Machine Learning (ML) Models to Enhance the Berlin Questionnaire (BQ) Detection of Obstructive Sleep Apnea (OSA) at-Risk Patients

Luana Conte , Giorgio De Nunzio * , Francesco Giombi , Roberto Lupo , Caterina Arigliani , Federico Leone , Fabrizio Salamanca , Cosimo Petrelli , Paola Angelelli , Luigi De Benedetto , Michele Arigliani

Posted Date: 17 June 2024

doi: 10.20944/preprints202406.1155.v1

Keywords: Obstructive Sleep Apnea; OSA; berlin questionnaire; Machine Learning; artificial intelligence

 doi:10.20944/preprints202406.1155.v1

Article

Machine Learning (ML) Models to Enhance the Berlin Questionnaire (BQ) Detection of Obstructive Sleep Apnea (OSA) at-Risk Patients

Luana Conte [1,2], Giorgio De Nunzio [1,2,*], Francesco Giombi [3,4], Roberto Lupo [5], Caterina Arigliani [6], Federico Leone [7], Fabrizio Salamanca [3,7], Cosimo Petrelli [8], Paola Angelelli [9], Luigi De Benedetto [10] and Michele Arigliani [11]

[1] Laboratory of Biomedical Physics and Environment, Department of Mathematics and Physics "E. De Giorgi", University of Salento, Via per Monteroni, 73100, Lecce, Italy
[2] Laboratory of Advanced Data Analysis for Medicine (ADAM) at the Laboratory of Interdisciplinary Research Applied to Medicine (DReAM), University of Salento and Local Health Authority (ASL) Lecce, Piazza Filippo Muratore, 73100 Lecce, Italy
[3] Department of Biomedical Sciences, Humanitas University, Via Rita Levi Montalcini 4, 20090 Pieve Emanuele – Milan, Italy
[4] Otorhinolaryngology Unit, IRCCS Humanitas Research Hospital, Via Manzoni 56, 20089 Rozzano – Milan, Italy
[5] Unit of Admitting and Emergency Medicine and Surgery, "San Giuseppe da Copertino" Hospital, Local Health Authority (ASL) Lecce, Via Carmiano, 73043 Copertino – Lecce, Italy
[6] Unit of Anesthesia, Fondazione Policlinico Universitario Campus Bio-Medico, Via Alvaro del Portillo, 00128 Rome, Italy
[7] Otorhinolaryngology Unit, Snoring & OSA Research Center, "Humanitas San Pio X" Hospital, Via Francesco Nava 31, 20159 Milan, Italy
[8] Unit of Internal Medicine, "San Giuseppe da Copertino" Hospital, Local Health Authority (ASL) Lecce, Via Carmiano, 73043 Copertino – Lecce, Italy
[9] Department of Experimental Medicine, College ISUFI –Ecotekne, Via per Monteroni s.n., 73100 - Lecce, Italy
[10] Unit of Integrated Therapies in Otolaryngology, Fondazione Policlinico Universitario Campus Bio-Medico, Via Alvaro del Portillo, 00128 Rome, Italy
[11] Unit of Otorhinolaryngology, "Vito Fazzi" Hospital, Local Health Authority (ASL) Lecce, Piazza Filippo Muratore, 73100 Lecce, Italy
* Correspondence: Giorgio De Nunzio, giorgio.denunzio@unisalento.it

Abstract: Objective: With just ten questions, the Berlin questionnaire (BQ) stands out as one of the simplest and most widely implemented non-invasive screening tools for detecting subjects at high risk for Obstructive Sleep Apnea (OSA), a still underdiagnosed syndrome characterized by partial or complete obstruction of the upper airways during sleep. The main aim of this study was to enhance the diagnostic accuracy of the BQ through Machine Learning (ML) techniques. **Methods**: A ML classifier (hereafter, ML-10) was trained using the ten questions of the standard BQ. A simplified variant of the BQ, BQ-2, which comprises only two questions out of the total of ten, was also assessed in a ML context. A 10-fold cross validation scheme was used. Ground truth was provided by the Apnea-Hypopnea Index (AHI) measured by Home Sleep Apnea Testing. Model performance was determined comparing ML-10 and BQ-2 with the standard BQ by the Receiver Operating Characteristic Curve (ROC), Area Under the Curve (AUC), sensitivity, specificity, and accuracy. **Results**: ML-10 demonstrated superior performance in predicting the risk for OSA compared to the standard BQ and was also capable in classifying OSA with two different AHI thresholds (AHI>=15, AHI>=30), typically used in clinical practice. Remarkably, BQ-2 was also better in sensibility to assess moderate to severe OSA (AHI≥ 15) compared to the ML-10. **Conclusions**: The study underscores the importance of integrating ML techniques for early OSA detection, suggesting

 doi:10.20944/preprints202406.1155.v1

a direction for future research to improve diagnostic processes and patient outcomes in sleep medicine.

Keywords: Obstructive Sleep Apnea; OSA; berlin questionnaire; Machine Learning; artificial intelligence

1. Introduction

Obstructive Sleep Apnea (OSA) is a syndrome characterized by partial or complete obstruction of the upper airways during sleep. This blockage leads to frequent awakenings to reopen the airway, disrupting sleep, causing excessive daytime sleepiness, and triggering a stress response in the body. The obstruction can also result in lowered blood oxygen levels during sleep [1], increased carbon dioxide levels, and potential damage to the cardiovascular system. OSA is also linked to a variety of health issues including stroke, high blood pressure, and even death [2–7]. These health problems are especially pronounced in individuals who are overweight and vary based on gender and age.

The occurrence of OSA, estimated between 9% to 38% of the Italian population, varies widely with a higher likelihood in older adults, men, and those who are obese [1,8,9]. Among older individuals, its prevalence may rise up to 84% [1]. Despite an increase in research and medical attention towards OSA in recent years, it remains a condition that is frequently not diagnosed. This underdiagnosis can be attributed to the lack of biomarkers for identifying the disease [10–14].

In 2019, the CERGAS (Research Center on Health and Social Care Management at the Bocconi University) released data estimating the annual costs associated with OSA in Italy at approximately 31 billion euros. On average, the cost for each patient with severe OSA was calculated to be around 3,850 euros. Despite having an estimated 12 million people with moderate to severe OSA, only about 460,000 individuals in Italy have been formally diagnosed, and merely half of these diagnosed patients have received treatment. This situation places Italy at the bottom among major countries in terms of the number of individuals diagnosed with OSA [15]. Considering that each patient is diagnosed many years after the onset of the disease, the direct and indirect healthcare costs determine a significant burden for the National Health System (NHS), which affects every single citizen. Prevention and early diagnosis are the only ways to achieve improved quality of life and cost containment [10,16].

