

Article

A Study on Improving Sleep Apnea Diagnoses Using Machine Learning Based on the STOP-BANG Questionnaire

Myoung-Su Choi ^{1,†} , Dong-Hun Han ^{2,†} , Jun-Woo Choi ³ and Min-Soo Kang ^{4,*} 

¹ Daejeon Eulji Medical Center, Department of Otolaryngology-Head and Neck Surgery, Eulji University School of Medicine, Daejeon 35233, Republic of Korea; mschoi@eulji.ac.kr

² Department of Medical Artificial Intelligence, Eulji University, Seongnam 13135, Republic of Korea; d555v@naver.com

³ Department of Medical IT, Eulji University, Seongnam 13135, Republic of Korea; chlwnsdn456@naver.com

⁴ Department of Bigdata Medical Convergence, Eulji University, Seongnam 13135, Republic of Korea

* Correspondence: mskang@eulji.ac.kr

† These authors contributed equally to this work.

Abstract: Sleep apnea has emerged as a significant health issue in modern society, with self-diagnosis and effective management becoming increasingly important. Among the most renowned methods for self-diagnosis, the STOP-BANG questionnaire is widely recognized as a simple yet effective tool for diagnosing and assessing the risk of sleep apnea. However, its sensitivity and specificity have limitations, necessitating the need for tools with higher performance. Consequently, this study aimed to enhance the accuracy of sleep apnea diagnoses by integrating machine learning with the STOP-BANG questionnaire. Research through actual cases was conducted based on the data of 262 patients undergoing polysomnography, confirming sleep apnea with a STOP-BANG score of ≥ 3 and an Apnea-Hypopnea Index (AHI) of ≥ 5 . The accuracy, sensitivity, and specificity were derived by comparing Apnea-Hypopnea Index scores with STOP-BANG scores. When applying machine learning models, four hyperparameter-tuned models were utilized: K-Nearest Neighbor (K-NN), Logistic Regression, Random Forest, and Support Vector Machine (SVM). Among them, the K-NN model with a K value of 11 demonstrated superior performance, achieving a sensitivity of 0.94, specificity of 0.85, and overall accuracy of 0.92. These results highlight the potential of combining traditional STOP-BANG diagnostic tools with machine learning technology, offering new directions for future research in self-diagnosis and the preliminary diagnosis of sleep-related disorders in clinical settings.

Keywords: sleep apnea; AI; STOP-BANG; machine learning; diagnostic accuracy

check for
updates

Citation: Choi, M.-S.; Han, D.-H.; Choi, J.-W.; Kang, M.-S. A Study on Improving Sleep Apnea Diagnoses Using Machine Learning Based on the STOP-BANG Questionnaire. *Appl. Sci.* **2024**, *14*, 3117. <https://doi.org/10.3390/app14073117>

Academic Editor: Juan A. Gómez-Pulido

Received: 12 March 2024

Revised: 4 April 2024

Accepted: 5 April 2024

Published: 8 April 2024



Copyright: © 2024 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/>).

1. Introduction

In contemporary society, the prevalence of irregular sleep patterns and insomnia due to excessive studying, 24 h shift work, workplace demands, and frequent international travel is on the rise. According to the National Institutes of Health in the United States, the prevalence of sleep disorders in the American population is 14.71%, leading to a productivity loss valued at USD 50 billion due to work-related slowdowns, in addition to USD 16 billion spent on health care costs related to drowsy driving and insomnia [1]. Disturbances in sleep functionality can result in decreased quality of life, diminished concentration, frequent illnesses and headaches, muscle pain, psychiatric issues, and an increase in mortality and sudden death [2]. Sleep apnea is a condition characterized by airway blockage during sleep, leading to symptoms such as snoring, episodes of apnea during sleep, and unstable respiration. With the recent rise in obesity and an aging population, the rate of sleep apnea is steeply increasing, necessitating further research into the condition. A diagnosis of sleep apnea requires a polysomnography test in a medical facility, which patients find burdensome because of the high costs and unfamiliar

environment [3]. Moreover, it is practically challenging to refer all patients suspected of having sleep apnea for testing, making alternative pre-screening tools such as the Berlin and STOP-BANG questionnaires crucial [4].

Of these, the STOP-BANG questionnaire is highly regarded for its use in the initial assessment of patients suspected of having sleep apnea, as well as for evaluating the risk of sleep apnea in patients scheduled for emergency surgery or those who are unable to undergo immediate polysomnography. Despite its high sensitivity, the STOP-BANG questionnaire's low specificity has been deemed problematic [5,6]. This concern was highlighted by the research team led by Cinthya A. at the Mayo Clinic, who questioned the questionnaire's sensitivity for obstructive sleep apnea (OSA) screening [7]. This study was dedicated to evaluating the accuracy of the STOP-BANG questionnaire by analyzing data on sleep apnea outcomes. Furthermore, it aimed to compare the outcomes of various classification models designed to improve specificity. In doing so, this study assessed a machine learning model that demonstrated both elevated sensitivity and specificity.

