OSA severity using TBS analysis	AR spectral-envelope, formant features	LDA	86.4% accuracy, 88.9% sensitivity, 84.6% specificity

2.6 Summary of Key Limitations in Current Literature

Although the mentioned studies have demonstrated the feasibility of wakefulness OSA screening using TBS, several methodological and practical limitations remain. These limitations are mainly related to the limited evaluation of previously developed frameworks across different recording hardware and the incomplete representation of heterogeneous OSA populations.

2.6.1 Limited Evaluation Across Different Recording Hardware

A key limitation of previous wakefulness TBS studies is most related frameworks were developed and evaluated using a specific recording setup. In particular, several studies used the Sony ECM77B microphone embedded in a small chamber placed over the suprasternal notch of the trachea, usually in a supine position and under structured breathing protocols [23,25,28–33,123]. Although the use of a consistent acquisition system facilitates comparison of acoustic characteristics within a study, it provides limited evidence regarding whether the extracted TBS features and developed classification frameworks remain transferable when a different microphone is used.

Different microphones may have different frequency-response characteristics, which can alter the recorded representation of TBS. In addition, recordings obtained over the suprasternal notch may contain environmental sounds and acoustic

components generated by oral and nasal breathing in addition to the target TBS. These components may exist regardless of microphone cost; however, different microphones may represent them differently because of their electroacoustic characteristics. Therefore, it remains unclear whether diagnostically relevant TBS information and frameworks developed using the Sony ECM77B can be directly transferred to recordings obtained using a different microphone.

The relatively high cost of the Sony ECM77B may also limit the accessibility and scalability of wakefulness OSA screening, motivating the investigation of lower-cost alternatives. However, microphone cost itself should not be considered a direct indicator of signal quality or environmental-noise sensitivity.

2.6.2 Limited Representation of Population Heterogeneity

A major limitation of previous studies is the limited ability of the datasets to represent the heterogeneity of the OSA population. OSA is associated with several anatomical, demographic, anthropometric, physiological, and lifestyle-related factors. However, several studies reported small sample size or insufficient numbers of subjects in important subgroups, such as females, individuals with lower neck circumference, higher BMI, or higher MpS [25,28,29,31,33,123]. This limitation can reduce the reliability of subgroup analysis and may cause models to learn patterns that are more representative of dominant groups in the dataset.

Previous studies attempted to address this variability by developing subgroup-specific classifiers based on anthropometric characteristics [25,33]. However, limited numbers of subjects in some groups prevented the development of classifiers for all desired subgroups. For example, the AWakeOSA framework [25] could not construct several complementary anthropometric subgroups because of insufficient numbers of participants within those groups.

Therefore, further investigation using a more heterogeneous study population is required to determine whether subgroup-specific TBS characteristics can be identified reliably and incorporated into a screening framework that provides robust performance across different population groups.

2.7 Summary

This chapter provided an overview of anatomical and physiological mechanisms underlying OSA, followed by a discussion of its major risk factors. Current clinical and practical OSA diagnostic methods, particularly PSG, were reviewed, and their limitations and implementation challenges were discussed. Questionnaire-based screening tools were then introduced as the initial screening methods, followed by a review of physiological modalities that have been used in machine learning OSA assessment. Wakefulness OSA screening via TBS was discussed in detail, including previous recording protocols, feature extraction and

selection methodologies, and reported results. Finally, the key limitations in prior studies on wakefulness OSA detection using TBS were summarized, particularly the limited evaluation across different recording hardware and the limited representation of heterogeneous OSA populations.

Chapter 3: Methodology

In this chapter, the methodology employed for developing the proposed wakefulness OSA classification framework is outlined. The first part of the chapter describes the dataset, study population, recording protocol, and anthropometric variables used in the proposed screening framework. The second part presents the methodology used to develop the proposed machine learning framework. This includes data preparation and quality control, signal preprocessing, acoustic feature extraction, sub-classifier definition, and training strategy. Overall, this chapter provides an in-depth explanation of the steps used to develop and evaluate the proposed wakefulness OSA screening framework using TBS and anthropometric information.

3.1 Dataset and Recording Protocol

Participants were recruited from individuals suspected of OSA who visited the Medical Medigas Center, Winnipeg, MB, Canada. Sleep apnea was assessed with either a portable HSAT at their home or with full PSG at Misericordia Sleep Disorder Center (MSDC); all verified by a respirologist. Individuals were excluded if they had any chronic respiratory disease, insomnia, or drug addiction. A total of 247 individuals aged 18 to 70 years old signed an informed consent and were enrolled in the study. The study was approved by the Biomedical Research Ethics Board of the

University of Manitoba. After obtaining the consent, each participant completed a questionnaire that recorded anthropometric information, including height, weight, age, sex, MpS, neck circumference, smoking history, high blood pressure, previously observed apnea, daytime fatigue, snoring, receding mandible, body fat distribution, and double chin.

TBS were recorded at the Remologie Sleep Clinic (at the Medigas Center), Winnipeg, Canada, during wakefulness in supine position (with a pillow placed under the head) using a low-cost microphone (EM273) at a sampling frequency of 44.1 kHz. The microphone was inserted into a custom 2-mm plastic chamber, placed over the suprasternal notch on the trachea. The participants were instructed to perform two deep breathing maneuvers. First, the nasal-breathing maneuver consisted of five breathing cycles performed through the nose, followed by an oral-breathing maneuver consisting of five deep breathing cycles performed through the mouth while wearing a nose clip. At the start of each maneuver, participants were instructed to hold their breath for a few seconds to record the background noise. A research assistant helped the participants with up/down hand movements for the breathing cycles.

3.2 Data Preparation and Quality Control

To prepare the dataset for our wakefulness OSA diagnostic framework, recordings with significant background noise (e.g., construction noise, speech) were excluded. The remaining recordings were segmented into inspiration and expiration phases by analyzing their logarithmic variance to detect the transitions between phases. The SNR for each phase was quantified using the breath-hold period of the corresponding signal. The raw dataset exhibited a mean SNR of 9.7 ± 5.6 dB for mouth breathings and 10.5 ± 5.7 dB for nose breathings, reflecting the challenging acoustic environment. Phases with SNR < 2 (approximately 3 dB) were excluded, particularly the inspiration and expiration phases adjacent to the breath-hold, to mitigate potential distortion of the acoustic characteristics.

Subjects were excluded if their data contained fewer than two phases per maneuver. In addition, it is important to note that unlike laboratory settings where noise is typically low-level and stationary, this dataset contained high-amplitude non-stationary artifacts (e.g., ambient speech, hammer sounds due to construction nearby). To ensure data quality, the mean power spectrum of the maneuvers was inspected for harmonic noise, characterized by prominent narrow-band peaks recurring at approximately regular frequency intervals, for each subject. After excluding 14 individuals due to low SNR or insufficient breathing phases and 5 individuals due to persistent harmonic noise, 228 subjects remained in the dataset.

The final dataset yielded a mean SNR of 10.7 ± 4.2 dB for mouth breathing and 11.8 ± 4.8 dB for nose breathing. To facilitate binary classification and account for the potential overlap between the two groups (OSA and non-OSA), subjects with AHI values between 10 and 15 were excluded. This exclusion removed 32 subjects, yielding a final sample of 196 participants, which each sample is defined as a unique subject with at least 2 phases per maneuver (e.g., mouth inspiration). The participant exclusion and dataset refinement steps are summarized in Figure 2. Subjects with an AHI < 10 were labeled as non-OSA (n = 47), and those with AHI > 15 were labeled as OSA (n = 149). Table 2 summarizes the demographic and clinical information of the final cohort included in this study.

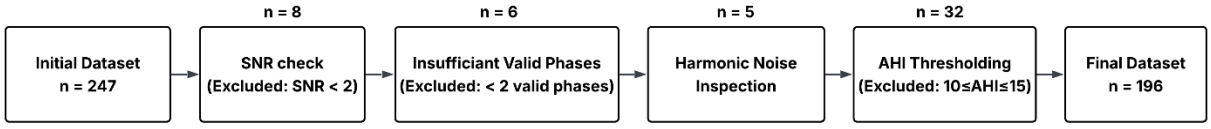

Figure 2. Dataset refinement process for the final OSA classification cohort.

Table 2. Demographic and clinical information of the study cohort. Data for age, BMI, and AHI are presented as mean ± standard deviation or n count (number of subjects). All other variables are reported as n count. BMI is the body mass index calculated as kg/m^2 and MpS is the Mallampati score; the numbers show the number of participants in each group of MpS level. Similarly, the number in smoking history shows the number of participants of non-smoker, former, and current smokers in each group.

Class	Total	Age (year)	Sex (M/F)	BMI	MpS (I, II, III, IV)	Smoking History (never, former, current)	AHI
OSA	149	48.1±12.0	(99, 50)	35.9±8.9	(32, 26, 38, 53)	(85, 37, 28)	42.6±26.7
Non-OSA	47	44.1±13.0	(19, 28)	31.9±6.8	(10, 11, 17, 9)	(34, 6, 7)	5.7±2.6

3.3 Data Preprocessing

The segmented TBS phases were passed through a 4th order Butterworth band-pass filter (75 – 5000 Hz) prior to down-sampling. Since above 5000 Hz, the power spectra of both OSA and non-OSA signals were highly overlapped and approached the noise floor, a down-sampling by a factor of 4 was applied to reduce the computational cost. Stationary portions of the signals were extracted using the method described in [39]. Briefly, the logarithmic variance was computed in 30 ms windows with 50% overlap to obtain the signal's envelope. The stationary portion was defined as the 50% of the signal duration centered on the maximum peak of the envelope. A 6th order Butterworth band-pass filter (75 – 3500 Hz) was subsequently applied to the stationary portion to restrict the subsequent analysis to the selected processing bandwidth and further attenuate residual high frequency components after down-sampling. Finally, 8-sample moving average and standard deviation normalizations were applied to minimize intra-segmental and inter-segmental variations, respectively. The preprocessing pipeline is shown in Figure 3.

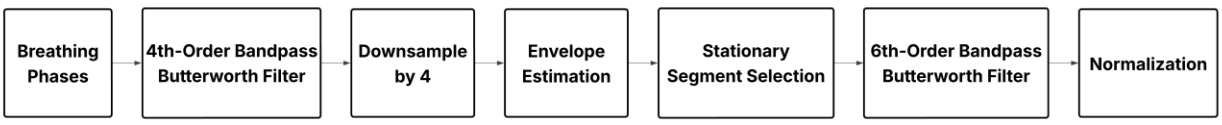

Figure 3. Preprocessing pipeline applied to breathing phases.

3.4 Acoustic Feature Extraction

Although deep learning approaches can automatically learn acoustic representations from raw signals, they generally require a larger number of training subjects; however, the number of subjects was limited in this study, particularly for the non-OSA class. Data augmentation and synthetic data generation were also not employed because the limited and highly variable clinical dataset may not provide a sufficient representative distribution from which to generate reliable physiologically meaningful samples. Therefore, handcrafted feature extraction followed by feature selection and conventional machine learning approaches were deployed in this study to reduce model complexity and improve physiological interpretability. The mean power spectra of the OSA and non-OSA classes overlapped substantially across most of the analyzed frequency range (Figure 4), indicating limited discriminatory power spectral analysis alone to capture the subtle acoustic features of OSA. The high degree of spectral overlap suggests that the discriminative information is not encoded in the average signal amplitude and motivates further characterization of localized sub-band signal properties. To overcome this limitation, Wavelet packet decomposition (WPD) was implemented to systematically decompose the signal into narrow frequency sub-bands and enable feature extraction from each sub-band, thereby reducing risk of overlooking localized discriminative information that may not be visually apparent in the power spectrum (Figure 5(a)). Several mother

wavelets were explored during preliminary analysis, and Symlet 8 (sym8) was retained due to its high symmetry, compact support, and a good balance between time and frequency resolution (Figure 5(b)). The corresponding WPD tree structure is shown in Figure 5.

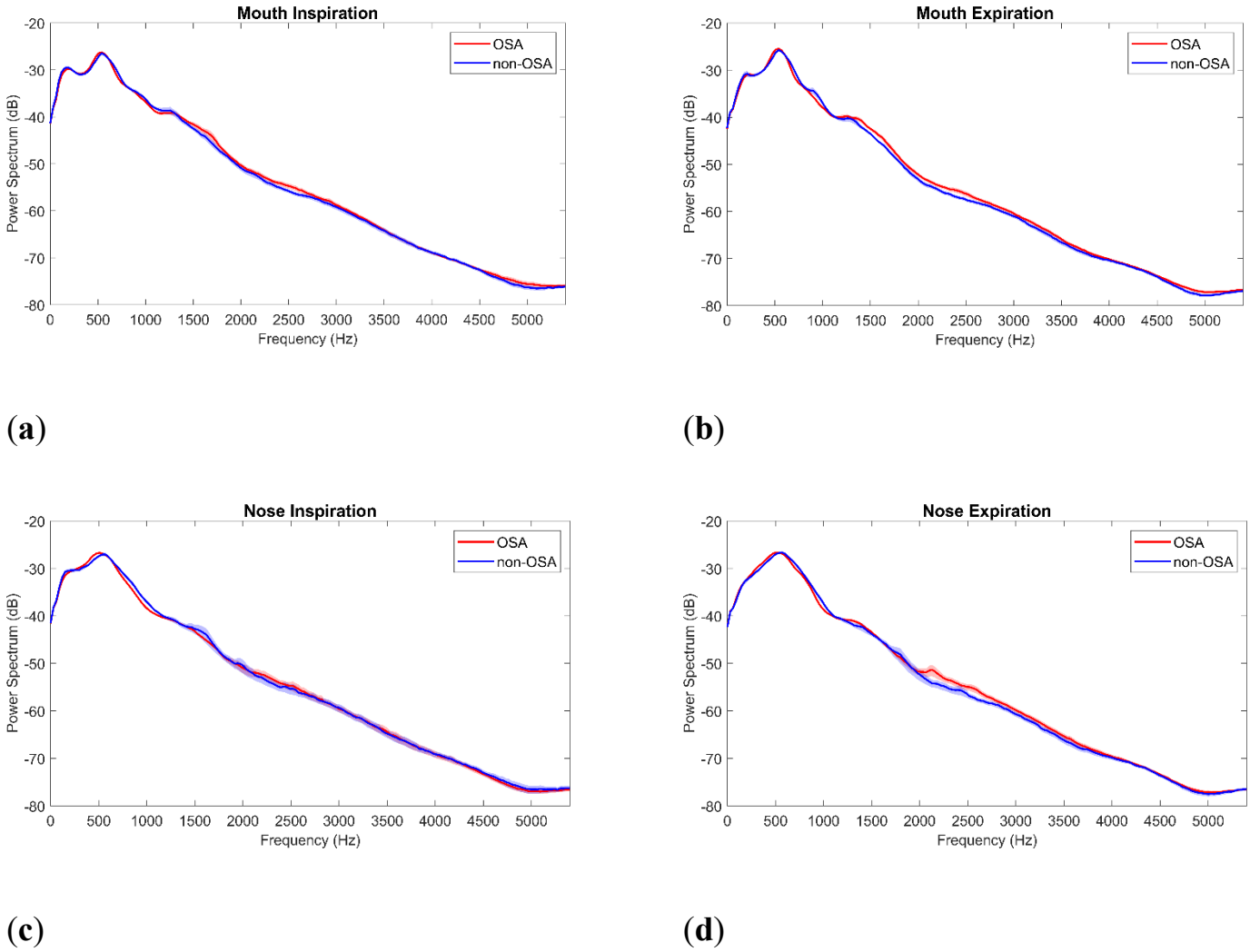

Figure 4. Mean power spectra (dB) of tracheal breathing sounds for OSA (red) and non-OSA (blue) subjects during (a) mouth inspiration, (b) mouth expiration, (c) nose inspiration, and (d) nose expiration maneuvers. Solid lines show the group mean and shaded bands indicate ±1 standard error of the mean (SEM).

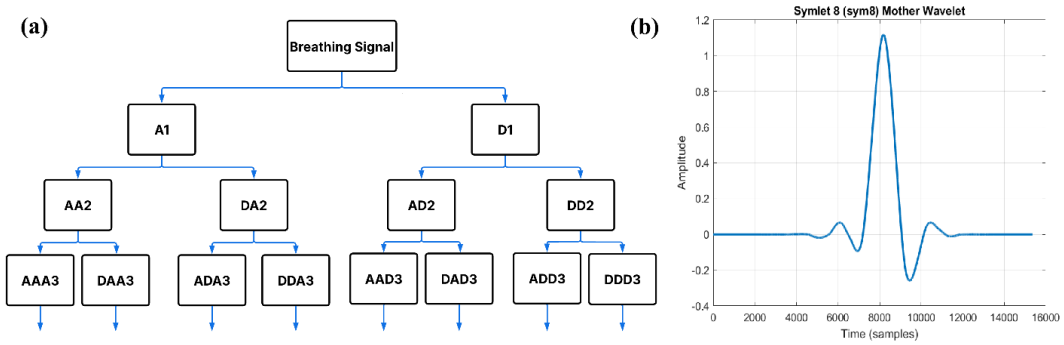

Figure 5. Wavelet packet decomposition tree (a) and Symlet 8 mother wavelet (b).