For the diagnosis of OSA, polysomnography (PSG) is considered the gold standard, and the severity of OSA is typically measured using the apnea-hypopnea index (AHI), with thresholds set at ≥5/h for OSA diagnosis, ≥15/h for moderate to severe OSA, and ≥30/h for severe OSA [1]. However, this method is expensive [10] and requires the patient to be monitored continuously by healthcare professionals [17], leading to a scarcity of available testing and, consequently, delays in diagnosis and an increase in the burden of disease [18–20]. Therefore, Home Sleep Apnea Testing (HSAT) is often used as an alternative. HSAT offers several advantages over traditional PSG. One of the foremost benefits of HSAT is the convenience it provides; patients can undergo testing in the familiar and comfortable setting of their own home. This not only reduces the anxiety and discomfort often associated with spending a night in an unfamiliar sleep lab environment but also removes the logistical challenges of arranging for an overnight stay away from home. Furthermore, HSAT stands out for its cost-effectiveness. Generally costing less than laboratory-based PSG, it becomes a more accessible option for a broader range of patients, breaking down financial barriers to obtaining a diagnosis.

Recent advancements in software technologies and Machine Learning (ML) methods have significantly enhanced the development of effective predictive and diagnostic tools, becoming increasingly prevalent in various fields of medical research and applications, including for OSA [12,13,21–28]. The prediction models described in existing research primarily utilize clinical data, such as demographic information (age and gender), comorbid conditions, anthropometric measures (Body Mass Index (BMI), waist and neck circumferences), symptoms of OSA, and physiological

 doi:10.20944/preprints202406.1155.v1

parameters (blood pressure, overnight pulse oximetry, and lung function tests). The effectiveness of these models in predicting OSA, as indicated by an AHI ≥5/h, has shown sensitivity rates ranging from 66% to 100% and specificity rates from 30.8% to 76.2%. For predicting more severe OSA (AHI ≥15/h), sensitivity ranges from 60.3% to 92.7%, with specificity between 33.3% and 90.7% [25]. The variability in these models ability to discriminate between cases may be due to factors such as the complexity of the models, sample size, OSA prevalence, and the proportion of cases with different severities of OSA. It is noted that most OSA prediction models prioritize higher sensitivity over specificity to facilitate early diagnosis, although this approach may result in a higher rate of false positives and potentially lead to unnecessary PSG testing [25].

The Berlin questionnaire (BQ) [29] stands out as one of the simplest and most widely implemented non-invasive screening tools for diagnosing OSA, demonstrating a sensitivity of 86% and a specificity of 95% for OSA diagnosis. Originally introduced in the United States (US), the BQ consists of a concise set of questions focused on risk factors and symptoms associated with OSA, aimed at identifying patients at high risk who might benefit from undergoing PSG to facilitate increased diagnosis rates. While the standard BQ comprises 10 questions, we previously introduced a streamlined version by using a trained classifier, known as the simplified Berlin questionnaire (BQ_2) [23], which reduces the questionnaire to just two questions. This abbreviated version has been shown to achieve results comparable to the original BQ, offering an efficient means of screening high-risk OSA patients.

The main aim of this research was to enhance the sensitivity, specificity, and accuracy of the conventional BQ by incorporating ML techniques. For this purpose, we developed a ML-enhanced BQ model (ML-10) capable of predicting the risk of OSA using the same items as the BQ. Additionally, we explored a simplified version of ML-10, called BQ-2 [23], which is based on the BQ but only preserves two of the original BQ items, to determine whether it yields comparable results. The predictive performance of these models was evaluated against the conventional BQ approach, which does not incorporate ML techniques. Furthermore, we utilized the ML-10 and the BQ-2 models to identify patients with OSA at two different AHI thresholds: ≥15/h, and ≥30/h, thereby assessing their efficacy across a spectrum of OSA severity.

In conclusion, the integration of a ML algorithm into the conventional BQ demonstrated a significant enhancement in the ability to predict the risk of OSA and also across various severity thresholds. This advancement underscores the potential of ML-enhanced diagnostic tools in improving early detection of OSA. The findings of this research validate the application of innovative ML approaches in enhancing the diagnostic processes for OSA, potentially leading to more timely and effective interventions for this widely prevalent but underdiagnosed condition.

2. Participants and Methods

2.1. Design

From January to December 2023, an observational multicenter study was conducted across two Italian hospitals: the Otorhinolaryngology Unit at the "Vito Fazzi" Hospital in Lecce and the Otorhinolaryngology Head & Neck Surgery Unit at the IRCCS Humanitas Research Hospital in Milan. A total of 462 subjects, comprising 112 from Lecce and 350 from Milan, were screened due to suspected symptoms of OSA and underwent a HSAT.

2.2. Participants

The inclusion criteria for this study were: (1) participants aged≥18 years and (2) who had undergone a HSAT recording. Before the HSAT examination, a baseline screening questionnaire was used to assess each participant's basic information, medication, and surgical history. The participants were measured for height, weight, and BMI (Kg/m^2) [29] at the time of registration.

2.3. OSA Diagnosis

 doi:10.20944/preprints202406.1155.v1

All the sleep-related signals were obtained using a HSAT device (Embletta Gold Portable Testing Device®, RemLogicE® Software 2015, Embla System Inc, Broomfield, US, used in Lecce, and the Embletta® Multi Parameter Recorder-Polygraph (MPR-PG), RemLogicE® 3.4.1, Embla Systems, Kanata, Ontario, Canada, used in Milan). This study adhered to the guidelines set forth by the American Academy of Sleep Medicine (AASM)[30,31].

2.4. The Berlin Questionnaire and the Simplified Berlin Questionnaire

The BQ [29] is structured into three categories that assess the risk of sleep apnea. Based on their responses to individual items and their cumulative scores within these symptom categories, patients are classified as either high-risk or low-risk for OSA. Category 1, comprising five items, focuses on snoring behaviors. Category 2, with three items, investigates daytime somnolence. Category 3 consists of a single item that evaluates the presence of hypertension. A positive score in the first two categories requires frequent symptom occurrence, defined as more than 3-4 times per week. In contrast, a positive score in the third category results from either a history of hypertension or a BMI greater than 30Kg/m^2 [29]. The overall assessment is based on the collective responses across these categories, with patients categorized as high-risk for OSA if they have positive scores in two or more categories; otherwise, they are deemed low-risk [29].