2. Related Research

2.1. STOP-BANG

In 2008, Frances Chung and her research team published "STOP questionnaire: a tool to screen patients for obstructive sleep apnea," which detailed the development and validation of a questionnaire tool aimed at distinguishing patients with sleep apnea among surgical patients. The STOP-BANG questionnaire consists of eight items, including questions about snoring, tiredness, observed apnea, and blood pressure, with additional criteria considering BMI ($\text{BMI} > 35 \text{ kg/m}^2$), age (age > 50 years), neck circumference (> 41 cm), and gender (male). Each 'yes' answer scores one point, with a total possible score of eight. A score of three or more is considered indicative of being at risk of sleep apnea. The STOP-BANG questionnaire, composed of questions related to snoring, daytime fatigue, observed apneas, and hypertension, was administered to 2467 patients. Sensitivities for Apnea–Hypopnea Index (AHI) thresholds of greater than 5, 15, and 30 were 65.6%, 74.3%, and 79.5%, respectively, for the STOP questionnaire cutoffs. Therefore, the STOP-BANG questionnaire can be considered useful for identifying patients at high risk of sleep apnea. [8] In 2011, Robert J. Farney, M.D., undertook an evaluation of the STOP-BANG questionnaire's capability to stratify obstructive sleep apnea (OSA) severity into four distinct categories, spanning from none to severe. The diagnostic utility of the STOP and STOP-BANG questionnaires for OSA was studied through logistic regression analysis on the study subjects, demonstrating the significance of classification prediction based on numerical regression analysis by predicting the severity of sleep apnea based on the AHI [9]. A 2019 study conducted by Hyun-Ju Yang established a statistically significant correlation between the prevalence of obstructive sleep apnea (OSA) and body mass index (BMI). This investigation reviewed literature spanning the previous decade, distinguishing between community-based populations and individuals seeking evaluation at sleep clinics, and it evaluated the efficacy of various diagnostic questionnaires. The findings revealed that the STOP-BANG questionnaire exhibited the highest sensitivity in diagnosing OSA among domestic patients, whereas the Berlin questionnaire demonstrated superior specificity. Consequently, it was proposed that the STOP-BANG questionnaire could be refined to improve its specificity by integrating additional physical measurements and physiological indices for individuals who are preliminarily diagnosed with OSA [10].

2.2. Application of Machine Learning

Machine learning is widely employed in clinical analysis across the medical field, utilizing probabilistic estimation to elucidate the relationships between a dependent variable and one or more independent variables [11]. An illustrative example was provided by the 2022 study conducted by Holfinger, S., et al. that juxtaposed the diagnostic efficacy of sleep apnea using machine learning tools against that of the STOP-BANG questionnaire [12]. This research highlighted that machine learning predictive models are capable of facilitating the

comprehensive identification of sleep apnea cases, including those where symptoms may not be reported or are unrecognized by patients [12]. Consequently, machine learning predictive tools are demonstrated to be effective in predicting sleep apnea in scenarios where patient survey errors or unacknowledged symptoms might otherwise impede diagnosis.

In 2023, Javeed, A., et al. introduced a machine learning model designed to provide accurate predictions and identify risk factors associated with the onset of sleep apnea [13]. Utilizing electronic health records from the Swedish National Study on Aging and Care (SNAC), this model integrated a comprehensive set of 75 features across 10,765 samples. The model's algorithm implemented a cross-validation approach, featuring an XGBoost module for assessing important features in the feature space and a BiLSTM (Bidirectional Long Short-Term Memory) module for classifying the probability of sleep apnea, achieving an accuracy of 97% with only the six most critical features from the dataset [13]. This underscores the model's utility in identifying risk factors and improving the diagnosis of sleep apnea. Furthermore, in 2023, Shi and his research team developed a machine learning model based on clinical characteristics and questionnaires to predict OSA risk and analyze risk factors. Using data from 1656 patients who underwent multiple polysomnographies (PSGs) between 2018 and 2021, the study included 23 variables and added 15 variables after univariate analysis. To evaluate predictive performance, six classification models were employed: Logistic Regression (LR), Gradient Boosting Machine (GBM), Extreme Gradient Boosting (XGBoost), Adaptive Boosting (AdaBoost), Bagging, and Multilayer Perceptron (MLP), with the GBM model showing the best performance [14]. This underscores the importance of integrating questionnaire-based approaches with machine learning models. Concurrently, in the same year, Han, H. proposed a novel screening method aimed at addressing the limitations inherent in polysomnographies (PSGs). The methodology leveraged data from 4014 patients, applying both supervised and unsupervised learning techniques for data clustering, utilizing hierarchical agglomerative clustering, K-means, bisecting K-means algorithms, and Gaussian mixture models. The study embarked on feature engineering, harnessing both medically established methodologies and machine learning techniques. Through the deployment of gradient boost-based models, including XGBoost, LightGBM, CatBoost, and Random Forest for classification tasks, it achieved notable success in predicting the severity of obstructive sleep apnea (OSA). The classification accuracies were reported as 88%, 88%, and 91% for Apnea-Hypopnea Index (AHI) thresholds of 5, 15, and 30, respectively [15]. This demonstrates the potential of machine learning in predicting the severity of obstructive sleep apnea. Lastly, Bazoukis, G., et al., in their 2023 work, summarized existing research on the role of machine learning in diagnosing, classifying, and treating sleep-related breathing disorders, reviewing studies up to January 2022 from Medline, EMBASE, and the Cochrane database [16]. Among the 132 studies reviewed, the use of characteristics derived from electrocardiograms, pulse oximetry, and sound signals showed promising performance in diagnosing sleep apnea and was significant in classifying tonsillar and central categories and predicting the severity of apnea [16].

These studies collectively affirm the significant role and implementation of machine learning technologies in the field of medicine and sleep-related breathing disorders.

