

Article

Machine Learning Identification of Obstructive Sleep Apnea Severity through the Patient Clinical Features: A Retrospective Study

Antonino Maniaci ^{1,2,*}, Paolo Marco Riela ^{3,†}, Giannicola Iannella ^{2,4}, Jerome Rene Lechien ^{2,5,6}, Ignazio La Mantia ¹, Marco De Vincentiis ⁴, Giovanni Cammaroto ^{2,7}, Christian Calvo-Henriquez ⁸, Milena Di Luca ¹, Carlos Chiesa Estomba ⁹, Alberto Maria Saibene ¹⁰, Isabella Pollicina ¹, Giovanna Stilo ¹, Paola Di Mauro ¹, Angelo Cannavici ⁷, Rodolfo Lugo ¹¹, Giuseppe Magliulo ⁴, Antonio Greco ⁴, Annalisa Pace ⁴, Giuseppe Meccariello ⁷, Salvatore Cocuzza ¹ and Claudio Vicini ^{7,12}

- ¹ Department of Medical and Surgical Sciences and Advanced Technologies “GF Ingrassia” ENT Section, University of Catania, 95123 Catania, Italy
 - ² Sleep Surgery Study Group of the Young-Otolaryngologists of the International Federations of Oto-rhino-laryngological Societies (YO-IFOS), 75001 Paris, France
 - ³ Department of Mathematics and Informatics, University of Catania, 95123 Catania, Italy
 - ⁴ Otorhinolaryngology Department, Sapienza University of Rome, Policlinico Umberto I, Viale del Policlinico 151, 00010 Rome, Italy
 - ⁵ Department of Human Anatomy and Experimental Oncology, Faculty of Medicine, UMONS Research Institute for Health Sciences and Technology, University of Mons (UMons), 7000 Mons, Belgium
 - ⁶ Department of Otorhinolaryngology and Head and Neck Surgery, Foch Hospital, School of Medicine, UFR Simone Veil, Université Versailles Saint-Quentin-en-Yvelines (Paris Saclay University), 75001 Paris, France
 - ⁷ Department of Head-Neck Surgery, Otolaryngology, Head-Neck, and Oral Surgery Unit, Morgagni Pierantoni Hospital, 47121 Forlì, Italy
 - ⁸ Service of Otolaryngology, Rhinology Unit, Hospital Complex of Santiago de Compostela, 15701 Santiago de Compostela, Spain
 - ⁹ Department of Otorhinolaryngology-Head and Neck Surgery, Hospital Universitario Donostia, 20001 San Sebastian, Spain
 - ¹⁰ Otolaryngology Unit Santi Paolo e Carlo, Hospital Department of Health Sciences, Università Degli Studi di Milano, 20021 Milan, Italy
 - ¹¹ Department of Otorhinolaryngology, Grupo Medico San Pedro, Monterrey 64660, Mexico
 - ¹² Department ENT and Audiology, University of Ferrara, 44121 Ferrara, Italy
- * Correspondence: tnmaniaci29@gmail.com; Tel.: +39-3204154576
- † These authors contributed equally to this work.



Citation: Maniaci, A.; Riela, P.M.; Iannella, G.; Lechien, J.R.; La Mantia, I.; De Vincentiis, M.; Cammaroto, G.; Calvo-Henriquez, C.; Di Luca, M.; Estomba, C.C.; et al. Machine Learning Identification of Obstructive Sleep Apnea Severity through the Patient Clinical Features: A Retrospective Study. *Life* **2023**, *13*, 702. <https://doi.org/10.3390/life13030702>

Academic Editors: Binh P. Nguyen and Luca Steardo

Received: 4 December 2022

Revised: 25 February 2023

Accepted: 2 March 2023

Published: 5 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Objectives: To evaluate the role of clinical scores assessing the risk of disease severity in patients with clinical suspicion of obstructive sleep apnea syndrome (OSA). The hypothesis was tested by applying artificial intelligence (AI) to demonstrate its effectiveness in distinguishing between mild–moderate OSA and severe OSA risk. Methods: A support vector machine model (SVM) was developed from the samples included in the analysis (N = 498), and they were split into 75% for training (N = 373) with the remaining for testing (N = 125). Two diagnostic thresholds were selected for OSA severity: mild to moderate (apnea–hypopnea index (AHI) ≥ 5 events/h and AHI < 30 events/h) and severe (AHI ≥ 30 events/h). The algorithms were trained and tested to predict OSA patient severity. Results: The sensitivity and specificity for the SVM model were 0.93 and 0.80 with an accuracy of 0.86; instead, the logistic regression full mode reported a value of 0.74 and 0.63, respectively, with an accuracy of 0.68. After backward stepwise elimination for features selection, the reduced logistic regression model demonstrated a sensitivity and specificity of 0.79 and 0.56, respectively, and an accuracy of 0.67. Conclusion: Artificial intelligence could be applied to patients with symptoms related to OSA to identify individuals with a severe OSA risk with clinical-based algorithms in the OSA framework.

Keywords: obstructive sleep apnea; artificial intelligence; machine learning; OSA severity; ESS; clinical OSA scores

1. Introduction

Obstructive sleep apnea (OSA) is a respiratory disorder characterized by the partial or total collapse of the upper airways with intermittent hypoxia, a chronic systemic inflammatory state, and an increased cardiovascular risk [1]. OSA is also associated with comorbidities, such as metabolic syndrome and olfactory or neurodegenerative disorders. Although OSA has a very high prevalence worldwide, up to 50% of the general population, subjects are often unaware of the disorder, and the diagnosis is performed when the associated comorbidities have already developed [2].

The strong connection between the severity of obstructive apnea identified through standard diagnostic tests, such as nocturnal polysomnography or pulse oximetry, and the associated cardiovascular or neurodegenerative risks are established in the literature [3,4]. Conversely, the different clinical features and scoring systems, such as age, BMI, the anatomical scores of palatal collapses, or validated questionnaires, although significantly correlating with pathology, do not possess adequate sensitivity or specificity to substitute instrumental diagnostics tools [5].