Our prior research showed that, out of the ten questions included in the standard BQ, two were sufficient to closely approximate the BQ output using a trained classifier. Further details are available in [23]. In summary, the first critical question assesses high blood pressure, asking, "Do you have high blood pressure?" This inquiry is followed by one of two options regarding fatigue: "How often do you feel tired or fatigued after your sleep?" or "During your waking time, do you feel tired, fatigued or not up to par?" These questions are designed to be selected independently yet provide insightful data for OSA risk assessment. Despite their independence, we arbitrarily opted to utilize the first fatigue-related question ("How often do you feel tired or fatigued after your sleep?"). This decision was indeed derived by the observation that both models yielded comparable results when applied independently, suggesting no significant advantage in the context of our study to favor one fatigue-related question over the other.

2.5. Statistical Analysis

The baseline characteristics and BQ items for all participants, encompassing both patients with confirmed OSA and those without, were subjected to descriptive statistical analysis. Continuous variables were summarized using the mean and standard deviation (SD), whereas categorical variables were described using frequencies and percentages. To explore associations between two categorical variables, Fisher's exact test was employed. Similarly, the Mann-Whitney U-test was utilized to assess the statistical significance of differences between the distributions of two continuous variables among participants categorized on the basis of their AHI values, specifically those who are not at risk of OSA (AHI < 5) and those who are (AHI ≥ 5), according to the threshold "implemented" in the BQ [29]. A p-value of less than 0.05 was deemed to indicate statistical significance. The scoring of the BQ and all statistical analyses, including evaluations of both qualitative and quantitative variables, were performed using the Matlab software, version 2023b.

2.6. Machine Learning Predictive Value

Calculating group statistics plays a crucial role in establishing the statistical relevance of variables within a diagnostic context, allowing for the assessment of risk factors and relationships with comorbidities. However, it is widely acknowledged that relevance does not equate to discriminant power, which is more critical for classification and prediction tasks. Variables that are statistically significant in a model do not necessarily guarantee superior prediction performance, and attributes deemed non-significant might prove to be predictive. Consequently, we chose to explore the BQ from the perspective of its predictive capabilities using ML techniques. To this end, six distinct classifiers were evaluated for their suitability in the predictive task: Naive Bayes, Support Vector

 doi:10.20944/preprints202406.1155.v1

Machine (SVM), Decision Trees, Error-correcting Output Codes (ECOC), Discriminant Analysis, Ensemble of decision trees, and Artificial Neural Networks (ANN). Among these, the Ensemble of decision trees demonstrated the best performance. This model underwent training first with the ten responses from the standard BQ and then with the two responses from the simplified version, BQ_2, independently, resulting in the development and testing of two distinct models.

A 10-fold cross-validation (CV) approach was used for training and quality assessment. For both models, the features were normalized to a 0-1 range using min-max normalization on the training dataset at each CV iteration, with the same normalization parameters applied to the corresponding test set.

The Receiver Operating Characteristic (ROC) curve was used to illustrate the diagnostic capability of the models at various decision thresholds, providing a graphical representation of the trade-off between sensitivity (true positive rate) and 1-specificity (false positive rate). We first identified the specific operating point on the ROC curve corresponding to the conventional BQ, which indicates the combination of sensitivity (the ability to correctly identify cases at high risk of OSA) and specificity (the ability to correctly identify low or non-OSA cases) achieved without the integration of ML techniques. Then, we compared this point with the performance of the ML enhanced models (both ML-10 and BQ_2) at equal specificity and at equal sensitivity, by moving (from the BQ point) vertically and horizontally respectively, until intersecting the ROC curve of the ML-10 model. This way we evaluated how ML-10 could improve sensitivity while maintaining the specificity of the conventional BQ, and vice versa.

Subsequently, we extended our analysis to evaluate the ML-10 and BQ-2 models across two different AHI thresholds. For this purpose, the ROC curve was utilized to assess the classifier performance and to determine an "optimal" prediction threshold that yields the highest accuracy. From this optimal operating point, binary classifiers were derived. Performance metrics including the Area Under the Curve (AUC), accuracy, sensitivity, and specificity were employed to measure the models effectiveness. All computational analyses were conducted using MATLAB software, version R2023b.

Inizio modulo

2.7. Ethical Considerations

The experimental protocol received approval from the Bioethics Committee of the Local Health Authority of Lecce (Protocol Number 74, dated 22/04/2022) and from Milan (Protocol Number CET Lombardia 5-PIO X-153 /23, dated 19/09/2023). Conducted in full compliance with the Helsinki Declaration for Human Research, this study ensured the ethical treatment and protection of all participants. Written informed consent was secured from each subject who agreed to partake in the study, underscoring our commitment to ethical research practices. The ethical considerations of the study were meticulously outlined in the questionnaire introduction, crafted in alignment with the principles set forth by the Italian Data Protection Authority (DPA). The participants were informed of their right to voluntary participation, with the explicit option to withdraw from the study at any point should they choose to. The process of obtaining informed consent was designed to reiterate the voluntary nature of participation, emphasizing the confidentiality and anonymity with which all collected information would be treated. This approach ensured that participants were fully aware of their rights and the study ethical standards, fostering an environment of trust and respect for individual autonomy.

3. Results

3.1. Sample Demographics

The baseline characteristics of the participants were analyzed. Overall, 460 subjects who had undergone HSAT recording, were enrolled in this study. Of these, 141 were women, 257 were over 60 years old and 310 (67%) had an AHI of ≥5 therefore considered positive. The median BMI was 27.38 Kg/m^2 (range 13-53 Kg/m^2).

 doi:10.20944/preprints202406.1155.v1

Clinical features were compared between patients with and without suspicion of OSA (cutoff AHI ≥ 5) and among the subgroups at three cutoffs (AHI≥10, AHI≥15, AHI≥30). The results are reported in Table 1. Compared to patients without OSA, those with suspect of OSA were older (<0.001***), more obese (<0.001***), sleepier (<0.001***), and had higher percentage of men.

Table 1. Baseline characteristics of the cohort. Data are expressed as mean ± standard deviation. The p-value represents the comparison between the AHI≥5 (positive) and AHI < 5 (negative) groups. A p-value <0.05 was considered statistically significant and labeled with one to three asterisks according to the p threshold (*p<0.05; **p<0.01; ***p<0.001).