3. Experiment

3.1. Data Analysis

This study was conducted from May 2022 to August 2023 at the Eulji University Hospital's Department of Otorhinolaryngology sleep study laboratory based on test data from 262 individuals. All data were anonymized and de-identified, and the study received ethical approval with the certification number 2022-03-010-002 from the Institutional Review Board (IRB). The definition of obstructive sleep apnea (OSA) followed the criteria set by the International Classification of Sleep Disorders-3 (ICSD-3) [17]. OSA was defined as a condition where, despite the observation of respiratory effort from the abdomen and chest,

there are five or more episodes per hour of apnea or hypopnea occurring in the oral and nasal passages, each lasting a minimum of 10 s.

Data were collected from individuals undergoing polysomnography who voluntarily chose to participate in this study, specifically targeting individuals aged 19 years and older. The exclusion criteria included the following. (1) Individuals under the age of 19. (2) Patients with a total sleep duration of less than 240 min. (3) Individuals with tattoos, pigmentation, or lesions on their wrists. (4) Individuals who are unable to clearly express pain. (5) Individuals suffering from severe anemia, undergoing chemotherapy, experiencing uncontrolled respiratory diseases, or afflicted with blood-related disorders. An overview of the data is presented in Table 1. Table 2 displays the number of respondents who selected either 1 or 0 for each question in the STOP-BANG questionnaire.

Table 1. Patient demographics (n = 262).

	Value
Age (years)	45.5 ± 13.3
Sex: Male	195 (74.4%)
Sex: Female	67 (25.6%)
BMI (kg/m ²)	28.2 ± 4.6
Neck circumference (cm)	39.2 ± 3.9
Waist-to-hip ratio	0.95 ± 0.1
PSQI score	11.8 ± 5.5
ESS score	8.5 ± 3.7
TST (min)	341.0 ± 50.7
AHI (events/h)	24.2 ± 20.0
ODI (events/h)	26.7 ± 23.9
STOP-BANG score	4.1 ± 1.5
SpO ₂ nadir (%)	80.9 ± 8.5
Diagnosis: OSA (AHI < 5)	53 (20.2%)
Diagnosis: OSA (AHI ≥ 5)	209 (79.8%)

Data are presented as mean ± standard deviation or n (%). BMI, body mass index; PSQI, Pittsburgh sleep quality index; ESS, Epworth sleepiness scale; TST, total sleep time; AHI, Apnea–Hypopnea Index; ODI, oxygen desaturation index; SpO₂, oxygen saturation; OSA, obstructive sleep apnea.

Table 2. STOP-BANG (n = 262).

	Snore	Tired	Observe	Pressure	BMI	Age	Neck	Gender
1	221	214	158	155	244	145	145	195
0	41	48	104	107	18	117	117	67

The diagnostic methodology reported an overall accuracy of 86.52%, indicating considerable reliability in the applied approach. Specifically, the sensitivity for accurately diagnosing obstructive sleep apnea (OSA) in patients with a STOP-BANG score of 3 or higher was ascertained to be 89.92%. This suggested that a significant proportion of patients classified within this score range genuinely suffered from OSA, with only a 10.08% likelihood of misdiagnosing healthy individuals. Conversely, the specificity for individuals with a STOP-BANG score of less than 3—the proportion of healthy individuals correctly identified as not having OSA—was determined to be 61.76%. This indicated a potential for 38.24% of individuals to be erroneously classified as OSA patients at scores below 3 while

accurately identifying non-OSA individuals at a rate of 61.76%. To validate the accuracy of these observations, a comparative analysis with existing research was conducted. A validation study of the STOP-BANG questionnaire in 2021 by Hwang M et al. reported an average sensitivity of 89.1% and a specificity of 32.3% for an Apnea–Hypopnea Index (AHI) of 5 or higher, with an area under the curve (AUC) of 0.86 [18].

3.2. Import Machine Learning

In this study, different machine learning techniques were applied and compared to evaluate the diagnostic possibility of obstructive sleep apnea (OSA). The algorithms used included Logistic Regression, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbor (KNN). Each algorithm was used to create a STOP-BANG enhancement model based on patients' STOP-BANG data and actual OSA diagnostic results. These models were used to assess the correlation and classification accuracy between the STOP-BANG questionnaire responses and the presence of OSA.

3.2.1. Logistic Regression

The evaluation of the diagnostic possibility of obstructive sleep apnea (OSA) using Logistic Regression was implemented as follows.

$$\hat{y} = \sigma(w^T x + b) \quad (1)$$

Here, “ \hat{y} ” represents the predicted probability, indicating the likelihood of belonging to class 1 (presence of obstructive sleep apnea). “ $\sigma(\cdot)$ ” denotes the logistic sigmoid function, a nonlinear function that transforms a linear combination into a probability. This function is defined as follows [19,20].

$$\sigma(z) = 1 / (1 + e^{-z}) \quad (2)$$

In this study, the term “ w ” represents the vector of model weights or regression coefficients, while “ x ” refers to the vector of independent variables or features. The term “ b ” is the model's intercept, also known as bias. The notation “ T ” signifies the transpose of a vector. The regularization strength, denoted as “ C ”, inversely affected the model's complexity; a lower “ C ” value meant stronger regularization, which helped to prevent the model from overfitting. By exploring various values for the hyperparameter “ C ” to adjust the regularization strength, this study aimed to mitigate overfitting and enhance the model's performance. The regression coefficients depict the influence of each variable on the log odds, indicating the importance of these variables. The coefficients derived in this study showed the degree to which each feature had a positive or negative effect on the dependent variable, specifically the diagnosis of OSA. The implementation of the code for this analysis is in Algorithm 1.