Recently, Sutherland et al. analyzed in a patient-level meta-analysis the relationship between craniofacial morphology and weight loss with sleep apnea severity [6]. The authors demonstrated a correlation between weight and AHI changes ($r_s = 0.3$, $p = 0.002$) and an increased maxilla–mandible relationship angle related to AHI improvement (β [95% CI] -1.7 [-2.9 , -0.5], $p = 0.004$) at linear regression.

In this regard, clinical prediction decision-making algorithms that use clinical information, objective scores, and easy-to-execute subjective questionnaires could be useful strategies for predicting OSA [7].

Maranate et al. in 2015 proposed a prioritization process of clinical risk factors of the severity of OSA using the analytical hierarchy process to select the patients with the greatest need for an in-depth diagnostic study [8]. The authors developed an algorithm used on 1042 suspected OSA patients who had undergone diagnostic PSG study. Moreover, 42 variables of disease severity were identified, with an overall sensitivity/specificity of the model for severe, moderate, and mild of 92.32%/91.76%, 89.52%/88.18%, and 91.08%/84.58%, respectively.

Artificial intelligence (AI) is rapidly gaining importance in medicine, opening promising new perspectives [6–8]. Through a deep learning system, different types of data can be exploited, such as clinical scores, questionnaires, or diagnostics [2,4,5].

Tsuiki et al. in 2021 developed a deep convolutional neural network (DCNN) using data from lateral cephalometric radiographs of 1389 subjects and tested on 10% of the enrolled sample ($n = 131$) [7]. The DCNN, as well as manual cephalometric analyses, significantly predicted the presence of severe OSA, with a full image sensitivity/specificity of 0.87/0.82 ($\chi^2 = 62.5$, $p < 0.01$), demonstrating promising prospects for AI in the triage of OSAS. Nevertheless, these forecasting tools' predictive performances and features differ significantly between literature studies, limiting their generalizability and practicality.

The aim of our study was to evaluate the usefulness of the application of the SVM algorithm predicting the severity of OSA through clinical parameters, subjective questionnaires, anatomical scores, and associated comorbidities.

2. Materials and Methods

2.1. Study Design and Data Collection

Guidelines on strengthening the reporting of observational studies in epidemiology (STROBE) were followed [9]. We carried out a multicentric retrospective study conducted at

the Ear, Nose, and Throat Unit (ENT) of our hospitals from 1 January 2010 to 31 December 2021 (Figure 1).

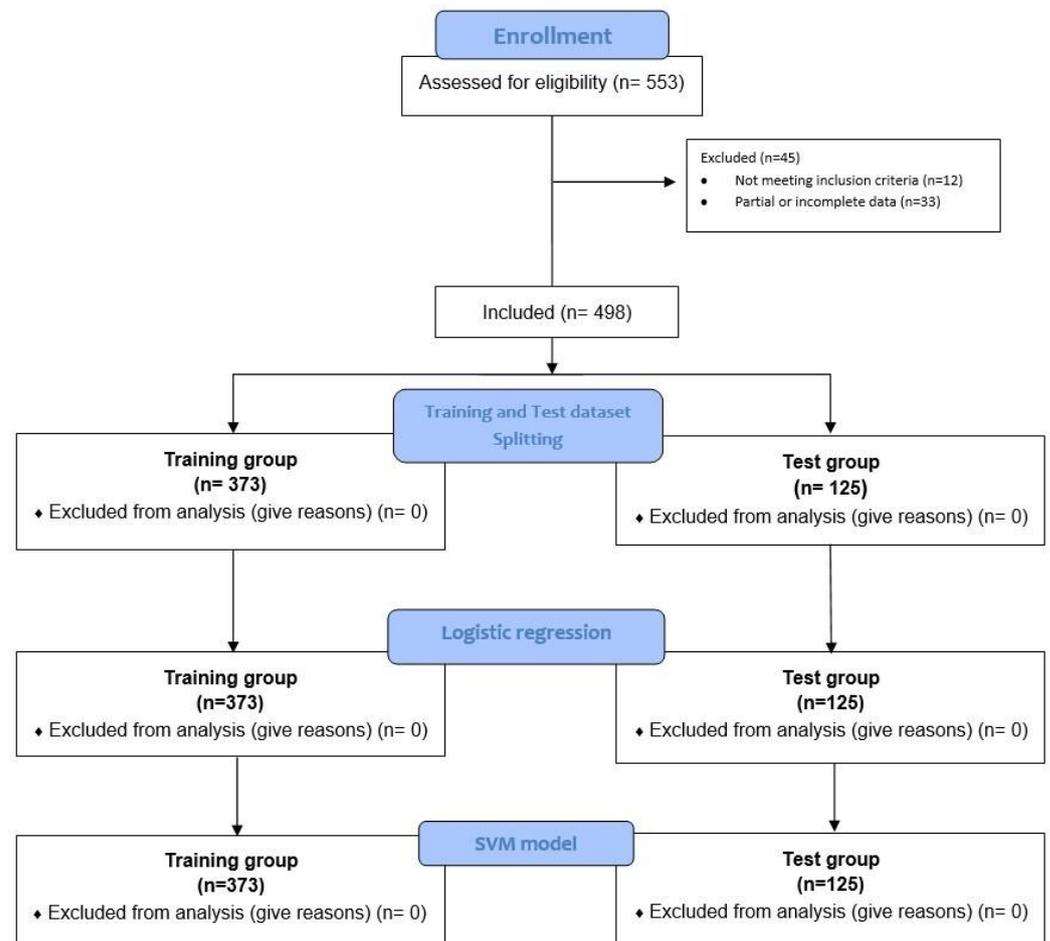


Figure 1. Flow diagram of preprocessing, dataset splitting, and SVM model training.

Participants aged ≥ 18 years who were referred to our units for sleep respiratory disorders were enrolled and subjected to clinical diagnostic evaluation and consequently phenotyped. A full clinical history of symptoms, such as morning headache and decreased libido, validated questionnaires (ESS), and anatomical or endoscopic scores were collected (Table 1) (Supplementary File S1–S3) [10–21].