	Negative (AHI≤5)	**Positive (AHI≥15)**	**Positive (AHI≥30)**	**p value** *(comparison between Positive vs Negative, cutoff AHI≥5)*
n	150	181	83	
Age >60 (n, %)	60 (40)	117 (64)	53 (63)	<0.001***
Female n (%)	52 (35)	39 (22)	11 (13)	0.19
Height (cm)	172.4 ± 9.8	174.1 ± 10.1	174.7 ± 8.06	0.94
Weight (kg)	77.6 ± 16.3	88.3 ± 16.9	93.4 ± 15.9	<0.001***
BMI (kg/m^2)	26.0 ± 4.6	29.2 ± 5.6	30.7 ± 5.6	<0.001***
HST				
AHI	2.5 ± 1.4	32.8 ± 15.7	45.7 ± 14.5	<0.001***
ODI	2.6 ± 1.8	31.3 ± 18.7	42.2 ± 14.8	<0.001***
LOS	92.6 ± 11.1	88.1 ± 10.3	85.5 ± 11.8	<0.001***
SO_2 mean (%)	89.8 ± 8.1	82.6 ± 7.7	81.1 ± 8.8	<0.001***

BMI = Body Mass Index; ODI = Oxygen Desaturation Index; LOS = Length of Stay.

3.2. Berlin Questionnaire Score and Metrics

The BQ was administered to all the participants, and the collected answers were analyzed. Before delving into the specifics of the BQ scores, we categorized the subjects into low versus high OSA risk groups based on the cutoff utilized in connection with the BQ, which considers an AHI of ≥5 as positive (high risk) [29]. Consequently, Table 2 compares the high versus low OSA risk groups as determined by this BQ cutoff. It is important to clarify that this initial categorization uses the 'ground truth' based on the AHI ≥5 threshold, rather than the metrics derived from the questionnaire itself. The latter will be examined subsequently to assess how well the BQ scores align with the established 'ground truth'. This approach allows for a direct comparison between the questionnaire categorization and the clinical benchmark, providing insight into the BQ's effectiveness in identifying patients at varying levels of OSA risk.

In our sample, the high-risk OSA group had a significantly larger proportion of respondents reporting frequent snoring compared to the low-risk group (p<0.001). The high-risk group also reported more breathing interruptions than the low-risk subjects (p<0.001). Fatigue, somnolence at awakening and during daytime are also symptoms significantly present in the high-risk group compared to the low-risk group (p <0.001). High blood pressure was also highly reported in subjects with high risk of OSA, and this difference was statistically significant (p<0.001). Following the administration of the BQ and the subsequent data collection, we calculated the BQ scores as prescribed by its guidelines. The outcomes of this analysis, including the accuracy, sensitivity, and specificity of the BQ, are detailed in Table 3. The ROC space is shown in Figure 1 where a red point indicates the BQ position. Notably, the classic BQ is positioned in the upper right corner of the evaluation plot, approaching the point (1,1) which represents maximum sensitivity and minimum specificity. This characteristic reflects the aim of the BQ to function as a high-sensitivity screening tool, intended to minimize false negatives even at the cost of accepting a higher number of false positives.

 doi:10.20944/preprints202406.1155.v1

Table 2. Differences between low vs high OSA risk groups. Statistical significance was determined by Mann-Whitney test (*p<0.05; **p<0.01; ***p<0.001).

Items	Z-value	Rank sum	p-value
Snoring category			
History of snoring	-3.31	32084	<0.001***
Very loud snoring	-2.30	31200	0.02
Snoring every night	-0.08	3.4004e+04	0.92
Bothersome snoring	-1.70	3.2403e+04	0.08
Interrupting night breathing	4.77	40315	<0.001***
Symptoms category			
Tired upon awakening	4.25	39622	<0.001***
Tired while daytime	3.29	38386	<0.001***
Dozing off while driving	-5.39	29418	<0.001***
Frequency of dozing off	4.33	37442	<0.001***
Hypertension category			
High blood pressure	-7.34	2.5476e+04	<0.001***

3.3. The ML-10 Model

To determine whether the ML-10 and the reduced version BQ-2 outperform the traditional BQ in predicting patients at high versus low risk of OSA, we conducted a comparative analysis using the same threshold used in the standard BQ (AHI≥5). By employing the same AHI≥5 threshold across all models, we ensure a consistent basis for comparison, enabling a clear understanding of the potential advantages offered by integrating ML techniques with the traditional BQ assessment. By maintaining the BQ level of specificity (through a vertical displacement in the ROC space), the ML-10 model showcased a remarkable specificity of 72%, significantly outperforming the BQ (53%). This improvement indicates the ML-10 model enhanced capability to correctly identify individuals without OSA, at fixed sensitivity, thereby reducing the incidence of false positives. Conversely, when aligning with the BQ sensitivity level (via a horizontal displacement), the ML-10 model demonstrated a sensitivity of 93%, a substantial increase from the BQ (82%). This indicates the ML-10 model has superior ability to accurately detect individuals at high risk of OSA, at fixed specificity, minimizing the risk of overlooking affected patients.

Additionally, the BQ-2 showed also improvements when compared to the conventional BQ, even if (as reasonable) its figures of merit are smaller than ML-10. Table 3 and Figure 1 presents metrics and the ROC curves comparing the three models.

Table 3. Comparing the performance of the standard BQ, the BQ enhanced through Machine Learning ML-10 and the simplified BQ enhanced through ML (BQ_2) using metrics such as AUC, Sensitivity, Specificity and Accuracy.

	BQ	ML-10	BQ-2
AUC	(not applicable)	86%	77%
Sensitivity	82%	93%	88%
Specificity	53%	73%	54%

AUC = Area Under the Curve.