Algorithm 1. Logistic Regression for OSA Diagnosis

```

1: data ← LoadData("dataFile.csv")
2: X ← data.drop(['data'])
3: Y ← data['OSA Diagnosis']
4: Xtrain, Xtest, Ytrain, Ytest ← SplitData(X, Y, testSize = 0.2)
5: C values ← [0.001, 0.01, 0.1, 1, 5, 10]
6: for each C in C values do
7:   logreg ← LogisticRegression(C = C)
8:   logreg.fit(Xtrain, Ytrain)
9:   Ypred ← logreg.predict(Xtest)
10:  accuracy ← CalculateAccuracy(Ytest, Ypred)
11:  Print("Accuracy for C=C:", accuracy)
12: end for

```

3.2.2. Random Forest

Random Forest comprises an ensemble of decision trees, where each tree is trained on a random sample of a set of independent variables [21]. This structure reduces a model's variance and prevents overfitting, thereby enhancing the overall predictive performance [22]. The model, which is an ensemble of decision trees, can be represented as follows.

$$f(x) = \sum (c_i \times I(x_i \in R_i)) \quad (3)$$

Here, " c_i " is the average of the response variable in leaf node " i ", " I " is the indicator function, " x_i " represents the independent variables, and " R_i " indicates leaf node " i " of a decision tree. A critical parameter of the random forest model is " $n_estimators$ ", which refers to the number of decision trees within the ensemble. This study assessed the model's performance by adjusting the " $n_estimators$ " values through a range of [10, 50, 100, 200, 500, 1000]. The goal was to identify the optimal " $n_estimators$ " value to achieve a model configuration that optimized the accuracy of OSA diagnosis. The code constructed for this purpose is described in Algorithm 2.

Algorithm 2. STOP-BANG Random Forest

```

1: data ← LoadData("dataFile.csv")
2: X ← data.drop(['data'])
3: Y ← data['OSA Diagnosis']
4: Xtrain, Xtest, Ytrain, Ytest ← SplitData(X, Y, testSize = 0.2)
5: n_estimators ← [10, 50, 100, 200, 500, 1000]
6: for each n in n_estimators do
7:   rf ← RandomForestClassifier(n_estimators = n)
8:   rf.fit(Xtrain, Ytrain)
9:   Ypred ← rf.predict(Xtest)
10:  accuracy ← CalculateAccuracy(Ytest, Ypred)
11:  Print("Accuracy for n_estimators:", accuracy)
12: end for

```

3.2.3. SVM

The Support Vector Machine (SVM) is a powerful machine learning model that belongs to the category of supervised learning and is widely used for both classification and regression problems [23]. This model maps data points to a high-dimensional space and identifies the decision boundary with the maximum margin within this space to classify the data [24]. The core concept of SVM is to find a hyperplane that maximizes the distance (margin) between data points, thereby maximizing the separation between two classes and minimizing the generalization error. The model used was as follows.

$$f(x) = w^T \varphi(x) + b \quad (4)$$

In this context, " w " represents the weight vector that determines the direction of the decision boundary, while " b " is the bias value that adjusts the position of the decision boundary. The function $\varphi(x)$ maps the input vector " x " to a higher-dimensional feature space, which is used to solve nonlinear classification problems. To optimize the performance of the Support Vector Machine, the regularization parameter " C " was adjusted. This parameter balanced the complexity of the model and the penalty for misclassification, aiming to find the optimal parameters. The range of " C " values explored was set to [0.001, 0.01, 0.1, 1, 10, 100], and the implemented settings described in Algorithm 3.

Algorithm 3. Support Vector Machine for OSA Severity Diagnosis

```

1: data ← read csv("stopbang")
2: X ← data.drop(['OSA'])
3: y ← data['OSA']
4: Xtrain, Xtest, ytrain, ytest ← train test split(X, y, test size = 0.2, random state = 42)
5: for C val in [0.001, 0.01, 0.1, 1, 10, 100] do
6:   model ← SV C(C = C val, kernel = 'linear')
7:   model.f it(Xtrain, ytrain)
8:   ypred ← model.predict(Xtest)
9:   cm ← confusion matrix(ytest, ypred)
10:  TN, FP, FN, TP ← cm.ravel()
11:  sensitivity ← TP/(TP + FN)
12:  specificity ← TN/(TN + FP)
13:  accuracy ← accuracy score(ytest, ypred)
14:  Print("Accuracy:", accuracy)
15: end for

```

3.2.4. KNN

The K-Nearest Neighbor (KNN) algorithm, a non-parametric approach, forecasts the classification of a given sample by considering the classes of the K nearest neighbors within its vicinity [25]. The crux of employing this model hinges on the selection of an optimal K value, which significantly impacts the prediction accuracy of the model [26]. The range of K values was established from 1 to 29 for experimental purposes. To achieve the most favorable outcome, the test size parameter was adjusted to 0.3. The implementation code is presented in Algorithm 4.

Algorithm 4. K-Nearest Neighbor for OSA Diagnosis

```

1: data ← read csv("stopbang")
2: X ← data.drop(['OSA'])
3: y ← data['OSA']
4: Xtrain, Xtest, Ytrain, Ytest ← train test split(X, Y, test size = 0.3, random state = 42)
5: for k in [1, . . . , 29] do
6:   model ← KNeighborsClassifier(n neighbors = k)
7:   model.f it(Xtrain, Ytrain)
8:   Ypred ← model.predict(Xtest)
9:   cm ← confusion matrix(Ytest, Ypred)
10:  TN, FP, FN, TP ← cm.ravel()
11:  sensitivity ← TP/(TP + FN)
12:  specificity ← TN/(TN + FP)
13:  accuracy ← accuracy score(Ytest, Ypred)
14:  Ypred proba ← model.predict_proba(Xtest)[: , 1]
15:  auc ← roc auc score(Ytest, Ypred proba)
16:  f1 ← f1 score(Ytest, Ypred)
17:  Print("Accuracy:", accuracy)
18: end for

```

4. Result

For the enhanced STOP-BANG model, various machine learning classification techniques were employed, including Logistic Regression, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbor (K-NN), to assess the performance in diagnosing obstructive sleep apnea. These methods were evaluated based on accuracy, sensitivity, specificity, and F1 score, aiding in determining the most effective model for obstructive sleep apnea diagnosis. The significant outcomes of this analysis are summarized in Tables 3–6, showcasing the results for each classification technique.