A five-level WPD was selected to provide sufficiently narrow frequency sub-bands for localized feature extraction in order to split the frequency range into 32 equally spaced sub-bands, each spanning a bandwidth of 172 Hz. Although both classes were highly overlapped in the frequency domain, the first 10 sub-bands (0 – 1,722 Hz) were selected as the regions of interest for feature extraction to limit the dimensionality of the candidate feature space while focusing on the frequency region containing the dominant energy. Although localized gaps were visible at some higher frequencies, these differences were relatively small and not consistently observed across all breathing maneuvers. Therefore, the selection of the first 10 sub-bands was used as a dimensionality-reduction choice rather than as an indication that no discriminatory information existed above the cutoff frequency. 16 distinct features were extracted directly from the wavelet coefficients of the 10 selected frequency sub-bands of each breathing cycle (e.g., a single mouth inspiration phase), as listed in Table 3. These features were intentionally limited to interpretable statistical, temporal, spectral, higher-order distributional, and entropy-based features to control

dimensionality relative to the available number of subjects; therefore, fractal, bispectral, and non-linear features were not included in this study. These features were motivated by the underlying physics of respiratory airflow and airway anatomy. Statistical moments (e.g., Mean, Variance, Energy) were calculated to quantify the overall signal power, as the intensity of tracheal sounds is known to correlate directly with airflow rate and airway resistance [125]. Spectral features (e.g., spectral centroid, spread) were extracted to capture variations in airway resonance [126]. Finally, information-theoretic features (e.g., Shannon and Tsallis entropies) were utilized to characterize signal complexity and distributional uncertainty of the sub-bands, providing information not directly represented by conventional statistical moments. For each breathing maneuver (e.g., mouth inspiration), this resulted in 160 features (16 feature types × 10 sub-bands). These features were then averaged across all available breathing cycles within a specific maneuver. Finally, the feature vectors from the four maneuvers (mouth inspiration, mouth expiration, nose inspiration, and nose expiration) were concatenated to yield a vector of 640 features per subject (160 features × 4 maneuvers). All features were standardized using z-score normalization across all subjects.

Table 3. Name of the features extracted from each frequency sub-band.

Category	**Feature List**
Statistical	1. Mean, 2. Variance, 3. Std Deviation, 4. Energy, 5. Median
Higher-Order	6. Skewness, 7. Kurtosis
Temporal	8. Zero-Crossing Rate

Spectral	9. Centroid, 10. Spread, 11. Skewness, 12. Kurtosis, 13. Entropy
Complexity	14. Shannon Entropy, 15. Tsallis entropy, 16. Rényi Entropy

3.5 Sub-Classifier Definition

To address the population heterogeneity in OSA presentation, subjects were stratified into nine subgroups, as shown in Figure 6. These subgroups included sex (male, female), age (age > 50, age ≤ 50), BMI (BMI > 35, BMI ≤ 35), MpS (MpS > 2, MpS ≤ 2), smoking history (non-current smokers or former-smokers). The sex, age, BMI, and MpS subgroups were selected based on their established relevance to OSA and their previous use in subgroup-based frameworks [25,33], while the age and BMI thresholds were consistent with the STOP-Bang criteria [98]. Smoking history was considered as one subgrouping criteria that the detailed supporting analysis is provided in Appendix A.

Neck circumference was retained as a continuous anthropometric feature rather than used as an additional subgrouping criterion to preserve its continuous predictive information. The choice was supported by the previous Awake OSA framework [25], in which neck circumference was significantly correlated with AHI ($r = 0.43$, $p < 0.01$) and was subsequently selected as a classification feature within the age ≤ 50 and male subgroup. The distribution of the subjects within each subgroup is presented in Table 4.

Table 4. Distribution of Study Subjects Stratified by Clinical and Demographic Subgroups

Subgroup	OSA	Non-OSA	Ratio*

Male	99	19	5.21
Female	50	28	1.79
High Age	64	18	3.56
Low Age	85	29	2.93
High BMI	66	11	6.00
Low BMI	83	36	2.31
High MpS	91	26	3.50
Low MpS	58	21	2.76
Non-Current Smokers	121	40	3.03

* Ratio is defined as number of OSA subjects over non-OSA subjects.

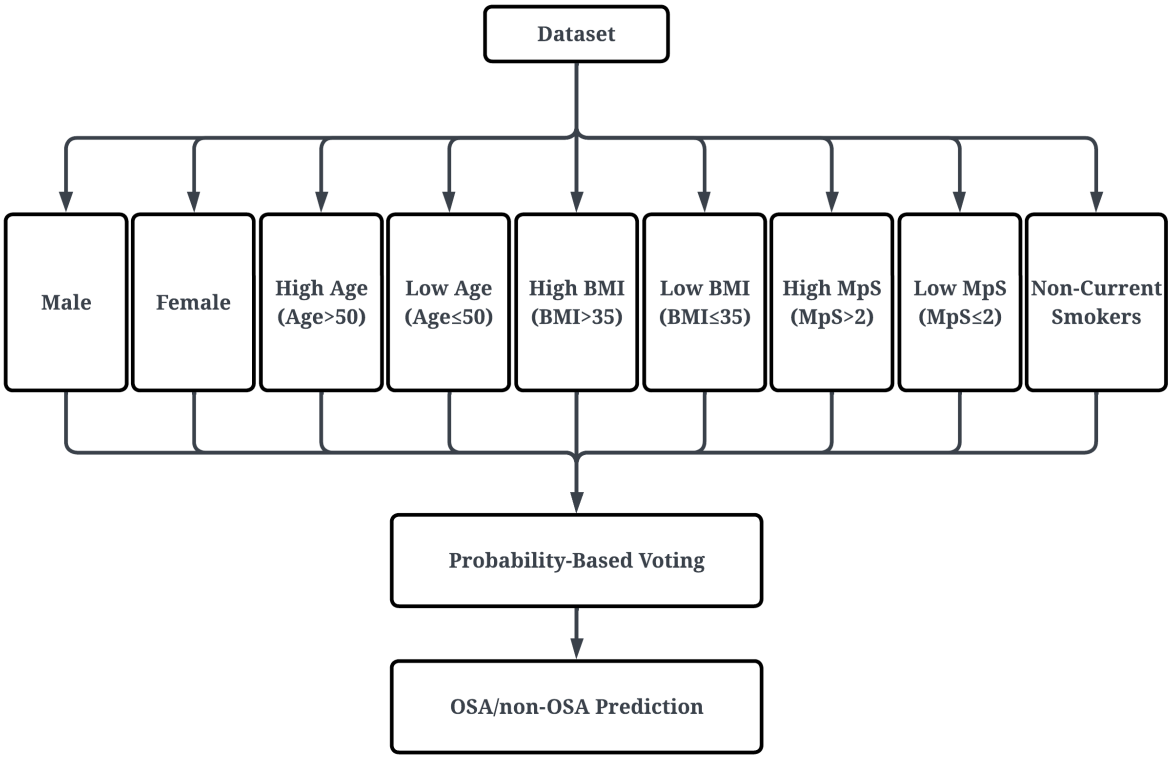

Figure 6. Ensemble voting across nine anthropometric sub-classifiers for final OSA vs. non-OSA prediction.

3.6 Training Strategy

A stratified 4-fold cross-validation was used to split the population into training and blind test sets for final model evaluation. For each subgroup, a dedicated

sub-classifier was defined and trained exclusively on subjects from that subgroup within the training folds. Within each sub-classifier, an additional repetitive stratified k-fold cross-validation was applied to partition the training and validation sets; the validation sets were used to evaluate the feature selection and to tune hyperparameters. The value of k ranged from 3 to 6, depending on the number of non-OSA subjects available within the subgroup.

To address the high dimensionality of the feature space and ensure reproducibility, a hierarchical feature stability framework was implemented. While OSA and non-OSA subjects were highly imbalanced across most subgroups, feature selection was performed based on repeatedly balanced subsets. To establish intra-fold stability for each subgroup, we generated $B = 1{,}000$ balanced subsets by combining all number of non-OSA subjects (minority class) with an equal number of randomly sampled OSA subjects (majority class). On each balanced subset, the minimum redundancy maximum relevance (mRMR) algorithm was utilized to select the top 7 informative acoustic feature set S [127]. The optimization objective of mRMR, which is to select features that maximize the mutual information with the target class y (relevance) while minimizing the mutual information shared with already selected features (redundancy), is defined as:

$$\max_{S \subset \Omega} \left[\frac{1}{|S|} \sum_{f_i \in S} I(f_i; y) - \frac{1}{|S|^2} \sum_{f_i, f_j \in S} I(f_i; f_j) \right] \quad (1)$$

Where Ω denotes as the pool of 640 acoustic features, $|S|$ is the size of the selected subset, and $I(.;.)$ represents mutual information. For each iteration, only the top 7 features were retained. Features with an mRMR importance score (ranged from 0 to 1) below 0.01 were discarded from the stability analysis. The value of 0.01 was used as a practical near-zero importance threshold in this study rather than as a standardized mRMR cutoff. Following the 1,000 iterations, a stability score was calculated for each feature based on its selection frequency across all subsets. The seven most stable acoustic features were combined with three predefined anthropometric features—neck circumference, snoring, and receding mandible—to form the final 10-feature set for each sub-classifier. Seven acoustic features were selected as a practical compromise between retaining complementary acoustic information and limiting model dimensionality relative to subgroup sample sizes. Neck circumference was retained as a continuous predictor because of its established association with OSA and its previous contribution to the AWakeOSA framework [25]. Snoring is an established OSA screening indicator included in STOP-Bang [98], while receding mandible represents a craniofacial risk factor for OSA [128]. These anthropometric features were selected a priori based on clinical relevance rather than through data-driven optimization.

The repeated sub-sampling strategy employed for feature selection was replicated for model training to reduce bias toward the majority class and enhance robustness across subject combinations. Each sub-classifier consisted of an ensemble of multiple bagged decision trees, with each tree trained on a distinct balanced subset. All minority-class subjects were included in every subset, while an equivalent number of majority-class subjects were drawn without replacement from the remaining pool until all majority-class subjects had been included at least once within the coverage cycle. Therefore, the minimum number of balanced subsets required for one complete coverage cycle was defined as $N_{min} = \lceil Ratio_{OSA/non-OSA} \rceil$. The coverage procedure was repeated three times as a practical design choice, resulting in $3 \times N_{min}$ trees and ensuring that each majority-class subject contributed to at least three balanced training subsets.

Once the number of ensemble trees was determined and the features were selected, the training phase commenced. Each decision tree within the ensemble was trained on its own dedicated balanced subset using the selected feature vector. For each sub-classifier, predictions of the models on the validation sets were then aggregated across trees to evaluate the selected features and tune hyperparameters. For the voting, average predicted-probability method was implemented. Hard voting was not adopted because it uses only binary class decisions and therefore discards information about the confidence of individual model predictions. Weighted voting

was also not deployed as the performance of individual model performance was sensitive to combination of subjects of the training set and potentially reducing the stability and generalizability of the voting. An average predicted-probability voting was implemented within each sub-classifier (C_k), where k denoted a specific anthropometric sub-classifier to obtain a single prediction per subject. The average predicted probability of a subject x belonging to the OSA class, denoted as $\hat{P}_k(y = 1 \mid x)$, was computed across all bagged trees (M_k) as follows:

$$\hat{P}_k(y = 1 \mid x) = \frac{1}{M_k} \sum_{m=1}^{M_k} P_m(y = 1 \mid x) \quad (2)$$

To construct the final predictive sub-classifier utilizing the entire training set (both training and validation sets), a consensus ranking strategy was applied to the stable features identified across the inner cross-validation folds. For each cross-validation partition, a robust candidate set of 7 features was identified via the internal stability selection process described above. The candidate sets from all k inner folds (with their internal repetitions) were then pooled to form a global collection of high-stability markers. We calculated the recurrence frequency of each feature across these candidate sets. The top 7 features with the highest recurrence in this global pool were selected as the final acoustic features for that sub-classifier. These robust features were then combined with three anthropometric features to form the final 10-dimensional feature vector, as it has been determined that the inclusion of

anthropometric features, in conjunction with acoustic features, yields a superior performance compared to acoustic features alone [28]. This feature combination also enabled us to compare the current study output with the previous proposed framework [25]. Subsequently, training and validation sets were concatenated to train the final sub-classifier using the same sub-sampling and ensemble strategy with this final feature vector.

To obtain the final prediction, a dynamic ensemble selection was applied. The probabilities from all sub-classifiers relevant to the subject's anthropometric profile (e.g., male, high age, high BMI) were averaged to produce the final probability $\hat{P}_{final}(x)$:

$$\hat{P}_{final}(y = 1 \mid x) = \frac{1}{|K_x|} \sum_{k \in K_x} \hat{P}_k(y = 1|x) \tag{3}$$

Where K_x is the set of indices for sub-classifiers relevant to subject x. Subjects with $\hat{P}_{final}(x) < 0.5$ were labeled as non-OSA, whereas subjects with $\hat{P}_{final}(x) \geq 0.5$ were labeled as OSA. Figure 7 provides a schematic overview of our entire algorithm.

Accuracy, sensitivity, specificity, and area under the curve (AUC) for each fold were computed to comprehensively assess the algorithm performance. The AUC was calculated from the continuous predicted probabilities for the test subjects in each fold. The mean and standard deviation across folds were also reported to

summarize variability. Balanced accuracy, defined as the average of sensitivity and specificity, was additionally reported because the dataset was imbalanced (149 OSA vs. 47 non-OSA). Unlike conventional accuracy, balanced accuracy gives equal importance to the performance of the OSA and non-OSA classes, thereby reducing the influence of the majority class on the overall performance measure. An aggregated confusion matrix, obtained by pooling predictions across all folds, was used to detail the counts of false positives and false negatives, which offered insight into error patterns. All preprocessing routines, feature extraction procedures, and classification algorithms were developed and executed in MATLAB R2023b (MathWorks, Natick, MA).

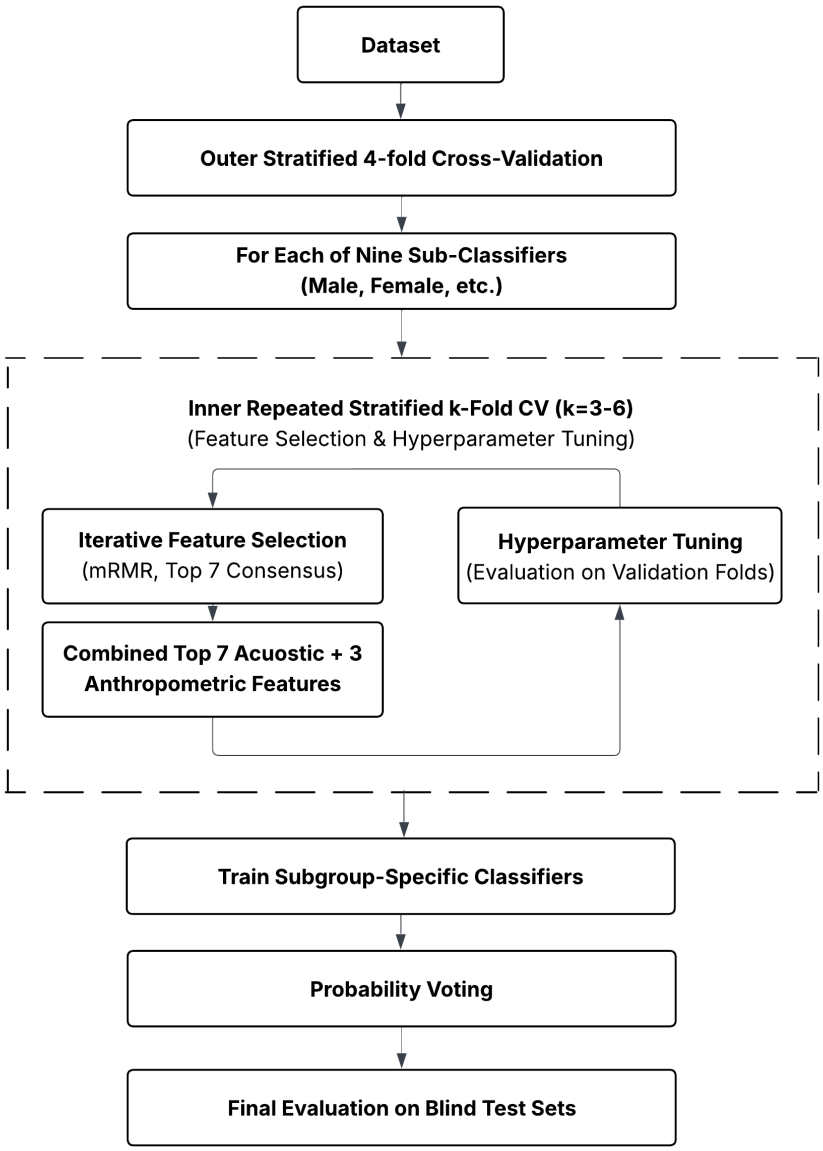

Figure 7. Schematic overview of the proposed Two-Level Ensemble Framework.

3.7 Summary

This chapter described the methodology used to develop and evaluate the wakefulness TBS-based OSA classification framework. First, the dataset, study population, anthropometric information, and tracheal breathing sound recording

protocol were introduced. The chapter then described the data preparation and quality control steps used to obtain reliable tracheal breathing sound recordings for classification. This included subject selection, signal quality assessment, breathing phase preparation, and preprocessing procedures. Acoustic features were then extracted from the prepared signals. Finally, the classification framework was described. The chapter explained the definition of sub-classifiers based on anthropometric and clinical characteristics, followed by the training strategy used to address class imbalance and population heterogeneity. These methodological steps provide the foundation for the classification results presented in the next chapter.

Chapter 4: Results

This chapter presents the results of the proposed wakefulness OSA detection framework. First, the selected acoustic features and anthropometric features used in the subgrouping framework are presented. The framework is then assessed through classification performance, voting strategy comparison, sub-classifier ablation analysis, baseline comparison, and misclassification analysis.