 doi:10.20944/preprints202406.1155.v1

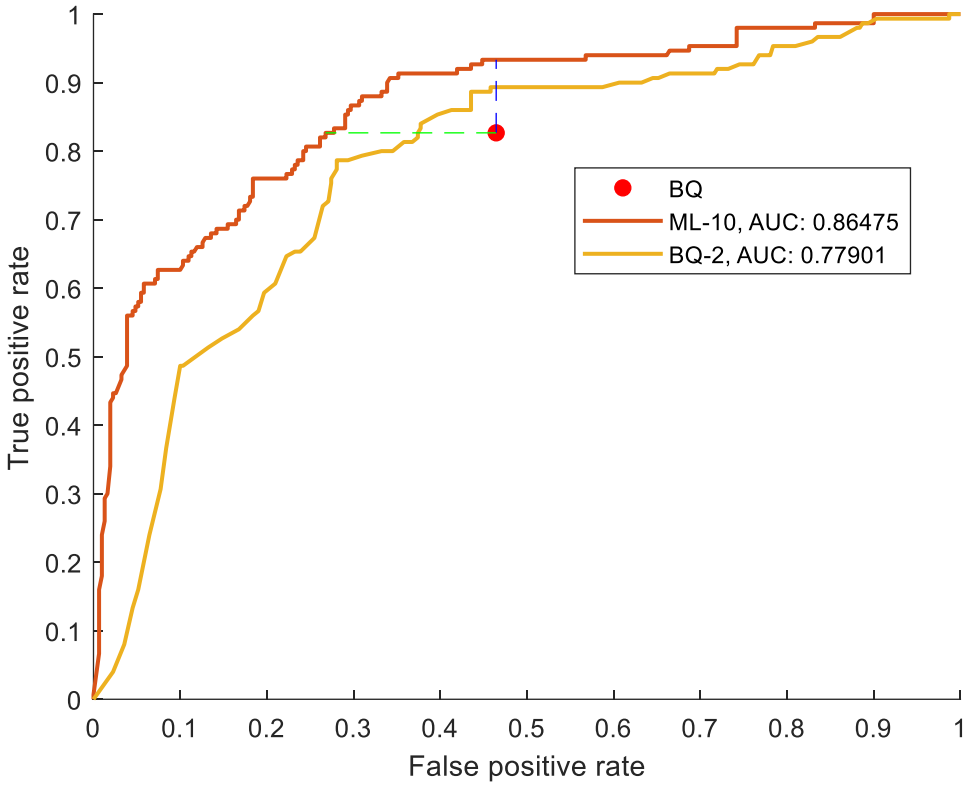

Figure 1. ROC comparison among the standard BQ, the BQ enhanced through Machine Learning (ML-10), and the simplified BQ enhanced through Machine Learning (BQ_2) [23]. AHI ≥5 was used as cutoff. Although some statistical software associates AUC values also to classifiers with binary output (when just one point exists in the ROC space), we preferred to neglect this feature and only drew the particular BQ working point (red point in the plot) which gives BW performance in terms of fixed sensitivity and specificity.

Recognizing the importance of a nuanced clinical evaluation, we expanded our analysis to investigate how well the ML-10 model and BQ-2 distinguish between patients across different levels of OSA severity. This step involved utilizing two AHI thresholds (AHI≥15, and AHI≥30) commonly referenced in the literature to categorize OSA severity as moderate and severe, respectively [1]. The outcomes of this comprehensive evaluation are presented in Figure 2 and Table 4. Before all we realized that AUCs for ML-10 (0.85 and 0.88) were (slightly) better than AUCs for BQ-2 (0.82 and 0.87), for AHI ≥ 15 and AHI ≥ 30 respectively . Then, by arbitrarily selecting as the optimal thresholds the ones that yield the highest accuracy, the performance of the ML-10 model consistently remained high across the two AHI thresholds and larger than that of BQ_2, except in one case because, compared to the ML-10 model, BQ-2 had higher sensitivity to assess moderate OSA with a cutoff of AHI≥15 (88% vs 70%).

Table 4. Metrics comparison among ML-10 model and BQ-2 model in assessing OSA severity with the two cutoffs (AHI≥15, AHI≥30).

	AHI ≥ 15	AHI ≥ 30
ML-10		
AUC	85%	88%
Sensitivity	70%	69%
Specificity	81%	93%
Accuracy	77%	89%

 doi:10.20944/preprints202406.1155.v1

BQ_2		
AUC	82%	87%
Sensitivity	88%	60%
Specificity	69%	92%
Accuracy	76%	86%

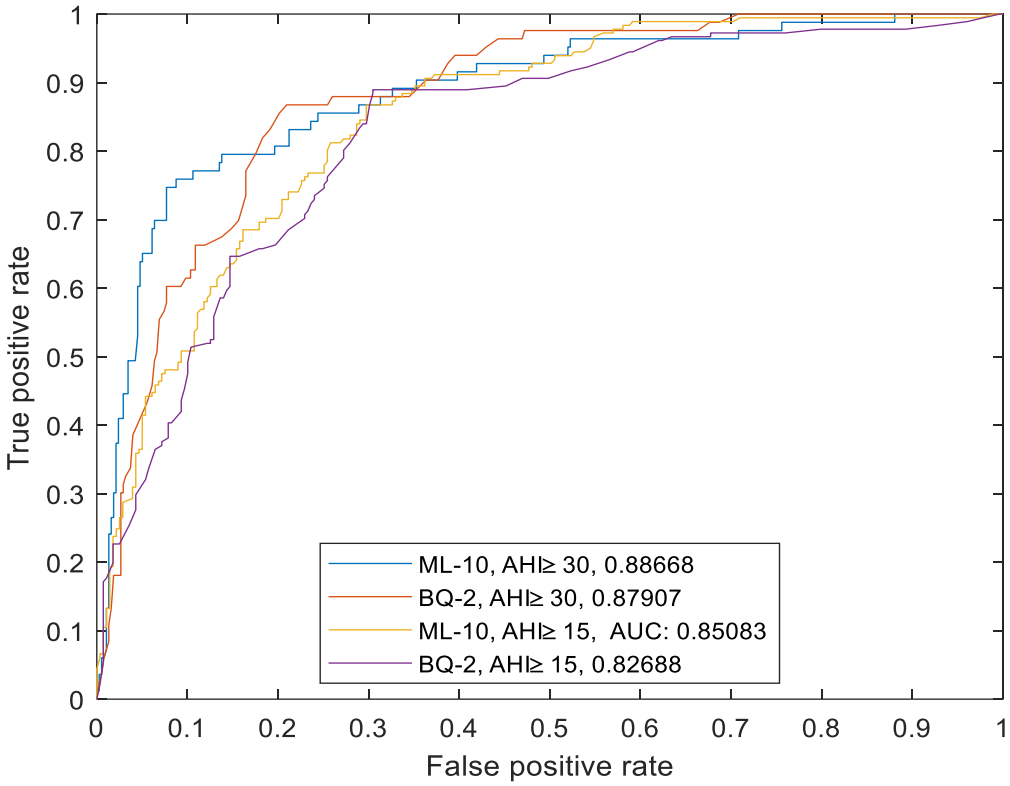

Figure 2. ROC curve comparison between the ML-10 and the BQ-2 models in assessing OSA severity at two cutoffs (AHI≥15, AHI≥30).