Table 3. Logistic Regression result.

C value	0.01	0.1	1	5
Accuracy	0.87	0.94	0.91	0.92
Sensitivity	1	0.96	0.94	0.96
Specificity	0	0.42	0.71	0.71
F1 Score	0	0.59	0.81	0.81

Table 4. Random Forest result.

n_estimators	10	50	100	200	500
Accuracy	0.84	0.84	0.85	0.84	0.85
Sensitivity	0.86	0.88	0.88	0.9	0.9
Specificity	0.71	0.57	0.71	0.42	0.57
F1 Score	0.78	0.69	0.78	0.58	0.69

Table 5. SVM result.

C value	0.1	1	10	100
Accuracy	0.87	0.87	0.92	0.92
Sensitivity	1	1	0.96	0.96
Specificity	0	0	0.71	0.71
F1 score	0.81	0.81	0.92	0.92

Table 6. K-NN result.

K value	7	9	11	13
Accuracy	0.92	0.92	0.94	0.96
Sensitivity	0.71	0.71	0.85	0.71
Specificity	0.89	0.89	0.92	0.92
F1 Score	0.96	0.93	0.95	0.96

The results for each model are as follows. When comparing the best-performing model with actual data and control group data, the comparison yielded in Table 7.

Table 7. Evaluation of enhanced STOP-BANG models for each classification algorithm.

	Collected Data	Comparison Data	K-NN (K Value 11)	Logistic Regression (C = 5)	SVM (C = 10)	Random Forest (n_estimators = 100)
Accuracy	0.86	0.86	0.92	0.92	0.92	0.85
Sensitivity	0.89	0.89	0.94	0.96	0.96	0.88
Specificity	0.61	0.32	0.83	0.71	0.71	0.71

Figure 1 showed that Logistic Regression varied in performance, with changes in the regularization strength C , particularly excelling at $C = 5$, where it achieved an accuracy of 0.92, sensitivity of 0.96, specificity of 0.71, and F1 score of 0.81. The Random Forest model displayed consistent performance with $n_estimators = 100$, resulting in an accuracy of 0.85, sensitivity of 0.88, specificity of 0.71, and F1 score of 0.78. The SVM model stood

out in terms of sensitivity and F1 score, showing an accuracy of 0.92, sensitivity of 0.96, specificity of 0.71, and F1 score of 0.92 at $C = 10$. The most significant findings came from the K-NN algorithm, which, at $K = 11$, recorded the highest performance metrics among all models used in this study, with an accuracy of 0.94, sensitivity of 0.94, specificity of 0.83, and F1 score of 0.95. This suggests that the K-NN model is a meaningful option for OSA diagnosis using the STOP-BANG questionnaire, capable of improving accuracy, sensitivity, and specificity simultaneously. A closer look at the results confirms that among the machine learning classification models applied with the STOP-BANG questionnaire and OSA diagnostic data, the K-NN model notably offers high sensitivity and specificity. Specifically, the model's specificity demonstrated a significant improvement, moving from a specificity of 0.61 in the research data and 0.47 in the control group data to 0.83, thereby proving how applying machine learning models to OSA diagnoses using the STOP-BANG questionnaire can effectively enhance sensitivity and specificity.

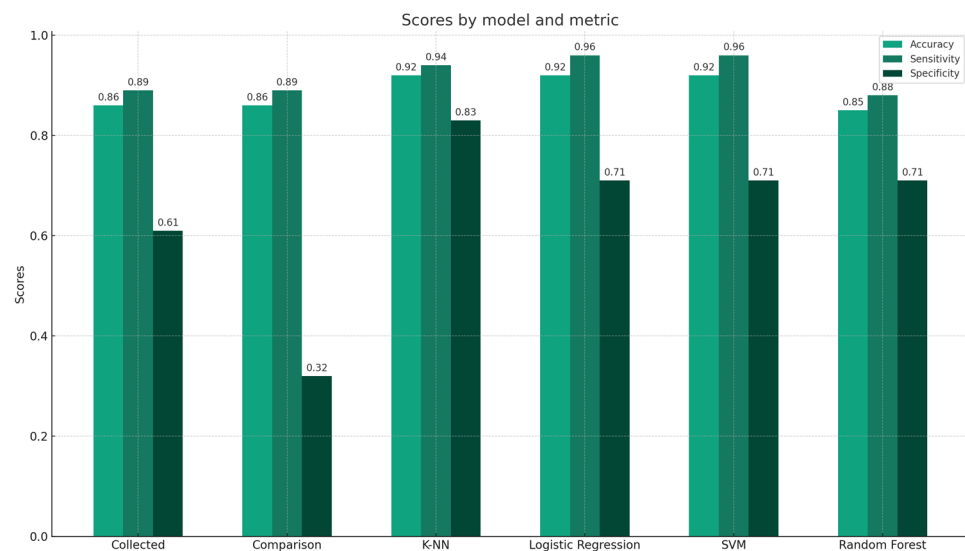


Figure 1. Evaluation of optimal models.