4.1 Predictive Feature Selection Results

From the initial 640 acoustic features extracted from four breathing maneuvers (nose and mouth inspiration/expiration), the seven most frequently selected acoustic features and the three anthropometric parameters (neck circumference, snoring, and receding mandible) were used to train the sub-classifiers. Table 5 summarizes the final stable acoustic features selected across all sub-classifiers, defined as those selected in at least three out of four folds.

Table 5. Summary of consistently selected acoustic features across outer cross-validation folds, appeared in at least three out of four folds for each anthropometric sub-classifier. Abbreviations: MI: Mouth Inspiration; ME: Mouth Expiration; NI: Nose Inspiration; NE: Nose Expiration.

Feature No.	Sub-classifier	Descriptor	Maneuver	Frequency Range (Hz)
9	Male	Spectral Centroid	MI	0 - 172
473	Male	Spectral Centroid	NI	1,550 - 1,722
276	Male	Kurtosis	ME	1,206 - 1,378
412	Female	Spectral Entropy	NI	861 - 1,033
410	Female – Low BMI	Spectral Kurtosis	NI	861 - 1,033
616	High Age	Median	NE	1,378 - 1,550

373	High Age	Standard Deviation	NI	517 - 689
522	Low Age	Spectral Kurtosis	NE	344 - 517
362	High BMI	Spectral Kurtosis	NI	344 - 517
113	Low BMI	Mean	MI	1,206 - 1,378
628	High MpS	Kurtosis	NE	1,550 - 1,723
599	Low MpS – Non-Current Smokers	Zero-Crossing Rate	NE	1,206 - 1,378

To evaluate the specific contribution of acoustic information relative to the three anthropometric input features, we analyzed the stability of the acoustic features within each subgroup. Rather than exhibiting random variation, which would be expected if the classification were driven primarily by the anthropometric inputs, distinct and high-stability acoustic features were observed across different subgroups. This variation implies that the selected acoustic features provide necessary discriminative resolution that complements, rather than duplicates, the baseline risk information captured by the anthropometric inputs. For example, in the male sub-classifier, f9 (spectral centroid in 0 – 172 Hz sub-band during mouth inspiration) was selected in all folds. The female sub-classifier frequently included f410 (spectral kurtosis) and f412 (spectral entropy) from the 861 - 1,034 Hz sub-band during nose inspiration. This divergence suggests that while the anthropometric features (e.g., neck circumference) provide a baseline risk profile, the model explicitly requires distinct spectral patterns to accurately classify the subject. Similarly, the high age and low age sub-classifiers, features such as f616 (the median

of nose expiration in the 1,378 – 1,550 Hz sub-band) and f522 (spectral kurtosis of the 344 – 516 Hz sub-band during nose expiration) were repeatedly selected. The high BMI sub-classifier consistently included f362, which quantifies the spectral kurtosis of the 344 – 516 Hz sub-band during nose inspiration, whereas the low BMI sub-classifier tended to favor mid to higher-frequency sub-band descriptors, particularly mean and spectra related features above 800 Hz. Both high MpS and low MpS sub-classifiers exhibited a preference for features from mid- to high-frequency ranges (around 1200 – 1,700 Hz), underscoring the relevance of higher frequency content in these groups. For the non-current smokers sub-classifier, f599 (zero-crossing rate of the 1,206 – 1,378 Hz sub-band during nose expiration) and higher-order statistical and/or spectral features were frequently included, highlighting the role of temporal and spectral complexity in this subgroup. The consistent selection of specific sub-band/descriptor pair serves as a complementary, high-resolution signal that distinguish OSA from non-OSA individuals within the distinct sub-classifiers.

To further investigate whether the repeatedly selected acoustic features were directly associated with the subgroup characteristics themselves, each feature was compared between the corresponding opposing subgroups. Most selected features did not show statistically significant differences between the opposing groups. Only feature 276 showed an unadjusted significant difference between male and female

subgroups (p = 0.032); however, no comparison remained significant after Benjamini-Hochberg correction (p = 0.161). These results indicate that the subgroup-specific feature selections were not simply explained by global differences between the opposing anthropometric groups; instead, their significance might be observed within the balanced training subsets used to construct the corresponding sub-classifiers.

The association between the selected acoustic features and AHI was also evaluated within each balanced training subset used to train the corresponding sub-classifier. Most selected features showed weak and non-significant associations with AHI according to Spearman correlation. Significant correlations were observed for feature 276 (ρ = -0.345) within male subsets, feature 616 (ρ = 0.351) and feature 373 (ρ = 0.428) within high age subsets, and feature 362 (ρ = 0.36) within high BMI subsets. These findings suggest that some subgroup-specific acoustic features may also reflect changes associated with OSA severity, although the observed relationships were moderate rather than strong.

4.2 Classification Performance of the Proposed Framework

4.2.1 Performance Across Four Folds

The proposed classification algorithm was evaluated using stratified 4-fold cross-validation to ensure a balanced distribution of OSA and non-OSA subjects in

each fold, and to assess its generalizability across different data splits. Table 6 presents the classification metrics across each fold, including balanced accuracy, accuracy, sensitivity, specificity, and area under the curve (AUC).

Table 6. Classification performance of the proposed OSA detection algorithm across 4-folds cross-validation using probability-based voting. Reported metrics include balanced accuracy (average of sensitivity and specificity), accuracy, sensitivity, specificity, and AUC. Values are given for each fold and as mean ± standard deviation across folds.

Metric	Fold 1	Fold 2	Fold 3	Fold 4	Average
Balanced Accuracy	74.5	68.4	72.4	69.6	72.1±2.0
Accuracy	75.5	73.5	79.6	79.6	77.1±3.1
Sensitivity	76.3	78.4	86.5	89.2	84.3±5.8
Specificity	72.7	58.3	58.3	50.0	59.9±9.4
AUC	0.80	0.78	0.74	0.74	0.77±0.01

The algorithm achieved an average balanced accuracy of 72.1 ± 2.0%, slightly lower than the overall accuracy of 77.1 ± 3.1%. This difference indicates better performance on OSA subjects than on non-OSA subjects. The sensitivity of 84.3 ± 5.8% shows that the proposed algorithm correctly identified most OSA subjects, while the specificity of 59.9 ± 9.4% implies that the framework was less effective in distinguishing non-OSA subjects.

The mean AUC was 0.77 ± 0.01, indicating fair to good discriminative capability irrespective of the decision threshold. Sensitivity was consistently high across folds, while specificity exhibited greater variability, ranging from 72.7% in fold 1 to 50.0% in fold 4. This variability suggests that the non-OSA classification

performance is more sensitive to the particular combination of training and testing datasets.

4.2.2 Probability-Based Voting Versus Majority Voting

Majority voting performance is presented in Table 7. Majority voting attained a mean balanced accuracy of 71.2 ± 5.5% and an overall accuracy of 78.1 ± 3.1% across the four folds. The approach exhibited high and stable sensitivity (84.6 ± 1.3%); however, its specificity was comparatively low (57.8 ± 10.7). When comparing the two voting approaches, both achieved comparatively similar sensitivity; however, probability-based voting yielded greater specificity (by approximately 2.1%) and slightly improved overall discriminative capability. This indicates that probability aggregation better utilizes the confidence levels of the individual sub-classifiers, whereas majority voting weights all predictions equally. Therefore, probability-based voting may be preferable for screening, as it maintains good sensitivity while reducing the false alarm rate.

Table 7. Classification performance of the proposed OSA detection algorithm across 4-folds cross-validation using majority voting. Reported metrics include balanced accuracy (average of sensitivity and specificity), accuracy, sensitivity, and specificity. Values are given for each fold and as mean ± standard deviation in percentage across folds.

Metric	Fold 1	Fold 2	Fold 3	Fold 4	Average
Balanced Accuracy	78.5	66.9	72.4	66.9	71.2±5.5
Accuracy	81.6	75.5	79.6	75.5	78.1±3.1
Sensitivity	84.2	83.8	86.5	83.8	84.6±1.3
Specificity	72.7	50.0	58.3	50.0	57.8±10.7

The decision threshold was also optimized to address the class imbalance and minimizing false negatives. Figure 8 illustrates the trade-off between accuracy, sensitivity, and specificity across thresholds ranging from 0.4 to 0.6. As shown, the default threshold of 0.5 yielded a high sensitivity (~84%) but moderate specificity (~60%). In-creasing the threshold toward 0.55 improves specificity to approximately 65% but causes a sharp decrease in sensitivity to 68%. The crossover point, where sensitivity and specificity are balanced, occurs at approximately 0.56; however, sensitivity drops below 70% at this point, which is not optimal for a screening application. Consequently, the threshold 0f 0.5 was retained to prioritize the detection of true OSA subjects.

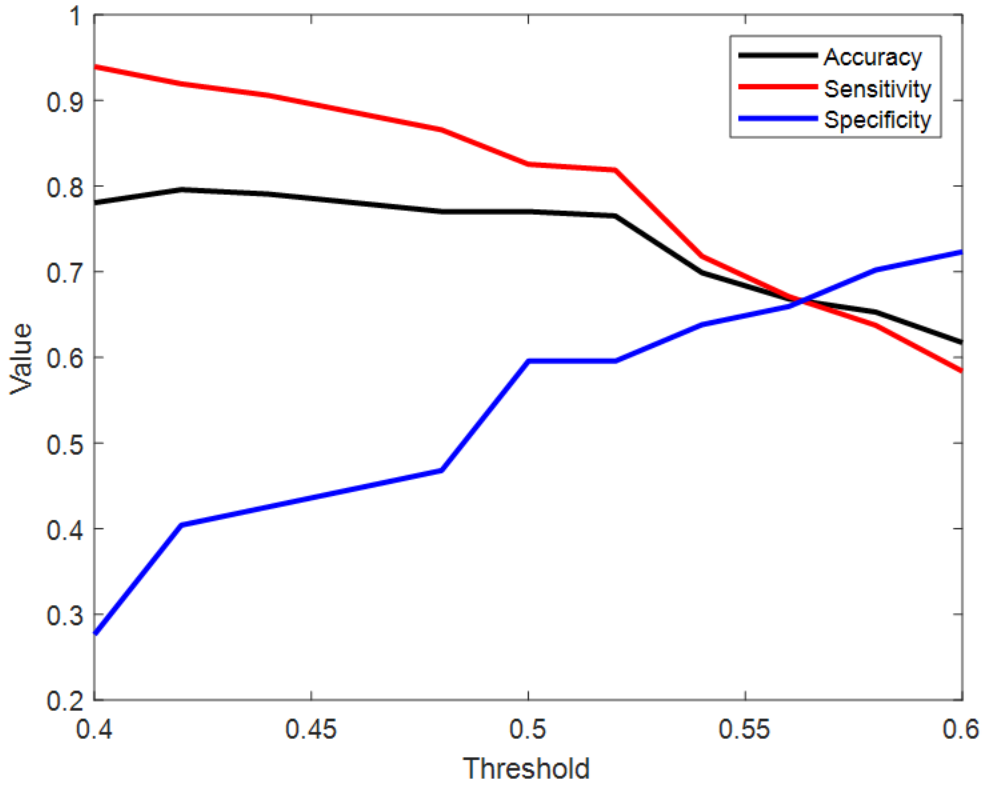

Figure 8. Performance trade-off analysis across varying probability thresholds, measured by accuracy (black line), sensitivity (red line), and specificity (blue line).

4.3 Sub-Classifier Ablation Analysis

To gain deeper insight into the relative contribution of each anthropometric sub-group to the final classification output, a sub-Classifier ablation analysis was performed, which is adapted from component-wise ablation procedure used in ensemble learning studies [129,130]. In this analysis, predictions of one sub-classifier (e.g., male, female) were removed at a time, and the final output was computed based on the remaining eight sub-classifiers as reported in Table 8. This

approach facilitates the investigation of each sub-classifier's contribution by considering the variation in performance resulting from excluding its predictions.

Table 8. Performance of the ensemble framework under sub-classifier ablation analysis. Each row shows results when predictions from the corresponding sub-classifier were excluded from the ensemble decision. Reported metrics included balanced accuracy, accuracy, sensitivity, and specificity (%).

Sub-Classifier	Balanced Accuracy	Accuracy	Sensitivity	Specificity
Male	68.9	74.0	79.2	57.8
Female	67.4	73.5	79.2	55.7
High Age	70.5	77.0	83.3	57.8
Low Age	71.2	77.0	82.6	59.9
High BMI	68.2	73.5	78.5	57.8
Low BMI	69.5	76.5	83.3	55.7
High MpS	71.2	77.0	83.6	59.9
Low MpS	70.2	76.5	82.6	57.8
Non-Current Smokers	68.3	77.6	85.2	51.3

Overall, removal of individual sub-classifiers substantially reduced the performance, indicating that each sub-classifier added unique and informative value to the overall ensemble. Removing the female sub-classifier, in particular, resulted in the lowest balanced accuracy rates (67.4%) and specificity (55.7%), suggesting that its predictive power is highly beneficial to the ensemble approach. Similarly, removing the high BMI sub-classifier lowered the balanced accuracy to 68.2%, consistent with the established association between obesity and OSA. Excluding either low age or high MpS sub-classifiers had smaller impacts with balanced

accuracy remaining close to 71%, which may reflect possible redundancy or overlapping with other subgroups.

Interestingly, removing the non-current smokers sub-classifier resulted in a slight increase in sensitivity (85.2%), but a noticeable reduction in specificity (51.3%), which indicates its role in distinguishing non-OSA subjects and controlling false positive rates. Taken together, the sub-classifier ablation analysis shows that while all sub-classifiers contribute to the ensemble's robustness, classifiers linked with established OSA risk factors (sex, age, and smoking status) play a critical role in maintaining balanced performance.

4.4 Comparison with Existing and Baseline Methods

In order to evaluate the proposed ensemble method, we compared its performance against three benchmarks: the STOP-Bang questionnaire, our previous AWakeOSA algorithm [25], and a global non-stratified baseline (a standard Bagged Trees ensemble without anthropometric stratification) using the same noisy dataset acquired in this study. Table 9 summarizes the results, showing that the proposed algorithm demonstrated a consistent superiority over all diagnostic methods.

Table 9. Comparison of different diagnostic methods with our proposed ensemble framework. Reported metrics include balanced accuracy, accuracy, sensitivity, and specificity, expressed as mean ± standard deviation in percentage across four folds.

Algorithm	Balanced Accuracy	Accuracy	Sensitivity	Specificity
STOP-Bang [98]	62.15±3.1	76.7±2.9	90.2±3.3	34.1±3.7

AWakeOSA [25]	48.4±1.9	44.9±6.1	37.8±3.6	59.1±5.7
Global Bagged Trees	56.6±4.7	78.6±1.2	98.7±1.6	14.6±10.5
Proposed Framework	**72.1±2.0**	**77.1±3.1**	**84.3±5.8**	**59.9±9.4**

As observed, although STOP-Bang achieved high sensitivity (90.2 ± 2.3%), it suffered from extremely low specificity (34.1 ± 3.7%), resulting in a large number of false positives. In contrast, the AWakeOSA algorithm produced higher specificity (59.1 ± 5.7%) but noticeably low sensitivity (37.8 ± 3.6%), leading to an unreliable OSA screening tool for our dataset. Crucially, the global baseline highlighted the necessity of the proposed two-level architecture. While the global baseline achieved high accuracy of 78.6%, this metric was misleadingly driven by a severe bias toward the majority class (OSA), failing to generalize to non-OSA subjects, resulting in a specificity of only 14.6%. The proposed framework achieved a balanced accuracy of 72.1 ± 2.0%, markedly higher than all comparators. By stratifying the problem into anthropometrically homogeneous subgroups, our framework successfully mitigated the class imbalance that caused the global baseline to fail, maintaining a robust trade-off between sensitivity (84.3 ± 5.8%) and specificity (59.9 ± 9.4%). This balance indicates that the proposed framework enables reliable OSA detection within an affordable and noise-tolerant experimental setup, outperforming questionnaire-based, previous acoustic-based, and global baseline methods.

4.5 Confusion Matrix and Misclassification Analysis

Further insight into the accuracy of the classification of the proposed ensemble algorithm was obtained by examining the confusion metrices of the test folds. The predicted probabilities were converted to binary labels using the selected probability threshold of 0.5. The overall out-of-fold confusion matrix shows that the ensemble correctly identified 123 OSA cases and 28 non-OSA cases, while misclassifying 26 OSA and 19 non-OSA cases. These results translate into an accuracy percentage of 77.0%, a sensitivity percentage of 82.6%, and a specificity percentage of 59.6%, resulting in a balanced accuracy of approximately 71.1%.

To better characterize misclassified subjects, participants were divided into four groups based on their classification results: correctly classified as OSA, correctly classified as non-OSA, false negatives, and false positives (Table 10). Subjects correctly classified as OSA (n = 123) had significantly higher mean AHI (44.7 ± 28.0 events/hour) and higher BMI (36.8 ± 7.1 kg/m²) compared with correctly classified as non-OSA (n = 28), who had low AHI values (5.3 ± 2.6 events/hour) and a smaller neck circumference (38.3 ± 3.8 cm).

Table 10. Clinical and demographic characteristics of subjects stratified by classification outcome (correct OSA, correct non-OSA, false negative, false positive). Continuous variables are expressed as mean ± standard deviation; categorical variables are shown as percentages or count.