Table 1. Main demographic features.

Characteristics	Total ($n = 498$)	Mild–Moderate OSA ($n = 220$)	Severe OSA ($n = 278$)	p -Value
Age	50.96 \pm 12.15	51.57 \pm 12.03	50.47 \pm 12.20	0.315
Gender				0.189
male	427/498 (87.76%)	179/498 (35.94%)	248/498 (49.79%)	
female	61/498 (12.24%)	31/498 (6.22%)	30/498 (6.02%)	
AHI	37.21 \pm 23.24	17.84 \pm 7.50	53.96 \pm 17.56	<0.001
ODI	35.37 \pm 24.79	17.76 \pm 17.82	49.38 \pm 20.13	<0.001
Mean SpO ₂	92.33 \pm 3.07	93.35 \pm 2.25	91.53 \pm 3.38	<0.001
Lower SpO ₂	75.92 \pm 12.13	80.05 \pm 11.87	72.65 \pm 11.31	<0.001
BMI	27.32 \pm 4.02	26.38 \pm 2.74	28.06 \pm 4.66	<0.001
ESS	7.97 \pm 4.92	7.26 \pm 4.43	8.54 \pm 5.20	0.003

All the participants underwent a sleep study that was carried out in an unattended way by means of a Polymesam Unattended Device 8-channel and then reviewed and scored by the same expert in sleep medicine according to the American Academy of Sleep Medicine (AASM) Guidelines [3].

We selected two diagnostic thresholds for OSA: mild to moderate ($AHI \geq 5$ events/h and $AHI < 30$ events/h) and severe ($AHI \geq 30$ events/h), according to the latest AASM guidelines on OSA management [3]. The algorithm was trained and tested using the two established threshold parameters to predict OSA patients' severity.

2.2. Statistical Analysis

Standard descriptive statistics were used, reporting mean and standard deviation for continuous variables and percentages for categorical ones. The independent t-test was performed for normally distributed values, while the Mann–Whitney U test was used for abnormally distributed values. The chi-square test was performed to test the observed and expected data differences. A value of $p < 0.05$ was deemed to be statistically significant. All analyses were performed using the Social Sciences Statistical Program (IBM SPSS Statistics for Windows, IBM Corp. Released 2017, Version 25.0 Armonk, NY, USA: IBM Corp).

2.3. Stratification Process

To perform the stratification of participants, we selected among independent variables assessed for disease severity the AHI cutoff according to the AASM guidelines to define disease severity. Thus, we stratified patients into a mild–moderate OSA group when the AHI value was ≤ 30 ; instead, a severe OSA group was defined when a value of $AHI > 30$ was found. Other variables identified were instead introduced into the predictive models as independent variables and converted into a binary value according to the respective cutoff definition.

Consequently, the splitting process of training and testing data divided the sample ($N = 498$) into two different homogeneous datasets. The training group included 75% ($N = 373$) of participants while the remaining 25% ($N = 125$) were used for testing. The performance of the logistic regression and SVM classifier models was tested according to the two AHI thresholds mild/moderate ($AHI \leq 30$) and severe ($AHI > 30$).

The study protocol was approved by the Ethical Committee and was conducted in accordance with the declaration of Helsinki.

We explored the data to improve their quality, handling missing or removing duplicate values, managing the existence of outliers or anomalies (data points differing substantially from the rest of the data), converting invalid or bad formatted values.

2.4. Logistic Regression Model

Logistic regression predictive models were used to classify patients and evaluate the performance of predictors. We used receiver operating characteristic (ROC) curves to assess the ability of the logistic regression models to identify patients with mild/moderate or severe AHI. Results were reported in terms of area under the curve (AUC) and 95% confidence interval (95% CI). A first multivariate logistic regression model was used to evaluate the model accuracy in an outcome prediction with the complete set of variables (full model). A second multivariate logistic model was the result of a backward stepwise elimination for selecting features and eliminating the ones that did not have a significant effect on the dependent variable or prediction of outcome to find a reduced model that best explained the data. We first worked on patients with a complete assessment of the following information: Age, gender, BMI, familiarity with OSAS, hypertension, cardiovascular disorders, diabetes, dyslipidemia, COPD, anxiety/depression, septoturbino-plasty, tonsillectomy, snoring, choking, morning headache, decreased libido, ESS, septal deviation, internal valve collapse, external valve collapse, lower turbinate hypertrophy, adenoid hypertrophy, Friedman tonsils score, Mallampati score, Friedman palate score, palate phenotype according to Woodson classification, endoscopic lingual tonsils score, retropalatal

Mueller maneuver, retrolingual Mueller maneuver, panting test, retrognathia, and upper jaw contraction. All tests were performed at a significance level $\alpha = 0.05$. After backward elimination, the reduced model included the following features: Age, gender, BMI, diabetes, anxiety/depression, choking, and septal deviation.

We consequently evaluated the multicollinearity of the logistic regression model due to the different features included. Therefore, a stepwise approach was performed in which previously removed features were individually reintroduced into the model to assess the risk of overfitting and the influence of individual values in the model.

2.5. SVM Model

The same training/test set and features from the full logistic regression were later used to develop the SVM model. Several datasets are not linearly separable even in a feature space, not satisfying all the constraints in the minimization problem of SVM. To fill this gap, Slack variables are introduced to allow certain constraints to be violated. By choosing very large slack variable values, we could find a degenerate solution that would lead to model overfitting [22]. To penalize the assignment of too large slack variables, the penalty is introduced in the classification objective:

- ε_i , indicates slack variables, one for each datapoint i , to allow certain constraints to be violated.
- C , indicates a tuning parameter that controls the trade-off between the penalty of slack variables ε_i and the optimization of the margin. High values of C penalize slack variables leading to a hard margin, whereas low values of C lead to a soft margin, which is a bigger corridor that allows certain training points inside at the expense of misclassifying some of them. In particular, the C parameter sets the confidence interval range of the learning model.