5. Discussion

Following the analysis of our model results, we conducted an in-depth evaluation of the correlations between obstructive sleep apnea (OSA) and various physiological and demographic factors, as depicted in Figure 2.

Our statistical analysis unveiled several noteworthy associations with the diagnosis of OSA. The correlation matrix heatmap provides a vivid visual representation of these relationships. A significant positive correlation emerged between snoring intensity and OSA diagnosis ($r = 0.38$), reinforcing snoring as a predominant symptomatic predictor of OSA. Conversely, the correlation between self-reported tiredness and OSA diagnosis was found to be weak ($r = 0.12$), suggesting that while tiredness is a common complaint among individuals with OSA, it may not serve as a definitive indicator of the disorder. Notably, the presence of observed apnea episodes demonstrated a significant positive correlation with OSA diagnosis ($r = 0.28$), highlighting its importance as a critical clinical sign of OSA. Moreover, we observed a moderate positive correlation between blood pressure and OSA diagnosis ($r = 0.21$), indicating a potential link that might contribute to the observed comorbidity in patients with OSA. Furthermore, BMI showed a positive correlation with OSA diagnosis ($r = 0.13$). Although BMI is a well-established risk factor, its predictive value seems less pronounced compared to other physical indicators within our cohort. Additionally, our study identified a positive correlation between age and OSA diagnosis ($r = 0.14$), implying that the likelihood of being diagnosed with OSA slightly increases with age in our study population. Neck circumference also revealed a moderate positive correlation with OSA diagnosis ($r = 0.23$), underscoring its relevance as a measurable risk

factor for OSA. Furthermore, gender exhibited a positive correlation with OSA diagnosis ($r = 0.22$), aligning with previous research that indicates a gender disparity in the prevalence of OSA.

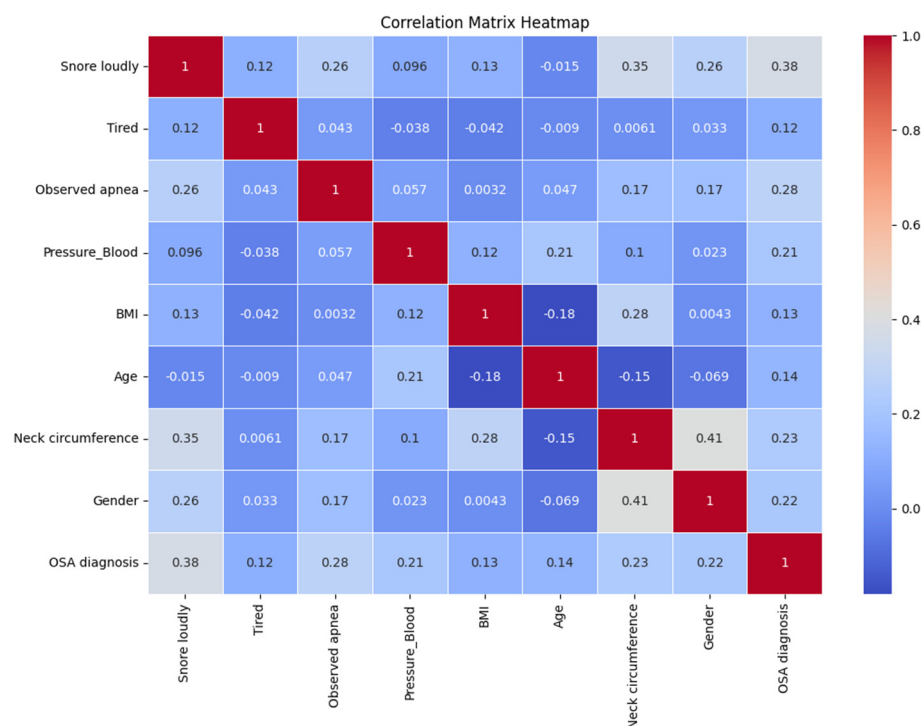


Figure 2. Correlation matrix heatmap.

In the landscape of OSA diagnostics, the deployment of ML models signifies a pivotal shift towards precision medicine. Our study's comparative analysis across Logistic Regression, Random Forest, SVM, and K-NN algorithms underscores the critical role of model selection, influenced by nuanced diagnostic criteria and the inherently complex nature of OSA. Particularly, the superior performance metrics of the K-NN algorithm highlight its efficacy as an OSA screening tool. This observation necessitates further exploration into optimizing distance metrics and weighting methodologies to augment diagnostic accuracy. Additionally, the observed variability in the performance of logistic regression, contingent on the regularization parameter C , emphasizes the importance of precise model tuning. This variability highlights the need for a thorough exploration of regularization techniques to mitigate model bias and variance, especially within the intricate context of medical data relationships. The SVM model's notable sensitivity and F1 score at optimal C values suggest its adeptness in distinguishing between patients with OSA and control groups, prompting further investigation into diverse kernel functions to refine the SVM's application in OSA diagnostics and enhance its predictive precision. Although the Random Forest model exhibits consistent performance, a deeper examination of hyperparameters is warranted to achieve a balance between model complexity and robustness, facilitating its practical application in clinical settings. In line with this perspective, the objective of this study was primarily focused on enhancing specificity, which led to the highest evaluation of the K-NN model. However, it is evident that there is a need to explore the viewpoints of other models.