Parameters	Correct OSA	Correct non-OSA	False Negative	False Positive
n	123	28	26	19
Male (%)	75.6	25	23.1	63.1

Age	47.7±12.0	43.9±13.3	50.2±2.8	46.3±12.9
BMI (kg/m²)	36.8±7.1	31.7±6.8	31.9±5.2	32.3±7.1
Current Smokers (%)	20.3	21.4	22.5	5.3
Neck Circumference	44.7±4.9	38.3±3.8	38.5±2.8	40.3±3.2
MpS	(25, 22, 34, 42)	(5, 8, 10, 5)	(7, 4, 4, 11)	(5, 3, 7, 4)
Snoring (%)	96.8	85.7	92.3	84.2
Receding Mandible	(4, 116, 3)	(3, 23, 2)	(0, 23, 3)	(1, 71, 1)
AHI	44.7±28.0	5.3±2.6	32.3±15.5	6.3±2.5

The false negative group (n = 26) had severe AHI values (32.3 ± 15.5), lower BMI (31.9 ± 5.2), and a relatively smaller neck size (38.5 ± 2.8 cm) compared with true OSA cases. False positives (n = 19) had mild AHI (6.3 ± 2.5), but their neck circumferences (40.3 ± 3.2 cm) and BMI (32.3 ± 7.1) were closer to the range of OSA, reflecting an anthropometric similarity with true OSA patients. In terms of sex distribution, the majority of false negatives were females (76.9%), while false positives were predominantly males (63.1%). Snoring remained high across all groups (> 84%). The distribution of MpS and receding mandible also indicated overlap between misclassified and correctly classified groups. Overall, misclassification occurred primarily in subjects whose anthropometric features mimicked the opposing class, leading to acoustic patterns that confounded the framework's decision boundary.

4.6 Summary

This chapter presented the results of the proposed wakefulness OSA classification framework based on TBS. The proposed awake OSA framework achieved balanced classification performance, maintained high sensitivity, and outperformed questionnaires, previous acoustic algorithm, and global baseline method. Overall, the findings highlight the importance of population heterogeneity and support the feasibility of wakefulness TBS combined with anthropometric information for accessible OSA screening.

Chapter 5: Discussion

This chapter discusses the main findings of this thesis in relation to OSA severity, population heterogeneity, and wakefulness OSA screening. The chapter discusses the proposed OSA classification framework, including the physiological relevance of selected acoustic features, the role of subgroup ensemble learning, classification performance, and sources of misclassification. Finally, the limitations of the study and future directions for improving practical OSA screening are presented.

5.1 Physiological Interpretation of Selected Acoustic Features

The feature selection analysis revealed that although specific feature sets varied across folds and anthropometric subgroups, certain frequency-band features recurred frequently, indicating a reproducible discriminative pattern. This result suggests that while the implemented methodology successfully extracted discriminative patterns representative of physiological information, features are highly linked to the combination of subjects in the training set due to various selected feature sets. Notably, the feature selection analysis revealed a balanced distribution of features between inspiratory and expiratory phases, suggesting that both parts of the respiratory cycle contribute effectively to discriminative performance. The selection of inspiratory features (e.g., f9) aligns with the established

pathophysiology of OSA, where airway collapse and wall vibration are driven by negative intraluminal pressure during inhalation [45,131]. Simultaneously, the equal recurrence of expiratory features indicates that valuable acoustic information is further present during exhalation due to greater turbulence, and flow–structure interaction during airway reopening [132]. In addition, the selected expiratory features are aligned with a study indicating that expiratory flow limitation (EFL), characterized by upper airway narrowing during expiration, is highly prevalent in OSA patients [133].

At the level of frequency composition, the selected features included both the lowest frequency band (0 – 172 Hz) and mid- to high-frequency ranges (344 – 1,548 Hz). We found that low-frequency measurements (e.g., f9) captured the fundamental resonance of the upper airway, which functionally linked to anatomical dimensions [134]. Conversely, mid- and high-frequency features align with the physics of turbulent airflow; as the airway narrows, localized flow velocity increases, shifting acoustic energy toward higher frequencies [125]. Therefore, the model appears to consider a combination of low-frequency structural resonance and high-frequency turbulent flow markers to distinguish OSA subjects from non-OSA subjects.

The analyses performed on a subgroup level confirmed the occurrence of physiologically important patterns in the sets of features. The most commonly selected feature among male participants was f9 (mouth-inspiration, 0 – 172 Hz,

spectral centroid), representing overall inspiratory load and the propensity for airway collapse even in the presence of increased airway caliber [135]. This interpretation is consistent with reported anatomical differences of the upper airway between males and females. Although upper airway dimensions are normally larger in males than in females [136]; upper airway length has been reported to be longer in males and associated with higher airway collapsibility [137]. A longer and more collapsible airway may therefore make low frequency mouth inspiration components more informative in males. By contrast, f410 (nose-inspiration 861 – 1,033 Hz, spectral kurtosis) and f412 (nose-inspiration 861 – 1,033 Hz, spectral entropy) were selected as best features in females; these features correspond to decreased upper airway dimensions and an enhancement of higher-frequency irregularities related to inspiration [138]. Harmonical and neuromuscular factors influence upper airway function in females; particularly, progesterone in combination with estrogen has been reported to improve upper airway dilator muscle [139,140]. Additionally, females have been found to have less collapsible and more stable upper airway during sleep than males [137]. Therefore, the selection of higher frequency features in female subgroup may reflect more localized airflow patterns during inspiration rather than the low frequency resonance pattern observed in males. Importantly, comparisons of the selected features between the male and female groups showed no statistically significant differences after Benjamini–Hochberg correction. Thus,

their selection should not be interpreted as evidence that these feature values are intrinsically different between sexes. Within the balanced male training subsets, however, f276 showed a moderate negative association with AHI ($\rho = -0.345$), indicating that its relevance within the male classifier may be related to variation in AHI within male subjects rather than to a general sex difference in the feature itself.

For high-BMI subgroups, the selected feature with the highest consistency was f362 (nose-inspiration 344 – 516 Hz, spectral kurtosis), representing the effects of mid-frequency nasal-inspiratory irregularity most likely due to obesity-induced soft-tissue loading [141]. Obesity can increase pharyngeal collapsibility and narrow the upper airway through mechanical loading of the pharyngeal soft tissues and reduced lung volumes, which reduce longitudinal traction on the upper airway [142]. By contrast, in low-BMI subgroup f410 (nose-inspiration, 861 – 1,033 Hz, spectral kurtosis) together with f113 (mouth-inspiration, 1,206 – 1,378 Hz, mean amplitude) were primarily selected. This suggests that OSA in lower-BMI individuals is less dominated by soft tissue loading and may instead be driven by other anatomical and physiological mechanisms [143]. Feature 362 did not differ significantly between the high and low BMI groups; however, within the balanced high BMI training subsets, it showed a moderate positive association with AHI ($\rho = 0.36$). Therefore, its repeated selection in the high BMI classifier appears to be related more to its

association with AHI within that subgroup than to a general difference in feature magnitude between high and low BMI subjects.

Likewise, age-related sub-classifiers agreed in consistently emphasizing nasal-expiration features f522 (nose-expiration, 344 – 516 Hz, spectral kurtosis) and f616 (nose-expiration, 1,378 – 1,550 Hz, median amplitude), posing the possibility that expiratory turbulence could represent an indicator of age-related changes in tissue compliance [144]. The opposing-group analysis showed that the repeatedly selected features did not differ significantly between the high and low age groups. However, within the balanced high age training subsets, f616 and f373 showed moderate positive associations with AHI ($\rho = 0.351$ and $\rho = 0.428$, respectively). These findings indicate that their relevance was not explained by a general age-related shift in feature values; rather, their values were associated with AHI within the high age subgroup.

As for the effect of smoking, the non-current/former smoker sub-classifier included mostly f599 (nose-expiration, 1,206 – 1,378 Hz, zero-crossing rate), suggesting that temporally irregular patterns of airflow are particularly relevant when smoking-related airway inflammation is absent. The high zero crossing rate, which counts how fast a sound wave changes from positive to negative, indicates that the TBS is changing rapidly. This rapid fluctuation reflects the uneven, complex acoustic created when air moves cleanly through unswollen upper airway structure

[145]. Additionally, this finding is aligned with the supporting analysis in Appendix A, where current smokers showed higher AHI values than non-smokers and former smokers, while no significant difference was observed between non-smokers and former smokers.

Crucially, the distinct feature patterns observed across all subgroups suggest that the classification was not driven solely by the three anthropometric features (neck circumference, snoring, and receding mandible). If the model relied predominantly on these anthropometric features, the acoustic features would be expected to contribute less consistently across the subgroup-specific classifiers. Instead, the emergence of distinct high-stability acoustic features suggests that the acoustic data provided complementary discriminative information beyond the anthropometric features alone [25,146]. The anthropometric subgroups therefore provide the context in which the sub-classifiers operate, while the acoustic features contribute additional information for distinguishing OSA from non-OSA subjects within those contexts.

The selected features can be mapped to physiologically interpretable properties of the upper airway. The spectral centroid expands the distribution of energy across frequencies. Kurtosis and skewness capture resonance and asymmetry, respectively. Entropy represents signal complexity, and the zero-crossing rate indicates rapid oscillatory changes, reflecting the high-frequency spectral content

associated with fine-scale flow disturbance and turbulent noise. These feature interpretations are consistent with respiratory sound studies that used spectral centroid, spectral kurtosis, spectral skewness, entropy, and zero-crossing rate for respiratory sound characterization and classification [147,148]. The recurrence of these features across folds and sub-classifiers highlights their stability and robustness within the proposed framework. Previous TBS studies further support that airway geometry, constriction, and airflow dynamics can influence tracheal breathing sound spectra [122,149]. Overall, our findings imply that the combined contribution of inspiratory and expiratory features captures discriminative information. This simultaneous contribution underscores the importance of including both inspiration and expiration phases in the recording protocols.

5.2 Performance of the Proposed TBS-Based Framework

Application of an ensemble learning approach was essential in boosting the model's robustness as well as generalizability. Nine sub-classifiers were constructed using stratified subsets defined by anthropometric factors, namely sex, age, BMI, MpS, and smoking history. The resultant subgroup-specialized training allowed for each classifier to exploit the best relevant features in their sub-population, thus reducing feature relevance heterogeneity across the entire population. A probability-based voting scheme was adopted to combine all the viable sub-classifiers determined by the anthropometric features of individuals. This voting mechanism

reduced abrupt transitions and provided smoother decision boundaries. In addition, it enabled the ensemble to exploit the varying confidence levels of sub-classifiers, reducing susceptibility to overfitting.

The ensemble fusion strategy resulted in an improved balanced accuracy (72.1 ± 2.0%) and a better sensitivity-specificity trade off compared with the STOP-Bang questionnaire, the previous AWakeOSA algorithm, and a global non-stratified baseline. STOP-Bang achieved high sensitivity but very low specificity, indicating that it tended to over-classify subjects as OSA [98]. Similarly, the global bagged tree model achieved very high sensitivity but poor specificity, suggesting that a single non-stratified classifier was biased toward the majority OSA class. The previous AWakeOSA algorithm also showed reduced performance on this dataset, likely because it was originally developed under different recording conditions and with different microphone characteristics [25]. Compared with previous wakefulness TBS studies, the proposed framework showed lower accuracy than studies conducted under more controlled recording conditions, such as the LR/RF-based model with 82.1% accuracy [30] and the anthropometric subgroup voting approach with 83.6% accuracy [33]. However, unlike those studies, the present framework was evaluated using a low-cost microphone in a noisy clinical environment, which likely reduced signal quality and classification performance. The enhanced performance supports the hypothesis that anthropometric-based sub-classifiers

enable the models to learn the acoustic and physiological variations associated with OSA within each sub-classifiers, which single global models fail to capture effectively. The sub-classifier ablation analysis also confirms the effectiveness of the fusion schema and the unique contribution of each sub-classifier to final classification.

These findings illustrate how the ensemble fusion leverages inter-subgroup diversity which is an essential property for achieving robust and unbiased predictions. It not only mitigated class imbalance and subject heterogeneity but also enhanced interpretability by revealing which anthropometric profiles contribute most to classification performance. The outcomes across multiple analysis confirm that the diversity and probabilistic integration of sub-classifiers have great potential to improve generalization and good sensitivity-specificity, even with a noisy recording.

5.3 Ensemble Fusion and Sub-Classifier Contribution

The developed ensemble algorithm achieved an average balanced accuracy of 72.1%, with a high sensitivity of 84.3% and a relatively lower specificity of 59.9%. The AUC of 0.77 indicates a fair to good discriminative capacity. When compared to other awake OSA screening modalities, such as speech-based frameworks which report accuracies higher than 80% [115], or previous TBS algorithms using advanced

machine learning in quiet environments [23,24], our performance appears lower. However, it is essential to consider the experimental setting; while mentioned studies were largely validated in soundproof and highly con-trolled environments, this study was conducted in a noisy clinical condition with an average SNR of approximately 10 dB. In such realistic conditions, the harmonic structures required for speech analysis are often masked, and high-complexity models are prone to overfitting ambient noise. Therefore, the 77.1% accuracy achieved here represents a robust performance baseline for real-world clinical deployment using inexpensive equipment.

The results imply that the framework is reliable in detecting OSA during wakefulness in a noisy environment and exhibits stable classification performance for various fold partitions. Notably, the higher sensitivity compared to specificity im-plies that the system is more effective at detecting OSA cases than at accurately excluding non-OSA subjects. Our threshold optimization analysis (Figure 8) highlighted the critical trade-off inherent to this class imbalance. While increasing the decision threshold could balance sensitivity and specificity, this adjustment would result in a substantial reduction in sensitivity. In clinical settings, this trade-off is acceptable for a screening tool, as minimizing false negatives is important in the prevention of adverse effects from undetected OSA, even at the expense of the possible increase in false positives that require confirmatory PSG follow-up [150].

Therefore, maintaining the threshold at 0.5, despite the lower specificity, prioritizes the capture of at-risk individuals, ensuring that the framework functions effectively. The observed performance trends can be attributed to OSA-related physiology as well as the limitations of the methodology adopted. For instance, OSA individuals have significantly elevated values of upper airway collapsibility, as measured by both acoustic analysis of breathing sounds [151], and anthropometric techniques like acoustic pharyngometry [152], leading to increased sensitivity. In contrast, the subjects who do not meet the criteria for OSA but have intermediate risk factors such as increased BMI or larger neck circumference might manifest similar acoustic and anthropometric features, thus reducing specificity [153,154].

These findings support the importance of accounting for population heterogeneity in OSA screening. Anthropometric features such as sex, age, BMI, MpS, neck circumference, and smoking history may influence both OSA presentation and acoustic breathing patterns. The supporting analysis in Appendix A further examines smoking history as one such factor.

5.4 Error Analysis and Practical Limitations

The aggregated confusion matrix (Table 10) showed that the ensemble correctly classified 123 individuals with obstructive sleep apnea (OSA) and 28 without OSA and misclassified 26 instances of OSA (false negatives) and 19

instances of non-OSA (false positives). Examining the clinical and demographic characteristics of these groups provided important insight into the sources of the errors.

False negatives (n = 26) consisted of patients with OSA with lower AHI values (32.3 ± 15.5 events/hour) compared with the correctly classified OSA subjects (44.7 ± 28.0 events/hour). This group also had lower anthropometric risk factors, including reduced BMI (31.9 ± 5.2 kg/m²) and lower neck circumference (38.5 ± 2.8 cm). In addition, a majority of this population were females (76.9%), suggesting that sex-related differences in acoustic markers could have contributed to their misclassification. While these patients were diagnosed clinically with OSA, their acoustic and anthropometric features were indistinguishable from those of non-OSA subjects, thereby impairing the sensitivity of the classifier.

False positives (n = 19) had low AHI levels (6.3 ± 2.5 events/hour) that validated their classification as non-OSA; however, their BMI (32.3 ± 7.1 kg/m²) and neck size (40.3 ± 3.2 cm) were similar to those patients with OSA, with an average BMI (36.8 ± 7.1 kg/m²) and neck circumference (44.7 ± 4.9 cm). This anthropometric similarity probably led to the acoustic characteristics of these patients being incorrectly interpreted as those of OSA. This interpretation is consistent with previous studies showing that obesity and increased neck circumference can increase pharyngeal collapsibility and upper-airway mechanical

loading, while male sex is also associated with increased upper airway mechanical instability during sleep [155,156]. A disproportionate number of these cases were male (63.1%), as expected for the known tendency for male airway anatomy to exhibit increased collapsibility in the absence of OSA [136,137,155]. Overall, misclassifications were found to be consistently associated with participants having borderline phenotypes. These would include OSA patients with decreased anthropometric measurements and lower AHI values, as well as non-OSA participants with anthropometric features usually characteristic of OSA.

These findings may also reflect the multifactorial pathophysiology of OSA, whereby anthropometric risk factors do not necessarily correspond directly to airway obstruction during sleep [43,48]. OSA can occur in individuals without pronounced obesity or increased neck circumference because airway collapsibility is also influenced by craniofacial anatomy, upper airway neuromuscular control, and other physiological mechanisms [157,158]. Conversely, individuals with obesity, larger neck circumference, or male sex may possess recognized anatomical risk factors without developing clinically significant airway obstruction during sleep [43,159].