The radial basis function (RBF) kernel function expression on two samples, $x \wedge x'$, is defined as $K(x, x') = \exp(-\gamma | |x - x'| |^2)$ where $| |x - x'| |^2$ is the squared Euclidean distance between the two feature vectors, and γ is a free parameter. The RBF can be applied to a dataset by choosing two parameters, C and γ . The classifier performance of SVM depends on the choice of these two parameters.

A grid search method was used to find the optimal parameters of the RBF for SVM. This method considered m values in C and n values in γ , according to the $M \times N$ combination of C and γ , by training different SVMs using K -fold cross-validation. Here, we used a grid search on a 5-fold cross-validation to optimize accuracy. Thus, we selected as optimal parameters for the RBF kernel in the SVM model a γ value of 0.5 and C of 100.

Consequently, a Shapley plot (SHAP) was calculated. SHAP, which stands for Shapley Additive exPlanations, is an interpretability method based on Shapley values, a solution concept in cooperative game theory named in honor of Lloyd Shapley, who introduced it in 1951. SHAP was consequently applied to explain individual predictions of any machine learning (ML) model. The explanation model is represented by a linear model—an additive feature attribution method—or just the summation of present features in the coalition game. After the generation of Shapley plots, each predictor contribution to the SVM model output was described in terms of the SHAP average value.

The independent variables were thus ranked in descending order of importance. Instead, the violin plot generated shows the impact of a value associated with higher or lower prediction and positive or negative correlation on the X-axis. The color correlates with the average feature value at the plot position: Red areas represent highly valued features while blue areas are low. The violin plot also shows the outliers drawn as points.

The analyses were conducted using Python 3.6.9 with Statsmodel 0.10.2, Scikit-learn 1.0.2, and Shap 0.40 libraries.

Shapley plots show the contribution of each predictor to the SVM model output in terms of SHAP value. The variables were ranked by importance in descending order and the color represents the average feature value at that position. The violin plot shows the average medium SHAP value of each independent variable; the violin plot shows the

impact of a value associated with higher or lower prediction and positive or negative correlation. The violin plot also shows the outliers drawn as points.

2.6. Models Test Analysis

Comparison of the accuracy of the diagnostic tests performed was analyzed by an analysis of the areas under the receiver operating characteristic (ROC) curves. Models with the same characteristics and the same training/test set were analyzed through a Z-test to evaluate a statistical difference. The Z-test is a parametric statistical test used to evaluate whether the mean of a given distribution differs significantly from a hypothesized value.

2.7. Reporting Completeness of Machine Learning Study

We evaluated the reporting completeness of this research study referring to the TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis; www.tripod-statement.org, accessed on 8 February 2023) checklist for prediction model validation (accessed on 6 February 2023).

This statement contains a 20-item checklist, for a total of 31 items, with all sub-items included. The checklist contains questions about the title, abstract, background, methods, results, discussion, Supplementary Material, and funding information.

Each included item received a score of “1” for adherence and a score of “0” for non-adherence. Multiple items (items 1, 2, 3a, 4b, 5a, 6a, 7a, 7b, 9, 10a, 10b, 10d, 13a, 13b, 14a, 15a, 16, 17, 20, and 22) in the TRIPOD analysis were derived from several sub-items (the sub-items for each number can be found in www.tripod-statement.org (accessed on 7 June 2021)). The results of each TRIPOD item for each paper and the level of reporting adherence for each TRIPOD item were documented systematically in a spreadsheet.

We thus obtained a TRIPOD adherence score of 93,45% (29/31 items) by dividing the sum of TRIPOD items adhered to by the entire number of applicable TRIPOD items in the study.

3. Results

3.1. Patients Features

After selection, a total of 498 participants, with an age of 50.96 ± 12.15 years, were included in the study, of which 427/498 (87.76%) were male vs. 61/498 (12.24%) female (Table 1). The mean BMI was 27.32 ± 4.02 kg/cm. A mean AHI value of 37.21 ± 23.24 events/h and a mean oxygen desaturation index (ODI) of 35.37 ± 24.79 events/h were reported. An overall 220/498 (44.17%) of the participants were identified as mild to moderate OSA (AHI < 30 events/h), while 278/498 (55.82%) of the participants as severe OSA (AHI threshold ≥ 30 events/h).

Subsequently, the participants were divided into two homogeneous training (N = 373) and test (N = 125) groups, homogeneous for the independent variables included in the analysis (Figure 1).

3.2. Logistic Regression Analysis, Full and Reduced Models

Through the traditional statistical analysis, the full model demonstrated a ROC curve with an AUC of 0.73 (95% CI = 0.64–0.82) and 0.68 accuracy (Figure 2a).

The sensitivity and specificity of the regression to distinguish among OSA severity were, respectively, 0.74/0.63.

The consequent features selection using the backward stepwise elimination demonstrated a reduced logistic regression model with a sensitivity/specificity of 0.79/0.56, an accuracy of 0.67, and an AUC of 0.69 (95% CI = 0.6–0.78) (Figure 2b).

The assessment of the multicollinearity of the model by re-introducing the removed features one-by-one did not demonstrate a statistically significant difference in model performance.