4. Discussion

OSA is increasingly recognized as a significant concern within global health and economic contexts, underlining the importance of its early detection and diagnosis in the realm of preventive medicine [1,18,32]. Prompt identification of OSA is essential for initiating timely interventions, which can mitigate a broad range of associated health risks and enhance patient outcomes. Given that the standard diagnostic test for OSA, such as in-laboratory PSG, is expensive and often subject to long wait times due to high demand, there is a clinical imperative to identify key factors and develop a simple, yet reliable tool for estimating OSA risk [18,19]. In general, BQ has expectedly high sensitivity, as this tool has been developed for identifying high-risk patients for OSA in primary care settings. Despite this advantage, the BQ low specificity and consequent high misclassification rate reveal its limited discriminatory capability, rendering its utility comparable to subjective clinical judgments [31,33]. In the quest for a straightforward questionnaire to ascertain OSA risk, clinicians demand enhancements to existing tools. Arunsurat et al. [34] posited that with certain modifications, the BQ could serve effectively as an OSA screening instrument. Furthermore, Stelmach-Mardas et al. [35] add to the growing body of evidence indicating the BQ inadequacy in distinguishing between high and low-risk patients, suggesting the need for the development of alternative protocols to heighten diagnostic precision for such individuals.

 doi:10.20944/preprints202406.1155.v1

In this research, we sought to advance the capabilities of the traditional BQ through the integration of ML techniques. Our research integrates ML models with the standard BQ to harness Artificial Intelligence capabilities for analyzing patterns and correlations in data that might not be immediately apparent to human evaluators. This method facilitates a more detailed assessment of risk factors, potentially identifying subtle signs of OSA risk overlooked by conventional approaches. To determine whether our ML-10 and the simplified two-item version BQ-2 outperform traditional BQ in predicting patients at high versus low risk of OSA, we conducted a comparative analysis using the established threshold used in the standard BQ (AHI≥5) and by comparing points on the ROC curves. The findings underscore the efficacy in terms of sensitivity and specificity of the ML-10 model when contrasted with conventional BQ. A sensitivity of 93% at the same specificity as conventional BQ indicates that the model can correctly identify 93% of individuals (at low or high risk of OSA), operating with the same TN-rate. This result is significant as it demonstrates that, while maintaining the same rate of false alarms (1 - Specificity), the ML-10 model is more effective in detecting OSA risk cases compared to conventional BQ. On the other hand, a specificity of 73% at the same sensitivity as conventional BQ emphasizes that the ML-10 model reduces the number of false positives (healthy individuals erroneously identified as at risk of OSA) compared to the conventional BQ, while still correctly detecting 82% of true positives. In this way, the ML-10 model shows excellent performance in identifying non-risk cases, surpassing conventional BQ.

These results indicate that the ML-10 model surpasses conventional BQ both in terms of sensitivity (when specificity is maintained) and specificity (when sensitivity is maintained). This implies that, depending on clinical or screening needs, the ML-10 model can be adjusted to optimize the ability to detect OSA risk cases (by maximizing sensitivity) or the ability to reduce false positives (by maximizing specificity), offering a more flexible and accurate approach in the diagnosis of OSA.

In the comparative evaluation between conventional BQ and the classifier based on its simplified version, the results indicate that BQ_2, despite the significant reduction in the number of questions to only two, slightly outperforms BQ in terms of sensitivity and specificity (fixing one of the two variables at the BQ value). Additionally, the use of BQ-2 offers the flexibility to adjust the operating point depending on the specific needs of clinical or screening applications, thus providing a potential advantage in terms of customizing the diagnostic approach.

After assessing the ML-10 and BQ-2 performance against traditional BQ using a single cutoff, we expanded our analysis to include two clinical-practice-relevant AHI cutoffs. This step involved utilizing two AHI thresholds (AHI≥15, and AHI≥30) commonly referenced in the literature to categorize OSA severity as moderate to severe, and severe, respectively [1]. The decision to employ these specific AHI thresholds is rooted in their widespread acceptance and use in clinical practice and research for defining the severity of OSA. Such a differentiated approach allows for a more detailed assessment of the models performance, providing insights into their predictive capabilities across a spectrum of OSA severity. This is particularly relevant for clinicians and healthcare providers seeking to tailor interventions and management strategies based on the severity of the condition. By choosing the optimal threshold for maximum accuracy, the ML-10 model performance consistently demonstrated its strength at both AHI thresholds.

Remarkably, BQ-2 exhibited even better outcomes in identifying moderate to severe risk for OSA cases (AHI≥15), achieving higher sensitivity compared to ML-10 at the ROC space point of maximum accuracy. These results highlight the potential for a more streamlined and efficient screening process. By examining whether a simplified model can retain or surpass the full BQ predictive accuracy, this study suggests the possibility of more accessible and less cumbersome OSA screening approaches. This is especially pertinent in primary care environments or areas with limited access to specialized sleep medicine services, where a rapid and dependable screening tool could significantly improve the early detection of individuals at risk for OSA. However, we should take into account that using only two questions likely makes the test sensitive but not specific, as various diseases could present with the same broad symptoms.

The present study is subject to several limitations that merit consideration. Firstly, the participant cohort was drawn exclusively from two hospitals in Italy, limiting the data set

 doi:10.20944/preprints202406.1155.v1

representativeness of the broader population. Consequently, the predictive model developed herein might not possess widespread generalizability, potentially limiting its applicability to populations beyond the initial study setting or to diverse ethnic groups [25,36]. Secondly, this observational study did not account for undiagnosed medical conditions commonly associated with OSA, such as neurological, cardiovascular, and pulmonary disorders. The absence of these variables could impact the model predictive accuracy. Furthermore, our model lacked detailed anthropometric imaging or measurements, which might have restricted its capability to identify disease-specific causes of OSA accurately.

In light of these limitations, there is a clear need for further research to enhance the model robustness and applicability. To this end, we are planning a prospective clinical trial aimed at evaluating ML-10 and BQ-2 across a more representative sample of the general population. This forthcoming trial is expected to address the current study limitations by incorporating a broader range of demographic and clinical variables, thereby improving the model predictive performance and generalizability.

5. Conclusions

Given the substantial proportion of individuals still undiagnosed with OSA, coupled with the current absence of definitive diagnostic biomarkers for the condition, there is a pressing need for improved screening methodologies. The BQ, when enhanced with ML techniques, stands out as significant advancements in this regard. The study discovered that the ML-10 model was particularly effective in identifying individuals at risk of OSA with greater accuracy than the traditional BQ. By integrating ML techniques, we achieved a notable improvement in sensitivity and specificity, highlighting the potential of ML to refine diagnostic processes. This suggests that the ML-10 model can more effectively distinguish between high-risk and low-risk individuals, thereby reducing the likelihood of false positives and negatives. Furthermore, BQ_2, with its reduced question set, also showcased its utility by maintaining comparable diagnostic accuracy to the full BQ while offering a more streamlined and accessible screening tool. This adaptation could facilitate wider screening efforts, particularly in primary care settings or areas with limited access to sleep medicine specialists. Additionally, the flexibility of the classifier allows for adjustments across different operating points, enabling the selection of an optimal threshold that best balances sensitivity and specificity for the targeted population. This adaptability is crucial in tailoring the screening process to diverse clinical environments and patient needs, optimizing the early detection and management of OSA.