Non-OSA subjects with AHI values in the intermediate range ($10 \leq AHI \leq 15$), who were excluded from the training phase to ensure clearer separation between OSA and non-OSA classes, were evaluated separately as an independent test group. The majority of these subjects were classified as OSA subjects; the average was

39.16 ± 2.03 % across four folds. This reduced performance was anticipated and reflects the inherent ambiguity of this AHI range which lies very close to the diagnostic threshold of OSA (AHI = 15). It is therefore debatable whether individuals with an AHI of 10 – 15 exhibit clinically meaningful pathological differences compared to those with an AHI of 15 – 20. Nevertheless, AHI remains the most widely used diagnostic metric for identifying patients requiring treatment [41]. In contrast, we hypothesize that tracheal breathing sounds capture underlying physiological alterations associated with OSA more directly, rather than relying on a rigid threshold-based definition. It should further be noted that the BMI, age, and neck circumference of these participants were also at the borderline values between the corresponding groups; thus, the borderline nature of anthropometric parameters of these subjects could also be contributing to the low accuracy for this group's classification.

In addition, recording instrumentation may have further impacted classification performance. The Primo EM273 capsule was used in this study, while Sony ECM77B microphone was employed in the previous studies [23,25,28–33,123]. Although both microphones are omnidirectional electret condenser microphones, they differ in their electroacoustic characteristics and frequency-response profiles. The microphone frequency response showed that the ECM77B provided a more uniform response across the frequency range of interest, whereas

the primo EM273 exhibited greater frequency-dependent variation. Consequently, the same underlying mixture of tracheal breathing sounds, oral/nasal breathing components, and environmental sounds may be represented differently by the two recording systems. However, the presence of environmental noise and harmonic artifacts should not be attributed directly to the lower cost of the primo EM273, as similar acoustic contamination may occur with either recording system and the independent contribution of microphone type was not isolated experimentally.

Despite these acquisition differences, the results demonstrate that the primo EM273 microphone preserved diagnostically useful wakefulness TBS information, as the proposed framework achieved a balanced accuracy of 72.1 ± 2.0%, with 84.3 ± 5.8% sensitivity and 59.9 ± 9.4% specificity. However, applying the previously developed AWakeOSA framework [25] directly to the present recordings resulted in substantially lower performance (48.4 ± 1.9% balanced accuracy, 44.9 ± 6.1% accuracy, 37.8 ± 3.6% sensitivity, and 59.1 ± 5.7% specificity), indicating that the previous framework did not transfer effectively to the new acquisition system. The proposed subgroup-based framework also outperformed the global non-stratified baseline (56.6 ± 4.7% balanced accuracy), supporting the development of a dedicated framework for the present recordings rather than direct reuse of the previous model. These findings support the feasibility of low-cost wakefulness TBS

acquisition while also demonstrating that changes in recording hardware may require adaptation or redevelopment of the associated classification framework.

Chapter 6: Conclusion

This thesis investigated the feasibility of wakefulness OSA screening framework using tracheal breathing sounds (TBS). OSA remains a major clinical concern because of its high prevalence, health consequences, and the limited accessibility of standard diagnostic methods such as polysomnography. Therefore, accessible and non-invasive screening methods are needed to identify individuals at risk of OSA before confirmatory sleep assessment.

A wakefulness OSA screening machine learning framework was developed using TBS recordings obtained with a low-cost primo EM273 microphone in a practical clinical recording environment. Wavelet packet decomposition was used for acoustic feature extraction, mRMR was applied for feature selection, and subgroup ensemble classifiers were combined using probability-based voting. The proposed framework achieved 72.1% balanced accuracy, 84.3% sensitivity, and 59.9% specificity, indicating its potential as a practical screening tool, particularly for identifying individuals with OSA. In contrast, applying the previously developed AWakeOSA framework [25] to the present recordings resulted in substantially lower performance, with a balanced accuracy of 48.4%, indicating limited transferability of the previous framework to the current acquisition setting. The proposed framework also outperformed the global non-stratified baseline, supporting the use

of subgroup-specific modeling for this heterogeneous population. However, these findings do not establish that the microphone alone caused the reduced performance of the previous framework, as differences in microphone response, acquisition conditions, and study population were not independently isolated.

This thesis also had several limitations. Although the final dataset included 196 individuals, the number of non-OSA subjects was limited (n = 47), resulting in class imbalance particularly in high BMI subgroup. In addition, the recording quality was heterogeneous. Several participants were excluded because of low signal-to-noise ratio, suboptimal breathing phases, or persistent harmonic noise, and the remaining recordings still contained background noise and artifacts that may have influenced feature extraction and classification performance. Finally, subjects with intermediate AHI values ($10 \leq AHI \leq 15$) were excluded to improve separation between OSA and non-OSA groups. While this exclusion helped the model learn clearer acoustic patterns, it limits the immediate generalizability of the framework to borderline clinical cases.

Future work should validate the proposed model on independent datasets with larger numbers of non-OSA subjects and wider ranges of anthropometric profiles. Subjects with intermediate AHI values should also be evaluated to determine whether their acoustic characteristics align more closely with non-OSA or OSA phenotypes. Paired recordings using the Primo EM273 and Sony ECM77B under

identical subject and recording conditions could help determine how differences in microphone frequency response affect TBS characteristics and model transferability. Future studies could also investigate frequency bands above the 0–1,722 Hz range used for feature extraction, since localized spectral differences were observed at higher frequencies. The candidate feature space could be expanded to include nonlinear and fractal descriptors, such as fractal dimensions and the Hurst exponent, as well as higher-order spectral features such as bispectral or bicoherence-based parameters, to determine whether they provide complementary information related to upper-airway obstruction. Adaptive signal decomposition techniques, including empirical mode decomposition (EMD) and variational mode decomposition (VMD), could also be investigated to isolate data-driven oscillatory components that may not be optimally represented by the fixed WPD sub-bands. Future multimodal extensions may additionally incorporate speech recordings, as speech acoustics can provide complementary information related to upper-airway anatomy and resonance characteristics that may not be fully captured by TBS alone.

Overall, this thesis shows that combining anthropometric information with wakefulness TBS features can support accessible and non-invasive OSA screening. The results demonstrate that diagnostically useful acoustic information can be obtained using a low-cost microphone, while the poor transferability of the previous AWakeOSA framework indicates that models developed using one acquisition

system may not generalize directly to another. The improved performance of the proposed subgroup-based framework further supports the importance of accounting for population heterogeneity in wakefulness TBS screening.

Appendix A: Statistical Analysis of Smoking Status and OSA severity

A.1 Introduction

Smoking remains popular among people worldwide, often regarded as a coping mechanism for stress. The act of smoking introduces harmful chemicals into the body, leading to inflammation, swelling, and increased mucus production in the upper airway [160]. These effects may contribute to airway narrowing or obstruction, which are directly related to OSA pathophysiology.

However, the relationship between smoking status and OSA severity remains controversial. Some studies have reported that current smokers are more likely to experience severe OSA and that higher smoking exposure is associated with increased AHI and reduced overnight oxygen saturation [75,76]. In contrast, other studies have found that the association between smoking and OSA severity becomes non-significant after adjusting for covariates such as age, sex, and BMI [77,78]. These conflicting findings suggest that smoking status should be interpreted alongside other demographic, anthropometric, and physiological risk factors.

This appendix presents a supporting statistical analysis of the association between smoking status and OSA severity in the study dataset. AHI values were compared across non-smokers, former smokers, and current smokers, both in the

overall cohort and within selected anthropometric subgroups. In addition, a covariate-adjusted generalized linear model was used to examine whether the association between smoking status and AHI remained after accounting for factors such as age, sex, BMI, and MpS. This analysis provides additional context for smoking history as one anthropometric subgrouping criteria in the broader OSA dataset.

A.2 Methodology

This appendix used the PSG reports and anthropometric information from the same cohort to examine the association between smoking status and OSA severity. The dataset included age, sex, BMI, neck circumference, MpS, and smoking history. Smoking history was categorized into three levels: 0 (non-smoker), 1 (former-smoker), and 2 (current-smoker). AHI was used as the outcome variable representing OSA severity. Table 11 presents a descriptive analysis of the parameters of our dataset used in this analysis.

Table 11. Dataset description used in smoking status relationship with OSA severity analysis.

Class	n	Age	Sex (M/F)	BMI	MpS (<=II, >II)	Neck Circumference	AHI Categories (<5, 5-15, 15-30, >30)
Non-Smokers	150	47.5±13.1	(87, 63)	34.2±8.9	(69, 81)	41.9±5.3	(16, 42, 39, 53)
Former-Smokers	57	52.4±11.0	(34, 23)	33.4±7.6	(25, 32)	42.4±5.4	(2, 15, 25, 15)
Current-Smokers	40	42.4±11.2	(27, 13)	37.5±8.5	(12, 28)	44.3±4.7	(2, 7, 7, 24)

The AHI values were compared across the three smoking groups (non-smoker, former-smoker, current-smoker). Initially, the AHI distributions were evaluated to assess whether they follow a normal distribution using the Shapiro-Wilk test. If the test confirmed that the AHI distribution was normal, a one-way ANOVA was used; otherwise, the Kruskal-Wallis test, a robust non-parametric alternative [161], was employed. In all instances, a p-value below 0.05 was considered statistically significant. If the statistical test showed a significant difference across the smoking groups, the Benjamin-Hochberg (BH) procedure [162] was applied to adjust for multiple comparisons. The BH procedure was chosen because it controls the false discovery rate (FDR) while being less conservative than methods, such as the Bonferroni correction [163].

Then, the subjects were divided into ten subgroups based on their anthropometric features: male, female, high age (age > 50), low age (age ≤ 50), high BMI (BMI > 35), low BMI (BMI ≤ 35), High MpS (MpS ≥ 3), low MpS (MpS < 3), high NC (NC ≥ 40), and low NC (NC < 40). These stratifications were aligned with clinical decision threshold from the STOP-Bang questionnaire [98], to enable direct comparison with prior studies using similar categorization. The AHI values were then compared across the smoking history groups within each subgroup, using the same procedure described above.

Lastly, generalized linear models (GLM) [164] were used to conduct an adjusted analysis where the dependent variable was AHI and independent variables included smoking history, with adjustment variables including age, sex, BMI, and MpS. If the AHI distribution did not follow a normal distribution, a Gamma distribution with a logarithmic link function was applied.

All statistical analyses and methods were implemented in RStudio (version 2024.04.2+764) to ensure reproducibility and consistency.

A.3 Results

Initially, the Shapiro-Wilk test was conducted to assess the normality of AHI values across smoking groups. The results indicated that non-smokers ($p < 0.001$), former-smokers ($p < 0.001$), and current-smokers ($p = 0.043$) significantly deviated from a normal distribution. Therefore, the Kruskal-Wallis test was applied to evaluate significant differences in AHI values across smoking groups. This test revealed a significant difference in AHI values among the groups ($X2 = 8.39$, $p = 0.015$). Benjamin-Hochberg test determined that the current-smokers group was significantly different from non-smokers ($p = 0.007$) and from former-smokers ($p = 0.014$). There was no significant difference between former-smokers and non-smokers ($p = 0.451$).

The AHI values distribution within each smoking group and all of the anthropometric subgroups deviated from normal distribution ($p < 0.001$). Consequently, the Kruskal-Wallis test was applied to evaluate significant differences in AHI between smoking groups; the results are presented in Table 12. The difference of AHI values across smoking groups within male and high NC subgroups were found to be significant, within low age and low BMI to be marginally significant, and within other subgroups were not significant. The Benjamin-Hochberg comparison has shown that within the male subgroup the current-smokers group showed a significant difference compared to with former-smokers ($p < 0.001$) and non-smokers ($p < 0.001$); however, there is no significant difference between former-smokers and non-smokers ($p = 0.451$). Further, the high NC subgroup demonstrated a significant variability across smoking history groups ($X^2 = 8.98$, $p = 0.011$). The BH multiple comparison indicated that current-smokers significantly differed with former-smokers ($p < 0.001$) and non-smokers ($p < 0.001$), while no significant difference was observed between former-smokers and non-smokers ($p = 0.177$).

Table 12. Kruskal-Wallis test results across subgroups for differences in AHI values by smoking history groups. (* for p < 0.05, ** for p < 0.01).

Subgroup	X^2	p-value	Sample Size (non, former, current smokers)
Male	14.38	0.000**	(87, 34, 27)
Female	0.41	0.82	(63, 23, 13)
High Age	2.89	0.236	(65, 34, 10)
Low Age	5.72	0.057	(85, 23, 30)

High BMI	4.14	0.126	(58, 21, 18)
Low BMI	4.65	0.098	(92, 36, 22)
High MpS	4.86	0.088	(81, 32, 28)
Low MpS	3.06	0.216	(69, 25, 12)
High NC	8.98	0.011*	(101, 38, 33)
Low NC	3.96	0.138	(49, 19, 7)

As shown in Figure 9(a), the AHI distribution of our dataset was skewed to the right and appeared to follow a Gamma distribution as confirmed by the Q-Q plot illustrated in Figure 9(b). subsequently, a GLM using gamma distribution with logarithmic link was fitted to predict AHI values based on smoking groups as our non-adjusted model. The non-adjusted model indicated that current-smokers had significantly higher AHI values compared to non-smokers, with an estimated 43.18% increase ($p = 0.025$). However, there was no significant difference between non-smokers and former-smokers ($p = 0.466$). In the adjusted Gamma GLM, current-smokers had approximately 32.18% higher AHI values than non-smokers ($p = 0.059$), although the association was marginally significant. Former-smokers did not show a significant difference in AHI values compared to non-smokers ($p = 0.481$). BMI and sex were strongly associated covariates with AHI values. For each unit increase in BMI, AHI increased by 3.58% ($p < 0.001$). Females had significantly lower AHI values compared to males, with the reduction of 33.99% ($p < 0.001$). Age was also marginally significant, with each unit increase in age resulting in a 0.73% increase in AHI ($p = 0.092$). Furthermore, MpS did not show any significant

association with OSA severity in class 2 ($p = 0.166$), class 3 ($p = 0.666$), or class 4 ($p = 0.103$) compared to class 1.

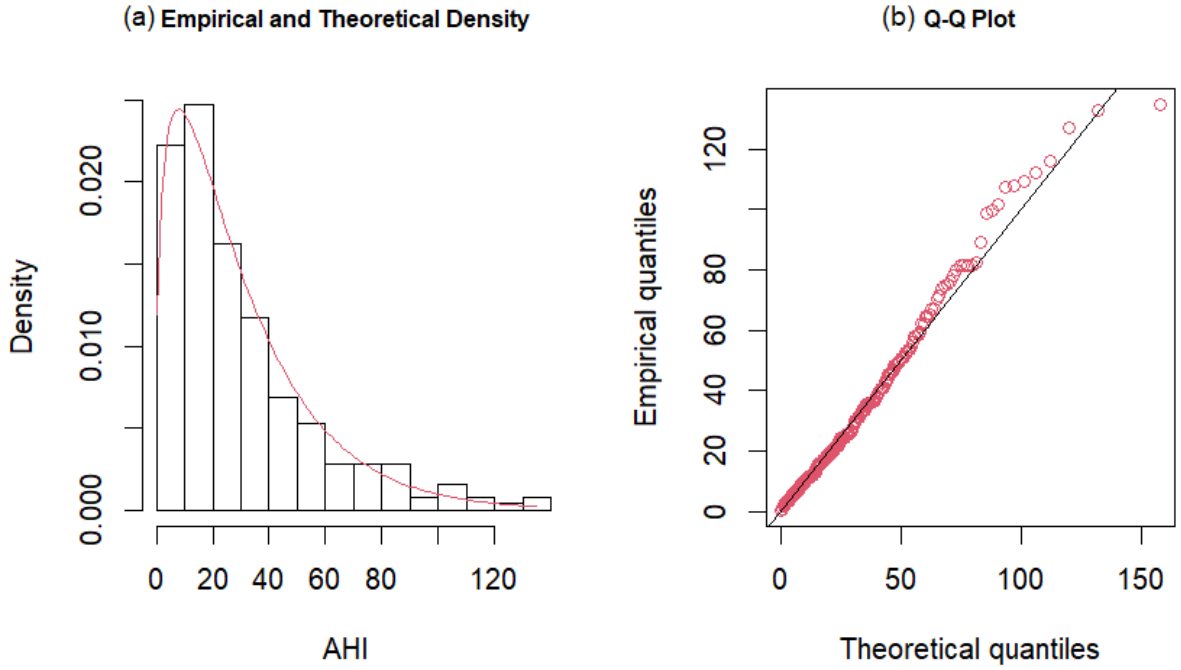

Figure 9. Empirical and theoretical fits for AHI data using the gamma distribution. a) Empirical histogram of AHI values with the theoretical gamma density overlaid in red, b) Q-Q plot comparing the empirical quantiles of AHI data with the theoretical quantiles.

A.4 Discussion

This supporting analysis explored the relationship between smoking history and AHI values to provide additional context on smoking status as one anthropometric factor associated with OSA severity. We found a significant difference between AHI values between current-smokers group and both non-smokers and former-smokers. However, no significant difference was observed

between the non-smokers and former-smokers. These results suggest that the physiological changes due to smoking, such as airway inflammation, might partially reverse upon cessation of smoking. While former-smokers may retain some residual effects of smoking, these appear insufficient to elevate AHI values compared to non-smokers.

Among male and female subgroups, male subjects showed a significant difference in AHI values between smoking groups. The BH statistical comparison implied these differences were between current-smokers and both non-smokers and former-smokers. This may reflect the variations in upper airway anatomy, neurochemical regulation, and sex-specific hormones between males and females [59]. Smoking may change the neurochemicals and hormones and result in changes to upper airway anatomy, thereby elevating OSA severity.

The high age subgroup did not show a significant difference between smoking groups. This may indicate the airway collapsibility with respect to age and other risk factors suppresses the impact of smoking. Conversely, low age subgroups revealed a borderline significant difference, suggesting that smoking affects younger individuals more than older adults, possibly due to fewer competing risk factors in younger populations.