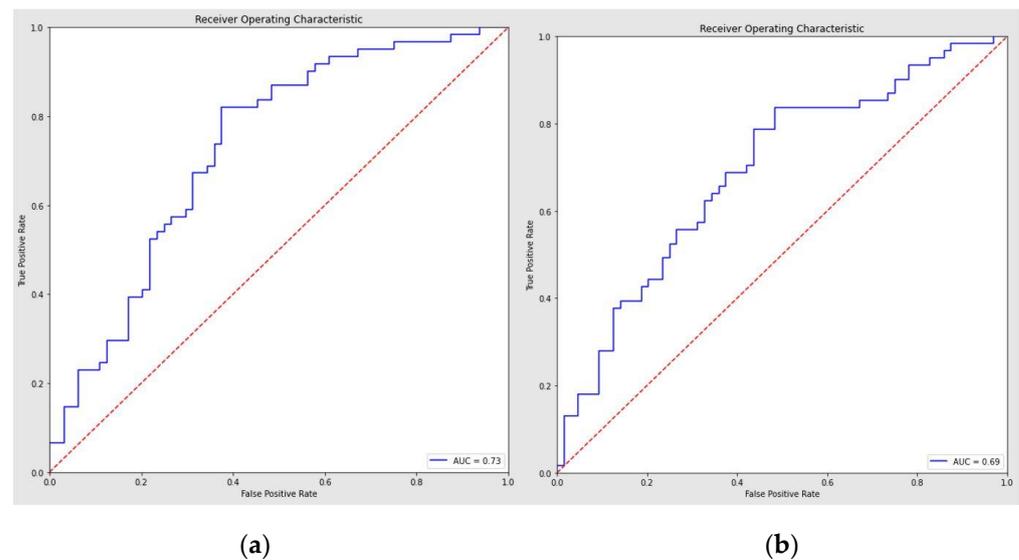


Figure 2. (a) Logistic regression full model. (b) Logistic regression classifier reduced model (features selection after backward elimination). List of selected features: Age, gender, BMI, diabetes, anxiety/depression, choking, septal deviation.

3.3. SVM Model Performance and ROC Curve Analysis

The SVM algorithm was adopted to discriminate OSA severity among participants. Through the algorithm, all the overall performance scores as accuracy (number of correct predictions/total number of predictions), ROC AUC (area under the curve), sensitivity (true positive rate), and specificity (true negative rate) were improved.

In particular, the SVM model demonstrated a recall and precision score of 0.80 and 0.93, respectively, to assess mild to moderate OSA; instead, the outcomes were 0.93 and 0.81 to identify severe OSA (Table 2).

Table 2. SVM model classification report. Accuracy, number of correct predictions/total number of predictions; ROC AUC, area under the curve; sensitivity, true positive rate; specificity, true negative rate. ^a SVM ROC AUC vs. Full regression model. ^b SVM ROC AUC vs. Reduced regression model. ^c Full regression ROC AUC vs. reduced regression.

Model	AHI	Precision	Recall	F1-Score	Sensitivity	Specificity	Accuracy	p-Value
Full Logistic	mild–moderate	0.71	0.62	0.67				
	severe	0.65	0.74	0.69	0.74	0.63	0.68	$p < 0.001^a$
Reduced Logistic	mild–moderate	0.73	0.56	0.64				
	severe	0.63	0.79	0.70	0.79	0.56	0.67	$p < 0.001^b$
SVM	mild–moderate	0.93	0.80	0.86				
	severe	0.81	0.93	0.87	0.93	0.80	0.86	$p = 0.541^c$

To avoid bias within the dataset, we evaluated the model score using a 10-fold stratified cross-validation strategy, obtaining an average accuracy score of 0.87 (95% CI = [0.83, 0.91]). The result of the stratified cross-validation is in line with the accuracy score of the model trained on 75% of the dataset and tested on the remaining 25%. The algorithm provided a ROC with an AUC of 0.92 (95% confidence interval = 0.87–0.97) (Figure 3).

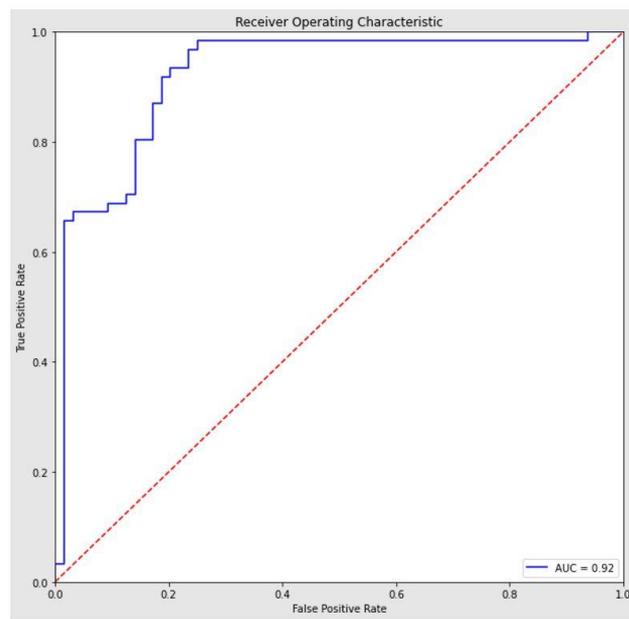


Figure 3. ROC SVM model.

The clinical variables included in the model and the Shapley plot consequently generated showed the contribution of each predictor to the SVM model, as summarized in Figure 4a,b.