Moreover, the application of the ML-10 model extends beyond the commonly used AHI threshold of the standard BQ (AHI≥5), demonstrating its utility across other clinically relevant AHI thresholds, specifically ≥15 and ≥30, which are frequently referenced in literature to categorize the severity of OSA as moderate to severe, and severe, respectively. This versatility underscores the model capability to adapt to varying clinical requirements, offering a nuanced approach to diagnosing OSA across its spectrum. Such adaptability ensures that the ML-10 model is not only a tool for preliminary screening but also a significant asset in stratifying OSA severity, thus enhancing the precision of diagnostic decisions and subsequent management plans.

By leveraging these insights, healthcare professionals can better stratify individuals based on their risk levels, paving the way for more tailored diagnostic and management strategies for sleep apnea. The ML-10 embodies the potential to transform the approach to diagnosing OSA, offering a more individualized assessment of risk. Looking forward, the insights gained from this research could serve as a foundation for further innovations in the field, ultimately leading to earlier detection, improved patient outcomes, and a reduction in the healthcare burden associated with OSA. In the end, this study underscores the value of combining traditional clinical assessment tools with cutting-edge technology to address complex health challenges, marking a significant stride towards the future of personalized medicine in sleep health.

References

1. Senaratna, C. V.; Perret, J.L.; Lodge, C.J.; Lowe, A.J.; Campbell, B.E.; Matheson, M.C.; Hamilton, G.S.; Dharmage, S.C. Prevalence of Obstructive Sleep Apnea in the General Population: A Systematic Review. *Sleep Med. Rev.* **2017**, *34*, 70–81, doi:10.1016/j.smrv.2016.07.002.
2. Marin, J.M.; Carrizo, S.J.; Vicente, E.; Agusti, A.G. Long-Term Cardiovascular Outcomes in Men with Obstructive Sleep Apnoea-Hypopnoea with or without Treatment with Continuous Positive Airway Pressure: An Observational Study. *Lancet* **2005**, *365*, 1046–1053, doi:10.1016/s0140-6736(05)71141-7.
3. Dyken, M.E.; Im, K. Bin Obstructive Sleep Apnea and Stroke. *Chest* **2009**, *136*, doi:10.1378/chest.08-1512.
4. Toraldo, D.; Benedetto, M.; Conte, L.; Nuccio, F. Statins May Prevent Atherosclerotic Disease in OSA Patients without Co-Morbidities? *Curr. Vasc. Pharmacol.* **2016**, *15*, 5–9, doi:10.2174/1570161114666161007164112.
5. Garbarino, S.; Scoditti, E.; Lanteri, P.; Conte, L.; Magnavita, N.; Toraldo, D.M. Obstructive Sleep Apnea With or Without Excessive Daytime Sleepiness: Clinical and Experimental Data-Driven Phenotyping. *Front. Neurol.* **2018**, *9*, doi:10.3389/fneur.2018.00505.
6. Vicini, C.; Cannavicci, A.; Ciocciloni, E.; Meccariello, G.; Cammaroto, G.; Gobbi, R.; Sanna, A.; Toraldo, D.M.; Bonetti, G.A.; Passali, F.M.; et al. Treatment. In *Obstructive Sleep Apnea*; Springer International Publishing: Cham, 2023; pp. 85–104.
7. Toraldo, D.M.; de Benedetto, M.; Conte, L.; de Nuccio, F. Statins May Prevent Atherosclerotic Disease in OSA Patients without Co-Morbidities? *Curr. Vasc. Pharmacol.* **2017**, *15*, doi:10.2174/1570161114666161007164112.
8. 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.
9. Fietze, I.; Laharnar, N.; Obst, A.; Ewert, R.; Felix, S.B.; Garcia, C.; Gläser, S.; Glos, M.; Schmidt, C.O.; Stubbe, B.; et al. Prevalence and Association Analysis of Obstructive Sleep Apnea with Gender and Age Differences - Results of SHIP-Trend. *J. Sleep Res.* **2019**, *28*, e12770, doi:10.1111/jsr.12770.
10. Toraldo, D.M.; Passali, D.; Sanna, A.; De Nuccio, F.; Conte, L.; De Benedetto, M. Cost-Effectiveness Strategies in OSAS Management: A Short Review. *Acta Otorhinolaryngol. Ital.* **2017**, *37*, 447–453, doi:10.14639/0392-100X-1520.
11. Conte, L.; Greco, M.; Toraldo, D.M.; Arigliani, M.; Maffia, M.; De Benedetto, M. A Review of the "OMICS" for Management of Patients with Obstructive Sleep Apnoea. *Acta Otorhinolaryngol. Ital.* **2020**, *40*, 164–172, doi:10.14639/0392-100X-N0409.
12. Arigliani, M.; Toraldo, D.M.; Montevecchi, F.; Conte, L.; Galasso, L.; De Rosa, F.; Lattante, C.; Ciavolino, E.; Arigliani, C.; Palumbo, A.; et al. A New Technological Advancement of the Drug-Induced Sleep Endoscopy (Dise) Procedure: The "All in One Glance" Strategy. *Int. J. Environ. Res. Public Health* **2020**, *17*, 1–11, doi:10.3390/ijerph17124261.
13. Arigliani, M.; Toraldo, D.M.; Ciavolino, E.; Lattante, C.; Conte, L.; Arima, S.; Arigliani, C.; Palumbo, A.; De Benedetto, M. The Use of Middle Latency Auditory Evoked Potentials (MLAEP) as Methodology for Evaluating Sedation Level in Propofol-Drug Induced Sleep Endoscopy (DISE) Procedure. *Int. J. Environ. Res. Public Health* **2021**, *18*, 2070, doi:10.3390/ijerph18042070.
14. Arigliani, C.; Arigliani, M.; Ciavolino, E.; Conte, L.; Toraldo, D.M.; Passariello, S.; Arima, S.; Palumbo, A.; De Benedetto, M. Polygraphic Findings in Simplified Barbed Reposition Pharyngoplasty (BRP) as a Treatment for OSA Patients. *J. Interdiscip. Res. Appl. to Med.* **2021**, *5*, 19–26, doi:10.1285/i25327518v5i1p19.
15. Armeni, P.; Borsoi, L.; Costa, F.; Donin, G.; Gupta, A. Cost-of-Illness Study of Obstructive Sleep Apnea Syndrome (OSAS) in Italy. **2019**.
16. Toraldo, D.M.; Toraldo, S.; Conte, L. The Clinical Use of Stem Cell Research in Chronic Obstructive Pulmonary Disease: A Critical Analysis of Current Policies. *J. Clin. Med. Res.* **2018**, *10*, 671–678, doi:10.14740/jocmr3484w.
17. Gottlieb, D.J.; Punjabi, N.M. Diagnosis and Management of Obstructive Sleep Apnea. *JAMA* **2020**, *323*, 1389, doi:10.1001/jama.2020.3514.
18. Knauert, M.; Naik, S.; Gillespie, M.B.; Kryger, M. Clinical Consequences and Economic Costs of Untreated Obstructive Sleep Apnea Syndrome. *World J. Otorhinolaryngol. - Head Neck Surg.* **2015**, *1*, 17–27, doi:10.1016/j.wjorl.2015.08.001.