Neither high BMI nor low BMI subgroups exhibited significant differences in AHI across smoking groups. It is possible that the dominant effect of obesity in the high BMI subgroup and reduced baseline risk in the low BMI group minimizes the influence of smoking history.

The high MpS subgroup showed a trend towards significance, whereas the low MpS subgroup did not. This pattern suggests that smoking may interact with other physiological or structural factors related to MpS.

In the high neck circumference subgroup, current-smokers demonstrated significantly elevated AHI values compared to non-smokers and former-smokers. However, the low NC subgroup showed no significant differences. This highlights the combined effect of neck circumference and smoking history on the upper airway obstruction.

The non-adjusted Gamma GLM with logarithmic link showed current-smokers had significantly higher AHI values compared to non-smokers, with an estimated 43.2% increase. On the other hand, no significant difference was observed between former-smokers and non-smokers. AHI values for current-smokers were 28.7% higher than for non-smokers in the adjusted model, but this result was only marginally significant. Former-smokers again did not show a significant difference

compared to non-smokers. It can be concluded that smoking history is not a dominant risk factor when other factors are considered.

Covariates, including sex, age, and BMI were strongly associated with AHI. This association aligns with prior research indicating that these factors influence upper airway obstruction significantly [77]. In contrast, MpS did not show a significant effect on OSA severity, possibly because it overlaps with other anthropometric features. These findings highlight the importance of investigating physiological factors in determining AHI and diagnosing OSA in clinical evaluations.

A.5 Conclusion

This appendix presented a supporting analysis of the relationship between smoking history and obstructive sleep apnea (OSA) severity, measured using the apnea-hypopnea index (AHI). The results indicate that current-smokers have significantly higher AHI values compared to former-smokers and non-smokers, specifically within male and high neck circumference subgroups. Additionally, the effect of smoking status in the generalized linear model adjusted by age, sex, BMI, and MpS was attenuated compared to the non-adjusted model. However, the limited number of subjects in some subgroups may lead to unreliable results and additional studies are necessary to confirm the findings. Future research could investigate this

relationship on larger and more balanced datasets with sufficient numbers of subjects per subgroup. Given the association of smoking and cardiovascular diseases, it would be advantageous to consider the cardiovascular condition of the individuals with OSA as a confounding variable in future research. Moreover, the cessation time is not included in the current study. Exploring the relationship between the duration of smoking cessation and OSA severity may provide further insights into the residual impacts of smoking on the upper airway over time.

References

1. Abbasi, A.; Gupta, S.S.; Sabharwal, N.; Meghrajani, V.; Sharma, S.; Kamholz, S.; Kupfer, Y. A Comprehensive Review of Obstructive Sleep Apnea. *Sleep Sci.* **2021**, *14*, 142–154, doi:10.5935/1984-0063.20200056.

2. Shahar, E. Apnea-Hypopnea Index: Time to Wake Up. *Nat. Sci. Sleep* **2014**, 51, doi:10.2147/NSS.S61853.

3. Berry, R.B.; Abreu, A.R.; Krishnan, V.; Quan, S.F.; Strollo, P.J.; Malhotra, R.K. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*; Version 3.; American Academy of Sleep Medicine: Darien, IL, USA, 2023;

4. Al Lawati, N.M.; Patel, S.R.; Ayas, N.T. Epidemiology, Risk Factors, and Consequences of Obstructive Sleep Apnea and Short Sleep Duration. *Prog. Cardiovasc. Dis.* **2009**, *51*, 285–293, doi:10.1016/j.pcad.2008.08.001.

5. Strohl, K.P.; Altose, M.D. Oxygen Saturation during Breath-Holding and during Apneas in Sleep. *Chest* **1984**, *85*, 181–186, doi:10.1378/chest.85.2.181.

6. Bangash, A.; Wajid, F.; Poolacherla, R.; Mim, F.K.; Rutkofsky, I.H. Obstructive Sleep Apnea and Hypertension: A Review of the Relationship and Pathogenic Association. *Cureus* **2020**, doi:10.7759/cureus.8241.

7. Muraki, I.; Wada, H.; Tanigawa, T. Sleep Apnea and Type 2 Diabetes. *J. Diabetes Investig.* **2018**, *9*, 991–997, doi:10.1111/jdi.12823.

8. Arzt, M.; Young, T.; Finn, L.; Skatrud, J.B.; Bradley, T.D. Association of Sleep-Disordered Breathing and the Occurrence of Stroke. *Am. J. Respir. Crit. Care Med.* **2005**, *172*, 1447–1451, doi:10.1164/rccm.200505-702OC.

9. Kuczyński, W.; Kudrycka, A.; Pierzchała, K.; Grabska-Kobyłecka, I.; Pencina, M.; Sakowski, S.; Białasiewicz, P. Overall Mortality and Comorbidities in Obstructive Sleep Apnea in Poland. *Medical Science Monitor* **2025**, *31*, doi:10.12659/MSM.950826.

10. Yayan, J.; Rasche, K. A Systematic Review of Risk Factors for Sleep Apnea. *Prev. Med. Rep.* **2024**, *42*, 102750, doi:10.1016/j.pmedr.2024.102750.

11. Rundo, J.V.; Downey, R. Chapter 25 - Polysomnography. In *Clinical Neurophysiology: Basis and Technical Aspects*; Levin, K.H., Chauvel, P., Eds.; Handbook of Clinical Neurology; Elsevier, 2019; Vol. 160, pp. 381–392.

12. C. Iber; S. Ancoli-Israel; A. Chesson; S.F. Quan The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology, and Technical Specification. *American Academy of Sleep Medicine* **2007**.

13. Borsini, E.; Nogueira, F.; Nigro, C. Apnea-Hypopnea Index in Sleep Studies and the Risk of over-Simplification. *Sleep Sci.* **2018**, *11*, 45–48, doi:10.5935/1984-0063.20180010.

14. Natsky, A.N.; Vakulin, A.; Coetzer, C.L.C.; McEvoy, R.D.; Adams, R.J.; Kaambwa, B. Economic Evaluation of Diagnostic Sleep Studies for Obstructive Sleep Apnoea: A Systematic Review Protocol. *Syst. Rev.* **2021**, *10*, 104, doi:10.1186/s13643-021-01651-3.

15. Zou, D.; Vits, S.; Egea, C.; Ehrsam-Tosi, D.; Lavergne, F.; Azpiazu, M.; Fietze, I. A New Approach to Streamline Obstructive Sleep Apnea Therapy Access Using Peripheral

Arterial Tone-Based Home Sleep Test Devices. *Frontiers in Sleep* **2023**, *2*, doi:10.3389/frsle.2023.1256078.

16. Zhang, Y.; Chen, K.; Xu, Y.; Wu, H. Research Progress on the First-night Effect in Polysomnography Studies. *Sleep Research* **2024**, *1*, 19–26, doi:10.1002/slp2.12.

17. Kapur, V.K.; Auckley, D.H.; Chowdhuri, S.; Kuhlmann, D.C.; Mehra, R.; Ramar, K.; Harrod, C.G. Clinical Practice Guideline for Diagnostic Testing for Adult Obstructive Sleep Apnea: An American Academy of Sleep Medicine Clinical Practice Guideline. *Journal of Clinical Sleep Medicine* **2017**, *13*, 479–504, doi:10.5664/jcsm.6506.

18. Gurubhagavatula, I.; Sullivan, S.; Meoli, A.; Patil, S.; Olson, R.; Berneking, M.; Watson, N.F. Management of Obstructive Sleep Apnea in Commercial Motor Vehicle Operators: Recommendations of the AASM Sleep and Transportation Safety Awareness Task Force. *Journal of Clinical Sleep Medicine* **2017**, *13*, 745–758, doi:10.5664/jcsm.6598.

19. Bernhardt, L.; Brady, E.M.; Freeman, S.C.; Polmann, H.; Réus, J.C.; Flores-Mir, C.; De Luca Canto, G.; Robertson, N.; Squire, I.B. Diagnostic Accuracy of Screening Questionnaires for Obstructive Sleep Apnoea in Adults in Different Clinical Cohorts: A Systematic Review and Meta-Analysis. *Sleep and Breathing* **2022**, *26*, 1053–1078, doi:10.1007/s11325-021-02450-9.

20. Kim, D.; Park, J.Y.; Song, Y.W.; Kim, E.; Kim, S.; Joo, E.Y. Machine-Learning-Based Classification of Obstructive Sleep Apnea Using 19-Channel Sleep EEG Data. *Sleep Med.* **2024**, *124*, 323–330, doi:10.1016/j.sleep.2024.09.041.

21. Yang, Q.; Zou, L.; Wei, K.; Liu, G. Obstructive Sleep Apnea Detection from Single-Lead Electrocardiogram Signals Using One-Dimensional Squeeze-and-Excitation Residual

Group Network. *Comput. Biol. Med.* **2022**, *140*, 105124, doi:10.1016/j.compbiomed.2021.105124.

22. Yılmaz, D.; Yıldız, M.; Uyar Toprak, Y.; Yetkin, S. Obstructive Sleep Apnea Detection with Nonlinear Analysis of Speech. *Biomed. Signal Process. Control* **2023**, *84*, 104956, doi:10.1016/j.bspc.2023.104956.

23. Alqudah, A.; Moussavi, Z. Assessing Obstructive Sleep Apnea Severity During Wakefulness via Tracheal Breathing Sound Analysis. *Sensors* **2025**, *25*, 6280, doi:10.3390/s25206280.

24. Alqudah, A.M.; Moussavi, Z. Deep Learning Model for OSA Detection Using Tracheal Breathing Sounds During Wakefulness. *CMBES Proceedings* **2023**, *45*.

25. Elwali, A.; Moussavi, Z. A Novel Decision Making Procedure during Wakefulness for Screening Obstructive Sleep Apnea Using Anthropometric Information and Tracheal Breathing Sounds. *Sci. Rep.* **2019**, *9*, 11467, doi:10.1038/s41598-019-47998-5.

26. Alqudah, A.M.; Elwali, A.; Kupiak, B.; Hajipour, F.; Jacobson, N.; Moussavi, Z. Obstructive Sleep Apnea Detection during Wakefulness: A Comprehensive Methodological Review. *Med. Biol. Eng. Comput.* **2024**, *62*, 1277–1311, doi:10.1007/s11517-024-03020-3.

27. Ashraf, W.; Fredberg, J.J.; Moussavi, Z. Aeroacoustics of Breath Sounds in Trachea and Upper Airway. *Applied Acoustics* **2026**, *241*, 111021, doi:10.1016/j.apacoust.2025.111021.

28. Elwali, A.; Moussavi, Z. Obstructive Sleep Apnea Screening and Airway Structure Characterization During Wakefulness Using Tracheal Breathing Sounds. *Ann. Biomed. Eng.* **2017**, *45*, 839–850, doi:10.1007/s10439-016-1720-5.

29. Elwali, A.; Moussavi, Z. Predicting Polysomnography Parameters from Anthropometric Features and Breathing Sounds Recorded during Wakefulness. *Diagnostics* **2021**, *11*, 905, doi:10.3390/diagnostics11050905.

30. Hajipour, F.; Jozani, M.J.; Moussavi, Z. A Comparison of Regularized Logistic Regression and Random Forest Machine Learning Models for Daytime Diagnosis of Obstructive Sleep Apnea. *Med. Biol. Eng. Comput.* **2020**, *58*, 2517–2529, doi:10.1007/s11517-020-02206-9.

31. Hajipour, F.; Jozani, M.J.; Elwali, A.; Moussavi, Z. Regularized Logistic Regression for Obstructive Sleep Apnea Screening during Wakefulness Using Daytime Tracheal Breathing Sounds and Anthropometric Information. *Med. Biol. Eng. Comput.* **2019**, *57*, 2641–2655, doi:10.1007/s11517-019-02052-4.

32. Montazeri, A.; Giannouli, E.; Moussavi, Z. Assessment of Obstructive Sleep Apnea and Its Severity during Wakefulness. *Ann. Biomed. Eng.* **2012**, *40*, 916–924, doi:10.1007/s10439-011-0456-5.

33. Elwali, A.; Meza-Vargas, S.; Moussavi, Z. Using Tracheal Breathing Sounds and Anthropometric Information for Screening Obstructive Sleep Apnoea during Wakefulness. *J. Med. Eng. Technol.* **2019**, *43*, 111–123, doi:10.1080/03091902.2019.1617799.

34. Eckert, D.J.; Jordan, A.S.; Merchia, P.; Malhotra, A. Central Sleep Apnea. *Chest* **2007**, *131*, 595–607, doi:10.1378/chest.06.2287.

35. Khan, M.T.; Franco, R.A. Complex Sleep Apnea Syndrome. *Sleep Disord.* **2014**, *2014*, 798487, doi:10.1155/2014/798487.

36. Sönmez, I.; Vo Dupuy, A.; Yu, K.S.; Cronin, J.; Yee, J.; Azarbarzin, A. Unmasking Obstructive Sleep Apnea: Estimated Prevalence and Impact in the United States. *Respir. Med.* **2025**, *248*, 108348, doi:10.1016/j.rmed.2025.108348.

37. Rodin, J.G.; Harkins, T.; Kent, E.; Phung, C.; Khan, R.; Seay, E.; Keenan, B.T.; Dedhia, R.C. The Prevalence of Central Sleep Apnea in New Referrals to a Sleep Surgery Clinic. *Otolaryngology–Head and Neck Surgery* **2025**, *172*, 686–692, doi:10.1002/ohn.1083.

38. Zhang, J.; Wang, L.; Guo, H.-J.; Wang, Y.; Cao, J.; Chen, B.-Y. Treatment-Emergent Central Sleep Apnea: A Unique Sleep-Disordered Breathing. *Chin. Med. J. (Engl).* **2020**, *133*, 2721–2730, doi:10.1097/CM9.0000000000001125.

39. American Academy of Sleep Medicine and others Sleep-Related Breathing Disorders in Adults: Recommendations for Syndrome Definition and Measurement Techniques in Clinical Research. The Report of an American Academy of Sleep Medicine Task Force. *Sleep* **1999**, *22*, 667–689.

40. Berry, R.B.; Abreu, A.R.; Krishnan, V.; Quan, S.F.; Strollo, P.J.; Malhotra, R.K. A Transition to the American Academy of Sleep Medicine–Recommended Hypopnea Definition in Adults: Initiatives of the Hypopnea Scoring Rule Task Force. *Journal of Clinical Sleep Medicine* **2022**, *18*, 1419–1425, doi:10.5664/jcsm.9952.

41. Malhotra, A.; Ayappa, I.; Ayas, N.; Collop, N.; Kirsch, D.; Mcardle, N.; Mehra, R.; Pack, A.I.; Punjabi, N.; White, D.P.; et al. Metrics of Sleep Apnea Severity: Beyond the Apnea-Hypopnea Index. *Sleep* **2021**, *44*, doi:10.1093/sleep/zsab030.

42. Younes, M. Role of Respiratory Control Mechanisms in the Pathogenesis of Obstructive Sleep Disorders. *J. Appl. Physiol.* **2008**, *105*, 1389–1405, doi:10.1152/japplphysiol.90408.2008.

43. Eckert, D.J.; Malhotra, A. Pathophysiology of Adult Obstructive Sleep Apnea. *Proc. Am. Thorac. Soc.* **2008**, *5*, 144–153, doi:10.1513/pats.200707-114MG.

44. Hartfield, P.J.; Janczy, J.; Sharma, A.; Newsome, H.A.; Sparapani, R.A.; Rhee, J.S.; Woodson, B.T.; Garcia, G.J.M. Anatomical Determinants of Upper Airway Collapsibility in Obstructive Sleep Apnea: A Systematic Review and Meta-Analysis. *Sleep Med. Rev.* **2023**, *68*, 101741, doi:10.1016/j.smrv.2022.101741.

45. Remmers, J.E.; deGroot, W.J.; Sauerland, E.K.; Anch, A.M. Pathogenesis of Upper Airway Occlusion during Sleep. *J. Appl. Physiol.* **1978**, *44*, 931–938, doi:10.1152/jappl.1978.44.6.931.

46. Betts JG; DeSaix P; Johnson E; et al *Anatomy and Physiology, 2nd Ed*; OpenStax: Houston, TX, 2022;

47. Qureshi, A.; Ballard, R.D.; Nelson, H.S. Obstructive Sleep Apnea. *Journal of Allergy and Clinical Immunology* **2003**, *112*, 643–651, doi:10.1016/j.jaci.2003.08.031.

48. McNicholas, W.T.; Korkalainen, H. Translation of Obstructive Sleep Apnea Pathophysiology and Phenotypes to Personalized Treatment: A Narrative Review. *Front. Neurol.* **2023**, *14*, doi:10.3389/fneur.2023.1239016.

49. Tserenpil, G.; Gebre, M.; Zergham, A.S.; Sekhon, A.K.; Malik, B.H. Managements for Obstructive Sleep Apnea in Adults: Review. *Cureus* **2020**, doi:10.7759/cureus.9905.

50. Ogbu, I.; Menon, T.; Chahil, V.; Kahlon, A.; Devanand, D.; Kalra, D.K. Sleep Disordered Breathing and Neurocognitive Disorders. *J. Clin. Med.* **2024**, *13*, 5001, doi:10.3390/jcm13175001.

51. Yeghiazarians, Y.; Jneid, H.; Tietjens, J.R.; Redline, S.; Brown, D.L.; El-Sherif, N.; Mehra, R.; Bozkurt, B.; Ndumele, C.E.; Somers, V.K. Obstructive Sleep Apnea and Cardiovascular Disease: A Scientific Statement From the American Heart Association. *Circulation* **2021**, *144*, doi:10.1161/CIR.0000000000000988.