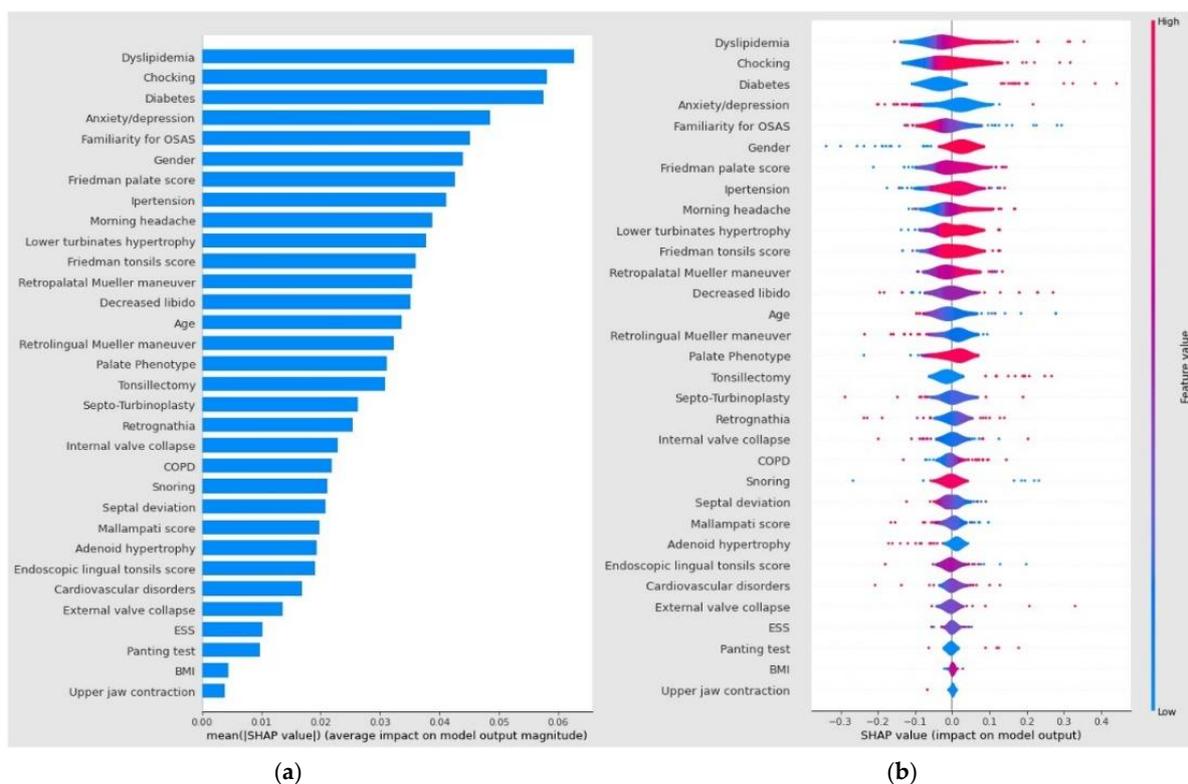


Figure 4. Shapley plots show the contribution of each predictor to the SVM model output in terms of SHAP value. The variables were ranked by importance in descending order, and the color represents the average feature value at that position. (a) The violin plot shows the average medium SHAP value of each independent variable; (b) the violin plot shows the impact of a value associated with higher or lower prediction and positive or negative correlation. The violin plot also shows the outliers drawn as points.

Among the independent variable predictors of OSAS severity assessed, dyslipidemia (0.0623), choking (0.0588), diabetes (0.0576), mood disorders (0.0486), and familiarity for OSAS (0.0452) demonstrated a higher average impact on model output magnitude in terms of an absolute mean SHAP value. The ROC AUC comparison between the complete logistic regression model (AUC = 0.73) and the SVM model (AUC = 0.92) through the Z-test confirmed a statistically significant difference among the two models ($p < 0.001$).

Instead, the ROC AUC comparison between full and reduced logistic regression model was not significant ($p = 0.541$).

4. Discussion

Our study demonstrated how artificial intelligence can be useful in assessing patients with OSA-related symptoms and determining the risk of disease severity with clinic-based algorithms. We reported a sensitivity and specificity significantly higher for the SVM model than classical logistic regression. Indeed, as reported in our results through the SVM model, it is possible to stratify OSA patients according to the severity with a higher accuracy of 0.86 compared with the full logistic regression mode accuracy of 0.68.

4.1. Diagnostic and Therapeutic Role

The use of AI technology in clinical practice is an emerging and debated topic, both for its possible diagnostic and therapeutic implications. Through ML, it is possible to exploit the multiple variables of easy and rapid extraction, such as a patient's clinical history and anthropometric or demographic characteristics, to facilitate the identification of otherwise complex pathologies [6].

The field of breathing sleep disorders could also benefit from improving ML Technology, using both its application in early diagnosis and the analysis of predictive factors of response to medical or surgical treatment [7,8,23].

Kim et al. in 2021, through a preoperative machine learning-based clinical mode, improved the prediction of the therapeutic sleep surgery outcome with higher accuracy for the gradient boosting model than the logistic models [24].

The application of the algorithm elaborating several parameters collected at baseline visits between adherent and non-adherent groups demonstrated a sensitivity of 68.6% and an AUC of 72.9% through the vector machine model.

4.2. Diagnostic Application of AI

Sleep breathing disorders lend themselves well to the diagnostic application of AI as they are often strongly correlated with elevated BMI, and cardiovascular, metabolic, or central nervous system disorders.

Holfinger et al. assessed the diagnostic performance of OSA machine learning prediction tools using readily available data, such as age, sex, BMI, and race, and compared the efficacy with a STOP-BANG-based model [23]. The authors included a wide cohort of 17,448 subjects in a retrospective study, demonstrating that AUCs (95% CI) of the kernel support vector machine (0.66 [0.65–0.67]) were significantly higher than logistic regression ones (0.61 [0.60–0.62]).

Machine learning-derived algorithms may also improve and simplify the widespread identification of OSA in the pediatric population, providing better diagnostic performance than logistic regression with patient-reported symptoms [8,24–26].

Gutiérrez-Tobal et al. performed a systematic review of studies assessing the reliability of a machine learning-based method implementation of OSA detection in clinical practice [25]. The authors retrieved 90 studies involving 4767 different pediatric subjects and demonstrated an improved ML diagnostic performance on OSA severity criteria (sensitivity = 0.652; specificity = 0.931; and AUC = 0.940).

However, an important aspect that has not been analyzed on the advantages of artificial intelligence in OSA is its possible use in identifying OSA pathology concerning the healthy

population, determining which patients are at risk, and optimizing the use of diagnostic resources, such as polysomnography [6,7].

Our analysis confirmed the superiority of vector models using SVM models in determining disease severity in patients with OSA compared to traditional logistic regression models. In fact, we found a significantly greater sensitivity/specificity of 0.93/0.80, and an accuracy of 0.86 for the SVM than the logistic regression full mode.

It is well-known in the literature that OSA is associated with obesity and cardiovascular and cerebrovascular disorders. Several mechanisms occur in patients with OSA, including a chronic inflammatory state, intermittent hypoxia, and even alterations of their lipid profile, probably due to a reduction in androgen levels. Moreover, OSA severity and lipid-related comorbidities, such as atherosclerosis, possess a well-known correlation.