19. Stewart, S.A.; Skomro, R.; Reid, J.; Penz, E.; Fenton, M.; Gjevre, J.; Cotton, D. Improvement in Obstructive Sleep Apnea Diagnosis and Management Wait Times: A Retrospective Analysis of a Home Management Pathway for Obstructive Sleep Apnea. *Can. Respir. J.* **2015**, *22*, 167–170, doi:10.1155/2015/516580.
20. 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. *J. Clin. Sleep Med.* **2017**, *13*, 479–504, doi:10.5664/jcsm.6506.
21. Davenport, T.; Kalakota, R. The Potential for Artificial Intelligence in Healthcare. *Futur. Healthc. J.* **2019**, *6*, 94–98, doi:10.7861/futurehosp.6-2-94.
22. Yu, K.-H.; Beam, A.L.; Kohane, I.S. Artificial Intelligence in Healthcare. *Nat. Biomed. Eng.* **2018**, *2*, 719–731, doi:10.1038/s41551-018-0305-z.
23. De Nunzio, G.; Conte, L.; Lupo, R.; Vitale, E.; Calabrò, A.; Ercolani, M.; Carvello, M.; Arigliani, M.; Toraldo, D.M.; De Benedetto, L. A New Berlin Questionnaire Simplified by Machine Learning Techniques in a Population of Italian Healthcare Workers to Highlight the Suspicion of Obstructive Sleep Apnea. *Front. Med.* **2022**, *9*, doi:10.3389/fmed.2022.866822.
24. Kuan, Y.-C.; Hong, C.-T.; Chen, P.-C.; Liu, W.-T.; Chung, C.-C. Logistic Regression and Artificial Neural Network-Based Simple Predicting Models for Obstructive Sleep Apnea by Age, Sex, and Body Mass Index. *Math. Biosci. Eng.* **2022**, *19*, 11409–11421, doi:10.3934/mbe.2022532.
25. Huang, W.-C.; Lee, P.-L.; Liu, Y.-T.; Chiang, A.A.; Lai, F. Support Vector Machine Prediction of Obstructive Sleep Apnea in a Large-Scale Chinese Clinical Sample. *Sleep* **2020**, *43*, doi:10.1093/sleep/zsz295.
26. Kirby, S.D.; Eng, P.; Danter, W.; George, C.F.; Francovic, T.; Ruby, R.R.; Ferguson, K.A. Neural Network Prediction of Obstructive Sleep Apnea from Clinical Criteria. *Chest* **1999**, *116*, 409–415, doi:10.1378/chest.116.2.409.
27. Zerah-Lancner, F.; Lofaso, F.; D'Ortho, M.P.; Delclaux, C.; Goldenberg, F.; Coste, A.; Housset, B.; Harf, A. Predictive Value of Pulmonary Function Parameters for Sleep Apnea Syndrome. *Am. J. Respir. Crit. Care Med.* **2000**, *162*, 2208–2212, doi:10.1164/ajrccm.162.6.2002002.
28. Zou, J.; Guan, J.; Yi, H.; Meng, L.; Xiong, Y.; Tang, X.; Su, K.; Yin, S. An Effective Model for Screening Obstructive Sleep Apnea: A Large-Scale Diagnostic Study. *PLoS One* **2013**, *8*, e80704, doi:10.1371/journal.pone.0080704.
29. 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, doi:10.7326/0003-4819-131-7-199910050-00002.
30. Oku, Y.; Okada, M. Periodic Breathing and Dysphagia Associated with a Localized Lateral Medullary Infarction. *Respirology* **2008**, *13*, 608–610, doi:10.1111/j.1440-1843.2008.01267.x.
31. Sert Kuniyoshi, F.H.; Zellmer, M.R.; Calvin, A.D.; Lopez-Jimenez, F.; Albuquerque, F.N.; van der Walt, C.; Trombetta, I.C.; Caples, S.M.; Shamsuzzaman, A.S.; Bukartyk, J.; et al. Diagnostic Accuracy of the Berlin Questionnaire in Detecting Sleep-Disordered Breathing in Patients with a Recent Myocardial Infarction. *Chest* **2011**, *140*, 1192–1197, doi:10.1378/chest.10-2625.
32. Salman, L.A.; Shulman, R.; Cohen, J.B. Obstructive Sleep Apnea, Hypertension, and Cardiovascular Risk: Epidemiology, Pathophysiology, and Management. *Curr. Cardiol. Rep.* **2020**, *22*, 6, doi:10.1007/s11886-020-1257-y.
33. Cowan, D.C.; Allardice, G.; Macfarlane, D.; Ramsay, D.; Ambler, H.; Banham, S.; Livingston, E.; Carlin, C. Predicting Sleep Disordered Breathing in Outpatients with Suspected OSA. *BMJ Open* **2014**, *4*, e004519, doi:10.1136/bmjopen-2013-004519.
34. Arunsurat, I.; Luengyosluechakul, S.; Prateephoungrat, K.; Siripaupradist, P.; Khemtong, S.; Jamcharoensup, K.; Thanapatkaiporn, N.; Limpawattana, P.; Laohasiriwong, S.; Pinitsoontorn, S.; et al. Simplified Berlin Questionnaire for Screening of High Risk for Obstructive Sleep Apnea Among Thai Male Healthcare Workers. *J. UOEH* **2016**, *38*, 199–206, doi:10.7888/juoeh.38.199.

 doi:10.20944/preprints202406.1155.v1

35. Stelmach-Mardas, M.; Iqbal, K.; Mardas, M.; Kostrzewska, M.; Piorunek, T. Clinical Utility of Berlin Questionnaire in Comparison to Polysomnography in Patients with Obstructive Sleep Apnea. In; 2017; pp. 51–57.
36. Kim, Y.J.; Jeon, J.S.; Cho, S.-E.; Kim, K.G.; Kang, S.-G. Prediction Models for Obstructive Sleep Apnea in Korean Adults Using Machine Learning Techniques. *Diagnostics* **2021**, *11*, 612, doi:10.3390/diagnostics11040612.