The cross-sectional design of this study, while providing initial insights, inherently limits the capacity for causal inference. Thus, we advocate for longitudinal research, incorporating temporal and possibly real-time monitoring data, to elucidate these models' predictive capabilities over time. The variability in model correlations across different demographics underscores the imperative for rigorous validation in varied clinical settings to ensure the generalizability and equity of ML-based diagnostic tools. Coupled with the

ML findings, our study's correlation analysis advocates for a multifaceted approach to the accurate diagnosis and management of OSA. It underscores the necessity of considering factors like snoring intensity, observed apnea episodes, and neck circumference in the development of diagnostic tools and intervention strategies. The relatively weaker correlations among tiredness, BMI, and OSA diagnostics suggest a composite evaluation of these indicators alongside more significant predictors. Recognizing the limitations of this study, its cross-sectional nature precludes definitive causal deductions. Longitudinal investigations are paramount to unravel the intricate nature of these associations over time. Furthermore, the observed variability in correlation strengths across different populations highlights the necessity of replicating this study in diverse demographic environments.

In summary, our findings make a significant contribution to the evolving understanding of the multifactorial etiology of OSA. They emphasize the need for a comprehensive evaluation of diagnostic and therapeutic factors. Further research, particularly employing longitudinal study designs, is essential for delving into these associations in greater detail, ultimately aiming to enhance OSA diagnosis and treatment modalities.

6. Conclusions

This study represents a significant advancement in the early diagnosis and management of obstructive sleep apnea (OSA), a condition affecting a vast global population and critically impacting patients' health and quality of life. Traditional diagnostic methods for OSA are often time-consuming and costly, leading to a considerable number of patients not receiving timely and appropriate diagnoses. To overcome this challenge, we propose an integrated approach that combines a machine learning model with the STOP-BANG questionnaire. This method represents a significant step forward in the early diagnosis and effective management of OSA. The results of our study indicate that utilizing a machine learning algorithm, specifically the K-Nearest Neighbor (K-NN) model, can significantly improve the accuracy, sensitivity, and specificity of OSA diagnoses made using the STOP-BANG questionnaire. These findings highlight the potential of machine learning technologies in the medical field, particularly for the early detection and categorization of diseases. The application of machine learning models like K-NN has the potential to enhance early diagnostic procedures for various medical conditions.

The practical implications of this study can be viewed from two perspectives. Firstly, the combined use of the STOP-BANG questionnaire and a machine learning model offers an effective method to enhance the precision of OSA diagnostics while reducing associated time and costs. This approach is expected to improve the efficiency of medical services and patient satisfaction. Secondly, this integrated approach opens new horizons for employing artificial intelligence (AI) technology in the medical sector. The application of machine learning models in the early detection and treatment of various illnesses has the potential to revolutionize disease prevention and management. Further research should aim to expand the scope of this inquiry by exploring a broader range of machine learning models, including deep learning algorithms, which could provide more detailed insights into OSA diagnosis.

In conclusion, our study demonstrates that integrating the STOP-BANG questionnaire with a machine learning model can have a positive impact on the early diagnosis and management of OSA. This suggests that the implementation of AI technology in the medical field can lead to beneficial outcomes, including reduced medical expenses, shorter diagnosis and treatment times, and improved patient quality of life. However, the journey towards optimizing the use of AI in medicine is ongoing. It requires the continuous exploration, validation, and adaptation of emerging technologies to meet the changing needs of patients and health care providers.

Author Contributions: M.-S.C. and D.-H.H.: Writing—original draft, Data curation, and Software. J.-W.C.: Visualization. M.-S.K.: Conceptualization, Validation, Writing—review and editing, Project administration. M.-S.C.: Formal analysis, Funding acquisition. D.-H.H. is the first author. All authors have read and agreed to the published version of the manuscript.

Funding: This paper was supported by Eulji University in 2023 (EJRG-23-16).

Institutional Review Board Statement: This study was conducted in accordance with and approved by the Institutional Review Board of Eulji Medical Center (protocol code 2022-03-010-002) for studies involving humans.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy and ethical restrictions.