4.3. OSA Risk Factors and Comorbidities

Our SVM model demonstrated dyslipidemia as the highest average SHAP feature value (0.0623) for OSA severity among clinical variables included in the diagnostic algorithm. Conversely, its predictive role was not confirmed in the logistic regression full model analysis ($p = 0.123$). Although OSA is commonly associated with craniofacial anomalies and palatal or base tongue disorders validated as upper airway obstruction sites by DISE, no research study has evaluated the predictive role of such features using a machine learning algorithm. Craniofacial variables represent noteworthy risk factors for OSA. The risk of OSA in adult subjects with altered craniofacial anatomy on lateral cephalograms is well-known in the literature [6]. Among the variables with significant heterogeneity, the position and length of the mandible (BNS: -1.49° and Go-Me: -5.66 mm), the area of the tongue (T: 366.51 mm²), and the soft palate (UV: $125, 02$ mm²), and upper airway length (UAL: 5.39 mm) were identified as strongly correlating with the presence of OSA. Our SVM analysis demonstrated a higher average SHAP feature value for the palate phenotype (0.0308), Friedman tonsils score (0.0363), and palate score (0.0623) for OSA severity than other clinical features assessed in the model. However, the predictive role in the logistic regression full model analysis was not significant for the palate phenotype ($p = 0.148$), while Friedman's tonsils score ($p = 0.041$) and Friedman's palate score ($p = 0.038$) were statistically significant.

A systematic review of the prevalence of OSA in an Asian population reported interesting data on highly related features, such as gender, older age, BMI increase, and arterial hypertension, that significantly correlated with the onset of sleep breathing disorders and severity [4]. The difference in gender prevalence rates was also supported by our analysis, with gender, age, and hypertension showing a SHAP value of 0.0435, 0.0336, and 0.0408, respectively. Conversely, in the logistic regression full model, the predictive role in the analysis was significant for age ($p = 0.017$) and gender ($p = 0.019$), while not for hypertension ($p = 0.649$).

4.4. Study's Limitations

Prediction tools that determine OSA risk include variables, such as patient-reported symptoms through symptom questionnaires; however, there are often unavailable or inaccurate data from large groups of individuals with multicenter studies using electronic medical records that are not similar across institutions.

Although the machine learning model demonstrated significant AUC and adequate sensitivity and specificity, our study's main shortcoming was the reduced sample with which the model was trained. In fact, it is known how artificial intelligence models become proportionally more effective as new subjects are added, with a refinement of the predictive algorithm. A consequence of this limitation is that in our model, highly significant features in the literature, such as BMI, Mueller maneuver, ESS, or cardiovascular disorders, do not take on a dimension in the model that reflects clinical severity.

Ultimately, although the verification of inputs and data values is important, we must also consider the validation of the data model itself. In fact, our data model could be not

correctly structured, leading to several biases. Therefore, a future perspective to enable the potential clinical application of our model will be to validate it using independent data.

5. Conclusions

The development of an AI model to predict the risk of developing OSA severity has shown promising prospects for application in a clinical setting after adequate training and sufficiently large samples. Using demographic data associated with easily detectable clinical or endoscopic scores, a practical model for predicting OSA severity in the future could be possible.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/life13030702/s1>, File S1: OSA clinical and endoscopic parameters; File S2: Mood and neurological Osa disorders; File S3: Osa comorbidities.

Author Contributions: Conceptualization, A.M., C.V. and P.M.R.; methodology, G.I.; software, J.R.L.; validation, I.L.M., M.D.V. and G.C.; formal analysis, S.C. and C.C.-H.; investigation, M.D.L., G.M. (Giuseppe Meccariello), and A.P.; resources, C.C.E.; data curation, A.M.S.; writing—original draft preparation, I.P.; writing—review and editing, G.S.; visualization, P.D.M.; supervision, A.C.; project administration, R.L.; funding acquisition, A.P., A.G. and G.M. (Giuseppe Magliulo). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of Rome (Prot.0305/2021).

Informed Consent Statement: Written informed consent has been obtained from the patient(s) to publish this paper.

Data Availability Statement: Data is unavailable due to privacy or ethical restrictions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. 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*, e56–e67. [[CrossRef](#)] [[PubMed](#)]
2. Stores, G. Clinical diagnosis and misdiagnosis of sleep disorders. *J. Neurol. Neurosurg. Psychiatry* **2007**, *78*, 1293–1297. [[CrossRef](#)]
3. 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. [[CrossRef](#)]
4. Mirrakhimov, A.E.; Sooronbaev, T.; Mirrakhimov, E.M. Prevalence of obstructive sleep apnea in Asian adults: A systematic review of the literature. *BMC Pulm. Med.* **2013**, *13*, 10. [[CrossRef](#)]
5. Amado-Garzón, S.B.; Ruiz, A.J.; Rondón-Sepúlveda, M.A.; Hidalgo-Martínez, P. Sensitivity and specificity of four screening sleep-disordered breathing tests in patients with and without cardiovascular disease. *Sleep Sci.* **2021**, *14*, 311–318. [[CrossRef](#)]
6. Sutherland, K.; Chapman, J.L.; A Cayanan, E.; Lowth, A.B.; Hoyos, C.M.; Wong, K.K.H.; Yee, B.J.; Grunstein, R.R.; A Cistulli, P.; Marshall, N.S. Does craniofacial morphology relate to sleep apnea severity reduction following weight loss intervention? A patient-level meta-analysis. *Sleep* **2021**, *44*, zsa207. [[CrossRef](#)]
7. Tsuiki, S.; Nagaoka, T.; Fukuda, T.; Sakamoto, Y.; Almeida, F.R.; Nakayama, H.; Inoue, Y.; Enno, H. Machine learning for image-based detection of patients with obstructive sleep apnea: An exploratory study. *Sleep Breath.* **2021**, *25*, 2297–2305. [[CrossRef](#)]
8. Maranate, T.; Pongpullponsak, A.; Ruttanaumpawan, P. The Prioritization of Clinical Risk Factors of Obstructive Sleep Apnea Severity Using Fuzzy Analytic Hierarchy Process. *Comput. Math. Methods Med.* **2015**, *2015*, 257856. [[CrossRef](#)] [[PubMed](#)]
9. von Elm, E.; Altman, D.G.; Egger, M.; Pocock, S.J.; Gøtzsche, P.C.; Vandenbroucke, J.P.; STROBE Initiative. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies. *Int. J. Surg.* **2014**, *12*, 1495–1499. [[CrossRef](#)] [[PubMed](#)]
10. de Aguiar Vidigal, T.; Martinho Haddad, F.L.; Gregório, L.C.; Poyares, D.; Tufik, S.; Azeredo Bittencourt, L.R. Subjective, anatomical, and functional nasal evaluation of patients with obstructive sleep apnea syndrome. *Sleep Breath.* **2013**, *17*, 427–433. [[CrossRef](#)] [[PubMed](#)]
11. Hwang, T.S.; Song, J.; Yoon, H.; Cho, B.P.; Kang, H.S. Morphometry of the nasal bones and pyriform apertures in Koreans. *Ann. Anat.* **2005**, *187*, 411–414. [[CrossRef](#)] [[PubMed](#)]