52. Benjafield, A. V; Ayas, N.T.; Eastwood, P.R.; Heinzer, R.; Ip, M.S.M.; Morrell, M.J.; Nunez, C.M.; Patel, S.R.; Penzel, T.; Pépin, J.-L.; et al. Estimation of the Global Prevalence and Burden of Obstructive Sleep Apnoea: A Literature-Based Analysis. *Lancet Respir. Med.* **2019**, *7*, 687–698, doi:10.1016/S2213-2600(19)30198-5.

53. Rizzo, D.; Baltzan, M.; Sirpal, S.; Dosman, J.; Kaminska, M.; Chung, F. Prevalence and Regional Distribution of Obstructive Sleep Apnea in Canada: Analysis from the Canadian Longitudinal Study on Aging. *Canadian Journal of Public Health* **2024**, *115*, 970–979, doi:10.17269/s41997-024-00911-8.

54. Pendharkar, S.R.; Sharpe, H.; Rosychuk, R.J.; Laratta, C.R.; Fong, A.; Duan, Q.M.; Ronksley, P.E.; MacLean, J.E. Temporal and Regional Trends in Obstructive Sleep Apnea Using Administrative Health Data in Alberta, Canada. *Ann. Am. Thorac. Soc.* **2023**, *20*, 433–439, doi:10.1513/AnnalsATS.202209-789OC.

55. Leppänen, T.; Kulkas, A.; Duce, B.; Mervaala, E.; Töyräs, J. Severity of Individual Obstruction Events Is Gender Dependent in Sleep Apnea. *Sleep Breath.* **2017**, *21*, 397–404, doi:10.1007/s11325-016-1430-0.

56. Kapsimalis, F.; Kryger, M.H. Gender and Obstructive Sleep Apnea Syndrome, Part 1: Clinical Features. *Sleep* **2002**, *25*, 412–419.

57. Bonsignore, M.R.; Saaresranta, T.; Riha, R.L. Sex Differences in Obstructive Sleep Apnoea. *Eur. Respir. Rev.* **2019**, *28*, doi:10.1183/16000617.0030-2019.

58. Votteler, S.; Knaack, L.; Janicki, J.; Fink, G.R.; Burghaus, L. Sex Differences in Polysomnographic Findings in Patients with Obstructive Sleep Apnea. *Sleep Med.* **2023**, *101*, 429–436, doi:10.1016/j.sleep.2022.11.025.

59. Lin, C.M.; Davidson, T.M.; Ancoli-Israel, S. Gender Differences in Obstructive Sleep Apnea and Treatment Implications. *Sleep Med. Rev.* **2008**, *12*, 481–496, doi:10.1016/j.smrv.2007.11.003.

60. Castelnuovo, A.; Marelli, S.; Mazzeo, S.; Casoni, F.; Proserpio, P.; Oldani, A.; Bombaci, A.; Bortolin, E.; Bruschi, G.; Agosta, F.; et al. Gender Differences in Obstructive Sleep Apnea: A Preliminary Clinical and Polysomnographic Investigation. *Neurol. Int.* **2025**, *17*, 85, doi:10.3390/neurolint17060085.

61. Ernst, G.; Mariani, J.; Blanco, M.; Finn, B.; Salvado, A.; Borsini, E. Increase in the Frequency of Obstructive Sleep Apnea in Elderly People. *Sleep Sci.* **2019**, *12*, 222–226, doi:10.5935/1984-0063.20190081.

62. Osorio, R.S.; Martínez-García, M.Á.; Rapoport, D.M. Sleep Apnoea in the Elderly: A Great Challenge for the Future. *European Respiratory Journal* **2022**, *59*, 2101649, doi:10.1183/13993003.01649-2021.

63. Thompson, C.; Legault, J.; Moullec, G.; Baltzan, M.; Cross, N.; Dang-Vu, T.T.; Martineau-Dussault, M.-È.; Hanly, P.; Ayas, N.; Lorrain, D.; et al. A Portrait of Obstructive Sleep Apnea Risk Factors in 27,210 Middle-Aged and Older Adults in the Canadian Longitudinal Study on Aging. *Sci. Rep.* **2022**, *12*, 5127, doi:10.1038/s41598-022-08164-6.

64. Young, T. Predictors of Sleep-Disordered Breathing in Community-Dwelling Adults<Subtitle>The Sleep Heart Health Study</Subtitle> *Arch. Intern. Med.* **2002**, *162*, 893, doi:10.1001/archinte.162.8.893.

65. Johansson, P.; Alehagen, U.; Svanborg, E.; Dahlström, U.; Broström, A. Sleep Disordered Breathing in an Elderly Community-Living Population: Relationship to Cardiac Function, Insomnia Symptoms and Daytime Sleepiness. *Sleep Med.* **2009**, *10*, 1005–1011, doi:10.1016/j.sleep.2009.01.011.

66. Antonaglia, C.; Fabozzi, A.; Steffanina, A.; Ture, R.; Palange, P.; Confalonieri, M. Walking the Fine Line between OSA and Aging. *Sleep and Breathing* **2025**, *29*, 195, doi:10.1007/s11325-025-03343-x.

67. Young, T.; Peppard, P.E.; Gottlieb, D.J. Epidemiology of Obstructive Sleep Apnea: A Population Health Perspective. *Am. J. Respir. Crit. Care Med.* **2002**, *165*, 1217–1239, doi:10.1164/rccm.2109080.

68. Tai, J.E.; Phillips, C.L.; Yee, B.J.; Grunstein, R.R. Obstructive Sleep Apnoea in Obesity: A Review. *Clin. Obes.* **2024**, *14*, doi:10.1111/cob.12651.

69. Peppard, P.E. Longitudinal Study of Moderate Weight Change and Sleep-Disordered Breathing. *JAMA* **2000**, *284*, 3015, doi:10.1001/jama.284.23.3015.

70. Malhotra, A.; Heilmann, C.R.; Banerjee, K.K.; Dunn, J.P.; Bunck, M.C.; Bednarik, J. Weight Reduction and the Impact on Apnea-Hypopnea Index: A Systematic Meta-Analysis. *Sleep Med.* **2024**, *121*, 26–31, doi:10.1016/j.sleep.2024.06.014.

71. Veasey, S.C.; Guilleminault, C.; Strohl, K.P.; Sanders, M.H.; Ballard, R.D.; Magalang, U.J. Medical Therapy for Obstructive Sleep Apnea: A Review by the Medical Therapy for Obstructive Sleep Apnea Task Force of the Standards of Practice Committee of the American Academy of Sleep Medicine. *Sleep* **2006**, *29*, 1036–1044, doi:10.1093/sleep/29.8.1036.

72. Messineo, L.; Bakker, J.P.; Cronin, J.; Yee, J.; White, D.P. Obstructive Sleep Apnea and Obesity: A Review of Epidemiology, Pathophysiology and the Effect of Weight-Loss Treatments. *Sleep Med. Rev.* **2024**, *78*, 101996, doi:10.1016/j.smrv.2024.101996.

73. Kashyap, R.; Hock, L.M.; Bowman, T.J. Higher Prevalence of Smoking in Patients Diagnosed as Having Obstructive Sleep Apnea. *Sleep And Breathing* **2001**, *05*, 167–172, doi:10.1055/s-2001-18805.

74. Najafabadi, V.B.; Moussavi, Z. The Impact of Smoking Status on Obstructive Sleep Apnea: Insights from Anthropometric and Physiological Covariates*. In Proceedings of the 2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); IEEE, July 14 2025; pp. 1–5.

75. Boussoffara, L.; Boudawara, N.; Sakka, M.; Knani, J. Smoking Habits and Severity of Obstructive Sleep Apnea Hypopnea Syndrome. *Rev. Mal. Respir.* **2013**, *30*, 38–43, doi:10.1016/j.rmr.2012.08.009.

76. Yosunkaya, S.; Kutlu, R.; Vatansev, H. Effects of Smokıng on Patıents Wıth Obstructıve Sleep Apnea Syndrome. *Clin. Respir. J.* **2021**, *15*, 147–153, doi:10.1111/crj.13278.

77. Hsu, W.-Y.; Chiu, N.-Y.; Chang, C.-C.; Chang, T.-G.; Lane, H.-Y. The Association between Cigarette Smoking and Obstructive Sleep Apnea. *Tob. Induc. Dis.* **2019**, *17*, 27, doi:10.18332/tid/105893.

78. Esen, A.D.; Akpinar, M. Relevance of Obstructive Sleep Apnea and Smoking: Obstructive Sleep Apnea and Smoking. *Fam. Pract.* **2021**, *38*, 180–185, doi:10.1093/fampra/cmaa112.

79. O'Dell, K. Predictors of Difficult Intubation and the Otolaryngology Perioperative Consult. *Anesthesiol. Clin.* **2015**, *33*, 279–290, doi:10.1016/j.anclin.2015.02.002.

80. Kumar, H.V.M.; Schroeder, J.W.; Gang, Z.; Sheldon, S.H. Mallampati Score and Pediatric Obstructive Sleep Apnea. *Journal of Clinical Sleep Medicine* **2014**, *10*, 985–990, doi:10.5664/jcsm.4032.

81. Nuckton, T.J.; Glidden, D. V; Browner, W.S.; Claman, D.M. Physical Examination: Mallampati Score as an Independent Predictor of Obstructive Sleep Apnea. *Sleep* **2006**, *29*, 903–908, doi:10.1093/sleep/29.7.903.

82. Chi, L.; Comyn, F.-L.; Mitra, N.; Reilly, M.P.; Wan, F.; Maislin, G.; Chmiewski, L.; Thorne-FitzGerald, M.D.; Victor, U.N.; Pack, A.I.; et al. Identification of Craniofacial Risk Factors for Obstructive Sleep Apnoea Using Three-Dimensional MRI. *European Respiratory Journal* **2011**, *38*, 348–358, doi:10.1183/09031936.00119210.

83. Mihaicuta, S.; Udrescu, L.; Udrescu, M.; Toth, I.-A.; Topîrceanu, A.; Pleavă, R.; Ardelean, C. Analyzing Neck Circumference as an Indicator of CPAP Treatment Response in

Obstructive Sleep Apnea with Network Medicine. *Diagnostics* **2021**, *11*, 86, doi:10.3390/diagnostics11010086.

84. Min, H.J.; Park, A.Y.; Kim, D.H.; Lee, J.-G.; Park, S.; Cho, H.-J. Neck Circumference and Lowest Oxygen Saturation Are Independently Associated with High Coexistence of Hypertension in Obstructive Sleep Apnea. *Yonsei Med. J.* **2014**, *55*, 1310, doi:10.3349/ymj.2014.55.5.1310.

85. Yenigün, M. Neck Circumference. Metabolic Syndrome and Obstructive Sleep Apnea Syndrome; Evaluation of Possible Linkage. *Medical Science Monitor* **2013**, *19*, 111–117, doi:10.12659/MSM.883776.

86. Markun, L.C.; Sampat, A. Clinician-Focused Overview and Developments in Polysomnography. *Curr. Sleep Med. Rep.* **2020**, *6*, 309–321, doi:10.1007/s40675-020-00197-5.

87. Togni, C.; Carpinelli, S.; Valko, P.O.; Bockisch, C.; Waldvogel, D.; Werth, E.; Weber, K.P.; Valko, Y. Rapid Eye Movements during REM Sleep Differentiate PSP from Parkinson's Disease. *Mov. Disord. Clin. Pract.* **2024**, *11*, 1281–1285, doi:10.1002/mdc3.14187.

88. Walters, A.S. Assessment of Periodic Leg Movements Is an Essential Component of an Overnight Sleep Study. *Am. J. Respir. Crit. Care Med.* **2001**, *164*, 1339–1340, doi:10.1164/ajrccm.164.8.2107127a.

89. Penzel, T.; Kantelhardt, J.W.; Bartsch, R.P.; Riedl, M.; Kraemer, J.F.; Wessel, N.; Garcia, C.; Glos, M.; Fietze, I.; Schöbel, C. Modulations of Heart Rate, ECG, and Cardio-Respiratory Coupling Observed in Polysomnography. *Front. Physiol.* **2016**, *7*, doi:10.3389/fphys.2016.00460.

90. Muzumdar, H.; Arens, R. Physiological Effects of Obstructive Sleep Apnea Syndrome in Childhood. *Respir. Physiol. Neurobiol.* **2013**, *188*, 370–382, doi:10.1016/j.resp.2013.05.006.

91. Garde, A.; Dehkordi, P.; Karlen, W.; Wensley, D.; Ansermino, J.M.; Dumont, G.A. Development of a Screening Tool for Sleep Disordered Breathing in Children Using the Phone Oximeter™. *PLoS One* **2014**, *9*, e112959, doi:10.1371/journal.pone.0112959.

92. Berry, R.B.; Brooks, R.; Gamaldo, C.; Harding, S.M.; Lloyd, R.M.; Quan, S.F.; Troester, M.T.; Vaughn, B. V. AASM Scoring Manual Updates for 2017 (Version 2.4). *Journal of Clinical Sleep Medicine* **2017**, *13*, 665–666, doi:10.5664/jcsm.6576.

93. Hossain, J.L.; Shapiro, C.M. The Prevalence, Cost Implications, and Management of Sleep Disorders: An Overview. *Sleep and Breathing* **2002**, *6*, 85–102, doi:10.1007/s11325-002-0085-1.

94. Flemons, W.W.; Douglas, N.J.; Kuna, S.T.; Rodenstein, D.O.; Wheatley, J. Access to Diagnosis and Treatment of Patients with Suspected Sleep Apnea. *Am. J. Respir. Crit. Care Med.* **2004**, *169*, 668–672, doi:10.1164/rccm.200308-1124PP.

95. Byun, J.-H.; Kim, K.T.; Moon, H.; Motamedi, G.K.; Cho, Y.W. The First Night Effect during Polysomnography, and Patients' Estimates of Sleep Quality. *Psychiatry Res.* **2019**, *274*, 27–29, doi:10.1016/j.psychres.2019.02.011.

96. Riedel, B.W.; Winfield, C.F.; Lichstein, K.L. First Night Effect and Reverse First Night Effect in Older Adults with Primary Insomnia: Does Anxiety Play a Role? *Sleep Med.* **2001**, *2*, 125–133, doi:10.1016/S1389-9457(00)00054-X.

97. Steinberg, R.; Spector, A.R.; McVeigh, T.; Fudim, M. Home Sleep Apnoea Testing: Advances, Challenges and Considerations in Heart Failure. *Card. Fail. Rev.* **2025**, *11*, doi:10.15420/cfr.2025.29.

98. Chung, F.; Abdullah, H.R.; Liao, P. STOP-Bang Questionnaire: A Practical Approach to Screen for Obstructive Sleep Apnea. *Chest* **2016**, *149*, 631–638, doi:10.1378/chest.15-0903.

99. Cho, T.; Yan, E.; Chung, F. The STOP-Bang Questionnaire: A Narrative Review on Its Utilization in Different Populations and Settings. *Sleep Med. Rev.* **2024**, *78*, 102007, doi:10.1016/j.smrv.2024.102007.

100. Johns, M.W. A New Method for Measuring Daytime Sleepiness: The Epworth Sleepiness Scale. *Sleep* **1991**, *14*, 540–545, doi:10.1093/sleep/14.6.540.

101. Onen, F.; Moreau, T.; Gooneratne, N.S.; Petit, C.; Falissard, B.; Onen, S.H. Limits of the Epworth Sleepiness Scale in Older Adults. *Sleep and Breathing* **2013**, *17*, 343–350, doi:10.1007/s11325-012-0700-8.

102. Netzer, N.C.; Stoohs, R.A.; Netzer, C.M.; Clark, K.; Strohl, K.P. Using the Berlin Questionnaire To Identify Patients at Risk for the Sleep Apnea Syndrome. *Ann. Intern. Med.* **1999**, *131*, 485–491, doi:10.7326/0003-4819-131-7-199910050-00002.

103. Scharf, M.T. Reliability and Efficacy of the Epworth Sleepiness Scale: Is There Still a Place for It? *Nat. Sci. Sleep* **2022**, *Volume 14*, 2151–2156, doi:10.2147/NSS.S340950.

104. Georgakopoulou, V.; Pantazis, N.; Tsiafaki, X.; Nena, E.; Amfilochiou, A.; Steiropoulos, P. Validation of NoSAS Score for the Screening of Obstructive Sleep Apnea. *Medicine International* **2023**, *3*, 14, doi:10.3892/mi.2023.74.

105. Peng, M.; Chen, R.; Cheng, J.; Li, J.; Liu, W.; Hong, C. Application Value of the NoSAS Score for Screening Sleep-Disordered Breathing. *J. Thorac. Dis.* **2018**, *10*, 4774–4781, doi:10.21037/jtd.2018.07.46.

106. Khan, A.; Biswas, S.Kr.; Chunka, C.; Kumar Das, A. A Machine Learning Model for Obstructive Sleep Apnea Detection Using Ensemble Learning and Single-Lead EEG Signal Data. *IEEE Sens. J.* **2024**, *24*, 20266–20273, doi:10.1109/JSEN.2024.3397395.

107. Nassehi, F.; Eken, A.; Atalay, N.B.; Firat, H.; Eroğul, O. Machine Learning Based Severity Classification of Obstructive Sleep Apnea Patients Using Awake EEG. *Biomed. Signal Process. Control* **2024**, *96*, 106566, doi:10.1016/j.bspc.2024.106566.