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

References

1. Kumar, V.M. Sleep and sleep disorders. *Indian J. Chest Dis. Allied Sci.* **2008**, *50*, 129–135. [[PubMed](#)]
2. Kim, H.C. The importance of diagnosis and treatment of snoring and obstructive sleep apnea and the role of dentists. *J. Korean Dent. Assoc.* **2010**, *48*, 178–183.
3. Kim, J.K.; Kim, H.J. Clinical assessment for obstructive sleep apnea and national health insurance criteria for polysomnography. *J. Korean Med. Assoc.* **2020**, *63*, 410–419. [[CrossRef](#)]
4. Chiu, H.Y.; Chen, P.Y.; Chuang, L.P.; Chen, N.H.; Tu, Y.K.; Hsieh, Y.J.; Wang, Y.C.; Guilleminault, C. Diagnostic accuracy of the Berlin questionnaire, STOP-BANG, STOP, and Epworth sleepiness scale in detecting obstructive sleep apnea: A bivariate meta-analysis. *Sleep Med. Rev.* **2017**, *36*, 57–70. [[CrossRef](#)] [[PubMed](#)]
5. Mashaqi, S.; Staebler, D.; Mehra, R. Combined nocturnal pulse oximetry and questionnaire-based obstructive sleep apnea screening—A cohort study. *Sleep Med.* **2020**, *72*, 157–163. [[CrossRef](#)] [[PubMed](#)]
6. Nagappa, M.; Liao, P.; Wong, J.; Auckley, D.; Ramachandran, S.K.; Memtsoudis, S.; Mokhlesi, B.; Chung, F. Validation of the STOP-BANG questionnaire as a screening tool for obstructive sleep apnea among different populations: A systematic review and meta-analysis. *PLoS ONE* **2015**, *10*, e0143697. [[CrossRef](#)]
7. Orbea, C.A.P.; Lloyd, R.M.; Faubion, S.S.; Miller, V.M.; Mara, K.C.; Kapoor, E. Predictive ability and reliability of the STOP-BANG questionnaire in screening for obstructive sleep apnea in midlife women. *Maturitas* **2020**, *135*, 1–5. [[CrossRef](#)]
8. Chung, F.; Yegneswaran, B.; Liao, P.; Chung, S.A.; Vairavanathan, S.; Islam, S.; Khajehdehi, A.; Shapiro, C.M. STOP Questionnaire: A Tool to Screen Patients for Obstructive Sleep Apnea. *Anesthesiology* **2008**, *108*, 812–821. [[CrossRef](#)] [[PubMed](#)]
9. Farney, R.J.; Walker, B.S.; Farney, R.M.; Snow, G.L.; Walker, J.M. The STOP-BANG Equivalent Model and Prediction of Severity of Obstructive Sleep Apnea: Relation to Polysomnographic Measurements of the Apnea/Hypopnea Index. *J. Clin. Sleep Med.* **2011**, *7*, 459–465. [[CrossRef](#)]
10. Yang, H.; Park, H. Usefulness of the Berlin, STOP, and STOP-BANG Questionnaires in the Diagnosis of Obstructive Sleep Apnea. *J. Sleep Med.* **2019**, *16*, 11–20. [[CrossRef](#)]
11. Rajendra, P.; Latifi, S. Prediction of diabetes using logistic regression and ensemble techniques. *Comput. Methods Programs Biomed. Update* **2021**, *1*, 100032. [[CrossRef](#)]
12. Holfinger, S.J.; Lyons, M.M.; Keenan, B.T.; Mazzotti, D.R.; Mindel, J.; Maislin, G.; Magalang, U.J. Diagnostic performance of machine learning-derived OSA prediction tools in large clinical and community-based samples. *Chest* **2022**, *161*, 807–817. [[CrossRef](#)]
13. Javeed, A.; Berglund, J.S.; Dallora, A.L.; Saleem, M.A.; Anderberg, P. Predictive power of XGBoost_BiLSTM model: A machine-learning approach for accurate sleep apnea detection using electronic health data. *Int. J. Comput. Intell. Syst.* **2023**, *16*, 188. [[CrossRef](#)]
14. Shi, Y.; Zhang, Y.; Cao, Z.; Ma, L.; Yuan, Y.; Niu, X.; Ren, X. Application and interpretation of machine learning models in predicting the risk of severe obstructive sleep apnea in adults. *BMC Med. Inform. Decis. Mak.* **2023**, *23*, 230. [[CrossRef](#)]
15. Han, H.; Oh, J. Application of various machine learning techniques to predict obstructive sleep apnea syndrome severity. *Sci. Rep.* **2023**, *13*, 6379. [[CrossRef](#)]
16. Bazoukis, G.; Bollepalli, S.C.; Chung, C.T.; Li, X.; Tse, G.; Bartley, B.L.; Armoundas, A.A. Application of artificial intelligence in the diagnosis of sleep apnea. *J. Clin. Sleep Med.* **2023**, *19*, 1337–1363. [[CrossRef](#)]
17. Sateia, M.J. International classification of sleep disorders-third edition. *Chest* **2014**, *146*, 1387–1394. [[CrossRef](#)]
18. Hwang, M.; Zhang, K.; Nagappa, M.; Saripella, A.; Englesakis, M.; Chung, F. Validation of the STOP-BANG questionnaire as a screening tool for obstructive sleep apnea in patients with cardiovascular risk factors: A systematic review and meta-analysis. *BMJ Open Respir. Res.* **2021**, *8*, e000848. [[CrossRef](#)] [[PubMed](#)]
19. Nusinovi, S.; Tham, Y.C.; Yan, M.Y.C.; Ting, D.S.W.; Li, J.; Sabanayagam, C.; Wong, T.Y.; Cheng, C.Y. Logistic regression was as good as machine learning for predicting major chronic diseases. *J. Clin. Epidemiol.* **2020**, *122*, 56–69. [[CrossRef](#)] [[PubMed](#)]
20. Christodoulou, E.; Ma, J.; Collins, G.S.; Steyerberg, E.W.; Verbakel, J.Y.; Van Calster, B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J. Clin. Epidemiol.* **2019**, *110*, 12–22. [[CrossRef](#)]
21. Rigatti, S.J. Random forest. *J. Insur. Med.* **2017**, *47*, 31–39. [[CrossRef](#)] [[PubMed](#)]
22. Biau, G.; Scornet, E. A random forest guided tour. *Test* **2016**, *25*, 197–227. [[CrossRef](#)]

23. Soman, K.P.; Loganathan, R.; Ajay, V. *Machine Learning with SVM and Other Kernel Methods*; PHI Learning Pvt. Ltd.: New Delhi, India, 2009.
24. Osisanwo, F.Y.; Akinsola, J.E.T.; Awodele, O.; Hinmikaiye, J.O.; Olakanmi, O.; Akinjobi, J. Supervised machine learning algorithms: Classification and comparison. *Int. J. Comput. Trends Technol. IJCTT* **2017**, *48*, 128–138.
25. Zhang, Z. Introduction to machine learning: K-nearest neighbors. *Ann. Transl. Med.* **2016**, *4*, 218. [[CrossRef](#)]
26. Guo, G.; Wang, H.; Bell, D.; Bi, Y.; Greer, K. KNN model-based approach in classification. In *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE, Proceedings of the OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, 3–7 November 2003*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 986–996.

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.