12. Patel, B.; Virk, J.S.; Randhawa, P.S.; Andrews, P.J. The internal nasal valve: A validated grading system and operative guide. *Eur. Arch. Otorhinolaryngol.* **2018**, *275*, 2739–2744. [[CrossRef](#)] [[PubMed](#)]
13. Poirrier, A.L.; Ahluwalia, S.; Kwame, I.; Chau, H.; Bentley, M.; Andrews, P. External nasal valve collapse: Validation of novel outcome measurement tool. *Rhinology* **2014**, *52*, 127–132. [[CrossRef](#)] [[PubMed](#)]
14. Camacho, M.; Zaghi, S.; Certal, V.; Abdullatif, J.; Means, C.; Acevedo, J.; Liu, S.Y.C.; Brietzke, S.E.; Kushida, C.A.; Capasso, R. Inferior turbinate classification system, grades 1 to 4: Development and validation study. *Laryngoscope* **2015**, *125*, 296–302. [[CrossRef](#)]
15. Cassano, P.; Gelardi, M.; Cassano, M.; Fiorella, M.L.; Fiorella, R. Adenoid tissue rhinopharyngeal obstruction grading based on fiberoendoscopic findings: A novel approach to therapeutic management. *Int. J. Pediatr. Otorhinolaryngol.* **2003**, *67*, 1303–1309. [[CrossRef](#)] [[PubMed](#)]
16. Kumar, D.S.; Valenzuela, D.; Kozak, F.K.; Ludemann, J.P.; Moxham, J.P.; Lea, J.; Chadha, N.K. The reliability of clinical tonsil size grading in children. *JAMA Otolaryngol. Head Neck Surg.* **2014**, *140*, 1034–1037. [[CrossRef](#)]
17. Friedman, M.; Hamilton, C.; Samuelson, C.G.; Lundgren, M.E.; Pott, T. Diagnostic value of the Friedman tongue position and Mallampati classification for obstructive sleep apnea: A meta-analysis. *Otolaryngol. Head Neck Surg.* **2013**, *148*, 540–547. [[CrossRef](#)]
18. Berg, L.M.; Ankjell, T.K.S.; Sun, Y.-Q.; Trovik, T.A.; Sjögren, A.; Rikardsen, O.G.; Moen, K.; Hellem, S.; Bugten, V. Friedman Score in Relation to Compliance and Treatment Response in Nonsevere Obstructive Sleep Apnea. *Int. J. Otolaryngol.* **2020**, *2020*, 6459276. [[CrossRef](#)]
19. Woodson, B.T. Palatal Anatomy and Phenotypes for the Treatment of OSA. *Curr. Otorhinolaryngol. Rep.* **2021**, *9*, 238–245. [[CrossRef](#)]
20. Friedman, M.; Yalamanchali, S.; Gorelick, G.; Joseph, N.J.; Hwang, M.S. A standardized lingual tonsil grading system: Interexaminer agreement. *Otolaryngol. Head Neck Surg.* **2015**, *152*, 667–672. [[CrossRef](#)]
21. Wu, M.J.; Ho, C.Y.; Tsai, H.H.; Huang, H.M.; Lee, P.L.; Tan, C.T. Retropalatal Müller grade is associated with the severity of obstructive sleep apnea in non-obese Asian patients. Retropalatal Müller grade and OSA in non-obese. *Sleep Breath.* **2011**, *15*, 799–807. [[CrossRef](#)] [[PubMed](#)]
22. Barchitta, M.; Maugeri, A.; Favara, G.; Riela, P.; Gallo, G.; Mura, I.; Agodi, A. Early Prediction of Seven-Day Mortality in Intensive Care Unit Using a Machine Learning Model: Results from the SPIN-UTI Project. *J. Clin. Med.* **2021**, *10*, 992. [[CrossRef](#)]
23. Holfinger, S.J.; Lyons, M.M.; Keenan, B.T.; Mazzotti, D.R.; Mindel, J.; Maislin, G.; Cistulli, P.A.; Sutherland, K.; McArdle, N.; Singh, B.; et al. Diagnostic Performance of Machine Learning-Derived OSA Prediction Tools in Large Clinical and Community-Based Samples. *Chest* **2022**, *161*, 807–817. [[CrossRef](#)] [[PubMed](#)]
24. Kim, J.Y.; Kong, H.-J.; Kim, S.H.; Lee, S.; Kang, S.H.; Han, S.C.; Kim, D.W.; Ji, J.-Y.; Kim, H.J. Machine learning-based preoperative datamining can predict the therapeutic outcome of sleep surgery in OSA subjects. *Sci. Rep.* **2021**, *11*, 14911. [[CrossRef](#)] [[PubMed](#)]
25. Gutiérrez-Tobal, G.C.; Álvarez, D.; Kheirandish-Gozal, L.; Del Campo, F.; Gozal, D.; Hornero, R. Reliability of machine learning to diagnose pediatric obstructive sleep apnea: Systematic review and meta-analysis. *Pediatr. Pulmonol.* **2021**, *57*, 1931–1943. [[CrossRef](#)] [[PubMed](#)]
26. Collins, G.S.; Reitsma, J.B.; Altman, D.G.; Moons, K.G. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *BMJ* **2015**, *350*, g7594. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.