108. Amen, S.; Rasool, B.; Al Lami, B.S.; Gamal, C.; Mohammad, A.N.; Maaroof, P.; Abdullah, R.M.; Subedi, R.; Al-Lami, R. Obstructive Sleep Apnea and Cardiovascular Diseases: A Systematic Review and Meta-Analysis of Prospective Studies. *Cureus* **2024**, doi:10.7759/cureus.71752.

109. Hemrajani, P.; Dhaka, V.S.; Rani, G.; Verma, S.; Kavita; Woźniak, M.; Shafi, J.; Ijaz, M.F. Integrating Physiological Signals for Enhanced Sleep Apnea Diagnosis with SleepNet. *Sci. Rep.* **2025**, *15*, 31715, doi:10.1038/s41598-025-16154-7.

110. Attar, E.T. Detailed Evaluation of Sleep Apnea Using Heart Rate Variability: A Machine Learning and Statistical Method Using ECG Data. *Front. Neurol.* **2025**, *16*, doi:10.3389/fneur.2025.1636983.

111. Sheta, A.; Turabieh, H.; Thaher, T.; Too, J.; Mafarja, M.; Hossain, M.S.; Surani, S.R. Diagnosis of Obstructive Sleep Apnea from ECG Signals Using Machine Learning and Deep Learning Classifiers. *Applied Sciences* **2021**, *11*, 6622, doi:10.3390/app11146622.

112. DAVIDSON, T.; SEDGH, J.; TRAN, D.; STEPNOWSKYJR, C. The Anatomic Basis for the Acquisition of Speech and Obstructive Sleep Apnea: Evidence from Cephalometric Analysis Supports The Great Leap Forward Hypothesis. *Sleep Med.* **2005**, *6*, 497–505, doi:10.1016/j.sleep.2005.03.007.

113. Goldshtein, E.; Tarasiuk, A.; Zigel, Y. Automatic Detection of Obstructive Sleep Apnea Using Speech Signals. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 1373–1382, doi:10.1109/TBME.2010.2100096.

114. TaghiBeyglou, B.; Geng, J.; McDaulid, D.; Gnaneswaran, P.; Yasokaran, O.; Chow, A.; Ng, R.; Chervin, R.D.; Brown, D.L.; Yadollahi, A. Self-Supervised Speech Representations for Sleep Apnea Severity Prediction. *IEEE Trans. Biomed. Eng.* **2026**, 1–11, doi:10.1109/TBME.2026.3656326.

115. Cao, S.; Rosenzweig, I.; Bilotta, F.; Jiang, H.; Xia, M. Automatic Detection of Obstructive Sleep Apnea Based on Speech or Snoring Sounds: A Narrative Review. *J. Thorac. Dis.* **2024**, *16*, 2654–2667, doi:10.21037/jtd-24-310.

116. Nakano, H.; Furukawa, T.; Tanigawa, T. Tracheal Sound Analysis Using a Deep Neural Network to Detect Sleep Apnea. *Journal of Clinical Sleep Medicine* **2019**, *15*, 1125–1133, doi:10.5664/jcsm.7804.

117. Darquea, B.; Vizcaíno-Imacaña, P.; Criollo, S.; Inga-Ipiales, A.; Zhinin-Vera, L.; Ramírez-Cando, L.; Cadena-Morejón, C.; Almeida-Galárraga, D.; Tirado-Espín, A.; Meneses, F.V.

Detection of Obstructive Sleep Apnea Using Deep Learning on Audio Signals. In; 2026; pp. 178–189.

118. Montazeri Ghahjaverestan, N.; Aguiar, C.; Hummel, R.; Cao, X.; Yu, J.; Bradley, T.D. Sleep Apnea Detection by Tracheal Motion and Sound, and Oximetry via Application of Deep Neural Networks. *Nat. Sci. Sleep* **2023**, *Volume 15*, 423–432, doi:10.2147/NSS.S397196.

119. Chen, H.; Aarab, G.; de Ruiter, M.H.T.; de Lange, J.; Lobbezoo, F.; van der Stelt, P.F. Three-Dimensional Imaging of the Upper Airway Anatomy in Obstructive Sleep Apnea: A Systematic Review. *Sleep Med.* **2016**, *21*, 19–27, doi:10.1016/j.sleep.2016.01.022.

120. Ciscar, M.A.; Juan, G.; Martínez, V.; Ramón, M.; Lloret, T.; Mínguez, J.; Armengot, M.; Marín, J.; Basterra, J. Magnetic Resonance Imaging of the Pharynx in OSA Patients and Healthy Subjects. *European Respiratory Journal* **2001**, *17*, 79–86, doi:10.1183/09031936.01.17100790.

121. Schwab, R.J.; Gupta, K.B.; Gefter, W.B.; Metzger, L.J.; Hoffman, E.A.; Pack, A.I. 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.* **1995**, *152*, 1673–1689, doi:10.1164/ajrccm.152.5.7582313.

122. Ashraf, W.; Fredberg, J.J.; Moussavi, Z. Anatomical Determinants of Tracheal Breathing Sounds: A Computational Study of Airway Narrowing and Obstructive Sleep Apnea. *Diagnostics* **2025**, *15*, 3108, doi:10.3390/diagnostics15243108.

123. Elwali, A.; Moussavi, Z. Determining Breathing Sound Features Representative of Obstructive Sleep Apnea During Wakefulness with Least Sensitivity to Other Risk Factors. *J. Med. Biol. Eng.* **2019**, *39*, 230–237, doi:10.1007/s40846-018-0410-1.

124. Sola-Soler, J.; Fiz, J.A.; Torres, A.; Jane, R. Identification of Obstructive Sleep Apnea Patients from Tracheal Breath Sound Analysis during Wakefulness in Polysomnographic Studies. In Proceedings of the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; IEEE, August 2014; pp. 4232–4235.

125. PASTERKAMP, H.; KRAMAN, S.S.; WODICKA, G.R. Respiratory Sounds: Advances beyond the Stethoscope. *Am. J. Respir. Crit. Care Med.* **1997**, *156*, 974–987, doi:10.1164/ajrccm.156.3.9701115.

126. Harper, P.; Kraman, S.S.; Pasterkamp, H.; Wodicka, G.R. An Acoustic Model of the Respiratory Tract. *IEEE Trans. Biomed. Eng.* **2001**, *48*, 543–550, doi:10.1109/10.918593.

127. Hanchuan Peng; Fuhui Long; Ding, C. Feature Selection Based on Mutual Information Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238, doi:10.1109/TPAMI.2005.159.

128. Finke, H.; Drews, A.; Engel, C.; Koos, B. Craniofacial Risk Factors for Obstructive Sleep Apnea—Systematic Review and Meta-analysis. *J. Sleep Res.* **2024**, *33*, doi:10.1111/jsr.14004.

129. Rahman, S.; Rahman, M.M.; Bhatt, S.; Sundararajan, R.; Faezipour, M. NeuroNet-AD: A Multimodal Deep Learning Framework for Multiclass Alzheimer's Disease Diagnosis. *Bioengineering* **2025**, *12*, 1107, doi:10.3390/bioengineering12101107.

130. Fang, Z.; Wang, Y.; Peng, L.; Hong, H. A Comparative Study of Heterogeneous Ensemble-Learning Techniques for Landslide Susceptibility Mapping. *International Journal of Geographical Information Science* **2021**, *35*, 321–347, doi:10.1080/13658816.2020.1808897.

131. Jing, H.; Wang, Y.; Wang, L.; Choi, S.; Xi, Z.; Cui, X. Analyzing the Inspiratory Airflow Unsteadiness in the Respiratory Tract Using Dynamic Mode Decomposition Method. *Physics of Fluids* **2024**, *36*, doi:10.1063/5.0230603.

132. Jing, H.; Ge, H.; Tang, H.; Farnoud, A.; Saidul Islam, M.; Wang, L.; Wang, C.; Cui, X. Assessing Airflow Unsteadiness in the Human Respiratory Tract under Different Expiration Conditions. *J. Biomech.* **2024**, *162*, 111910, doi:10.1016/j.jbiomech.2023.111910.

133. Stanescu, D.; Kostianev, S.; Sanna, A.; Liistro, G.; Veriter, C. Expiratory Flow Limitation during Sleep in Heavy Snorers and Obstructive Sleep Apnoea Patients. *European Respiratory Journal* **1996**, *9*, 2116–2121, doi:10.1183/09031936.96.09102116.

134. Sanchez, I.; Pasterkamp, H. Tracheal Sound Spectra Depend on Body Height. *American Review of Respiratory Disease* **1993**, *148*, 1083–1087, doi:10.1164/ajrccm/148.4_Pt_1.1083.

135. Mohsenin, V. Effects of Gender on Upper Airway Collapsibility and Severity of Obstructive Sleep Apnea. *Sleep Med.* **2003**, *4*, 523–529, doi:10.1016/S1389-9457(03)00168-0.

136. Brown, I.G.; Zamel, N.; Hoffstein, V. Pharyngeal Cross-Sectional Area in Normal Men and Women. *J. Appl. Physiol.* **1986**, *61*, 890–895, doi:10.1152/jappl.1986.61.3.890.

137. Malhotra, A.; Huang, Y.; Fogel, R.B.; Pillar, G.; Edwards, J.K.; Kikinis, R.; Loring, S.H.; White, D.P. The Male Predisposition to Pharyngeal Collapse: Importance of Airway Length. *Am. J. Respir. Crit. Care Med.* **2002**, *166*, 1388–1395, doi:10.1164/rccm.2112072.

138. Wimms, A.; Woehrle, H.; Ketheeswaran, S.; Ramanan, D.; Armitstead, J. Obstructive Sleep Apnea in Women: Specific Issues and Interventions. *Biomed Res. Int.* **2016**, *2016*, 1–9, doi:10.1155/2016/1764837.

139. Regensteiner, J.G.; Woodard, W.D.; Hagerman, D.D.; Weil, J. V.; Pickett, C.K.; Bender, P.R.; Moore, L.G. Combined Effects of Female Hormones and Metabolic Rate on Ventilatory Drives in Women. *J. Appl. Physiol.* **1989**, *66*, 808–813, doi:10.1152/jappl.1989.66.2.808.

140. Popovic, R.M.; White, D.P. Upper Airway Muscle Activity in Normal Women: Influence of Hormonal Status. *J. Appl. Physiol.* **1998**, *84*, 1055–1062, doi:10.1152/jappl.1998.84.3.1055.

141. Schwartz, A.R.; Patil, S.P.; Squier, S.; Schneider, H.; Kirkness, J.P.; Smith, P.L. Obesity and Upper Airway Control during Sleep. *J. Appl. Physiol.* **2010**, *108*, 430–435, doi:10.1152/japplphysiol.00919.2009.

142. Shelton, K.E.; Woodson, H.; Gay, S.; Suratt, P.M. Pharyngeal Fat in Obstructive Sleep Apnea. *American Review of Respiratory Disease* **1993**, *148*, 462–466, doi:10.1164/ajrccm/148.2.462.

143. Gupta, A.; Kumar, R.; Bhattacharya, D.; Thukral, B.; Suri, J. Craniofacial and Upper Airway Profile Assessment in North Indian Patients with Obstructive Sleep Apnea. *Lung India* **2019**, *36*, 94, doi:10.4103/lungindia.lungindia_303_18.

144. Mittman, C.; Edelman, N.H.; Norris, A.H.; Shock, N.W. Relationship between Chest Wall and Pulmonary Compliance and Age. *J. Appl. Physiol.* **1965**, *20*, 1211–1216, doi:10.1152/jappl.1965.20.6.1211.

145. Verbanck, S.; Schuermans, D.; Paiva, M.; Meysman, M.; Vincken, W. Small Airway Function Improvement after Smoking Cessation in Smokers without Airway Obstruction. *Am. J. Respir. Crit. Care Med.* **2006**, *174*, 853–857, doi:10.1164/rccm.200603-422OC.

146. Alqudah, A.M.; Moussavi, Z. Awake Insights for Obstructive Sleep Apnea: Severity Detection Using Tracheal Breathing Sounds and Meta-Model Analysis. *Diagnostics* **2026**, *16*, 448, doi:10.3390/diagnostics16030448.

147. Pramono, R.X.A.; Bowyer, S.; Rodriguez-Villegas, E. Automatic Adventitious Respiratory Sound Analysis: A Systematic Review. *PLoS One* **2017**, *12*, e0177926, doi:10.1371/journal.pone.0177926.

148. Monaco, A.; Amoroso, N.; Bellantuono, L.; Pantaleo, E.; Tangaro, S.; Bellotti, R. Multi-Time-Scale Features for Accurate Respiratory Sound Classification. *Applied Sciences* **2020**, *10*, 8606, doi:10.3390/app10238606.

149. Alqudah, A.M.; Ashraf, W.; Lithgow, B.; Moussavi, Z. Interpretable Acoustic Features from Wakefulness Tracheal Breathing for OSA Severity Assessment. *J. Clin. Med.* **2026**, *15*, 1081, doi:10.3390/jcm15031081.

150. Bailer, M.; Sprügel, M.I.; Stein, E.M.; Utz, J.; Mestermann, S.; Spitzer, P.; Kornhuber, J. Diagnostic Accuracy of Screening Questionnaires for Obstructive Sleep Apnea in Psychiatric Patients. *J. Psychiatr. Res.* **2025**, *186*, 280–288, doi:10.1016/j.jpsychires.2025.04.026.

151. Schwartz, A.R.; Smith, P.L.; Wise, R.A.; Bankman, I.; Permutt, S. Effect of Positive Nasal Pressure on Upper Airway Pressure-Flow Relationships. *J. Appl. Physiol.* **1989**, *66*, 1626–1634, doi:10.1152/jappl.1989.66.4.1626.

152. Opsahl, U.L.; Berge, M.; Lehmann, S.; Bjorvatn, B.; Opsahl, P.; Johansson, A. Acoustic Pharyngometry - A New Method to Facilitate Oral Appliance Therapy. *J. Oral Rehabil.* **2021**, *48*, 601–613, doi:10.1111/joor.13134.

153. Vana, K.D.; Silva, G.E.; Carreon, J.D.; Quan, S.F. Using Anthropometric Measures to Screen for Obstructive Sleep Apnea in the Sleep Heart Health Study Cohort. *Journal of Clinical Sleep Medicine* **2021**, *17*, 1635–1643, doi:10.5664/jcsm.9268.

154. Acet Ozturk, N.A.; Gorek Dilektasli, A.; Demirdogen Cetinoglu, E.; Ursavas, A.; Karadag, M. Diagnostic Accuracy of a Modified STOP-BANG Questionnaire with National Anthropometric Obesity Indexes. *Turk. Thorac. J.* **2019**, *20*, 103–107, doi:10.5152/TurkThoracJ.2018.18074.

155. Kirkness, J.P.; Schwartz, A.R.; Schneider, H.; Punjabi, N.M.; Maly, J.J.; Laffan, A.M.; McGinley, B.M.; Magnuson, T.; Schweitzer, M.; Smith, P.L.; et al. Contribution of Male Sex, Age, and Obesity to Mechanical Instability of the Upper Airway during Sleep. *J. Appl. Physiol.* **2008**, *104*, 1618–1624, doi:10.1152/japplphysiol.00045.2008.

156. Schwartz, A.R.; Patil, S.P.; Laffan, A.M.; Polotsky, V.; Schneider, H.; Smith, P.L. Obesity and Obstructive Sleep Apnea: Pathogenic Mechanisms and Therapeutic Approaches. *Proc. Am. Thorac. Soc.* **2008**, *5*, 185–192, doi:10.1513/pats.200708-137MG.

157. SUTHERLAND, K.; LEE, R.W.W.; CISTULLI, P.A. Obesity and Craniofacial Structure as Risk Factors for Obstructive Sleep Apnoea: Impact of Ethnicity. *Respirology* **2012**, *17*, 213–222, doi:10.1111/j.1440-1843.2011.02082.x.

158. Dempsey, J.A.; Veasey, S.C.; Morgan, B.J.; O'Donnell, C.P. Pathophysiology of Sleep Apnea. *Physiol. Rev.* **2010**, *90*, 47–112, doi:10.1152/physrev.00043.2008.

159. Antonaglia, C.; Passuti, G. Obstructive Sleep Apnea Syndrome in Non-Obese Patients. *Sleep and Breathing* **2022**, *26*, 513–518, doi:10.1007/s11325-021-02412-1.

160. Kimoff, R.J.; Hamid, Q.; Divangahi, M.; Hussain, S.; Bao, W.; Naor, N.; Payne, R.J.; Ariyarajah, A.; Mulrain, K.; Petrof, B.J. Increased Upper Airway Cytokines and Oxidative Stress in Severe Obstructive Sleep Apnoea. *European Respiratory Journal* **2011**, *38*, 89–97, doi:10.1183/09031936.00048610.

161. Ostertagová, E.; Ostertag, O.; Kováč, J. Methodology and Application of the Kruskal-Wallis Test. *Applied Mechanics and Materials* **2014**, *611*, 115–120, doi:10.4028/www.scientific.net/AMM.611.115.

162. Ferreira, J.A.; Zwinderman, A.H. On the Benjamini–Hochberg Method. *The Annals of Statistics* **2006**, *34*, doi:10.1214/009053606000000425.

163. Thissen, D.; Steinberg, L.; Kuang, D. Quick and Easy Implementation of the Benjamini-Hochberg Procedure for Controlling the False Positive Rate in Multiple Comparisons. *Journal of Educational and Behavioral Statistics* **2002**, *27*, 77–83, doi:10.3102/10769986027001077.

164. Engel, J.; Huele, A.F. A Generalized Linear Modeling Approach to Robust Design. *Technometrics* **1996**, *38*, 365–373, doi:10.1080/00401706.1996.10484548.