Novel Model for Detecting Sleep Apnea through Sleep Parameters

Maria Johnson[1]*, Dr. Durgansh Sharma[2], Dr. Smriti Mathur[3]

[1,2,3] School of Business and Management, CHRIST (Deemed to be University), Bangalore, India

[1]Email: mariamjk1627@gmail.com

Abstract

This paper addresses the prevalence and health risks of Obstructive Sleep Apnea (OSA) in India, emphasizing the need for improved diagnostic techniques. It explores the use of Machine Learning (ML) models for OSA screening and the impact of feature selection on model performance. Using a dataset from Kaggle, the study developed classification models such as Logistic Regression, Decision Tree, Random Forest, k-Nearest Neighbour and Support Vector Machine for OSA detection. The performance was measured after selecting features through ranking, forward selection and backward elimination. The features we selected were blood pressure, BMI category and occupation. With the selected features, SVM performed the best with an accuracy of 95.80%. With forward selection, Decision Tree performed the best with an accuracy of 94.54% whereas backward elimination did not give optimum results. The models can be trained and validated for OSA detection using diverse data from wearable devices and electronic health records.

Keywords: *Obstructive Sleep Apnea, Detection, Feature Selection, Support Vector Machine, Decision Tree, Optimized Model*

1. Introduction

According to a study conducted by pulmonary and medicine experts at AIIMS, 11% of adult Indians have OSA, with men having a higher risk (13% versus 5% for women). As per the AIIMS study, obstructive sleep apnea (OSA), which they described as "a common sleep disorder associated with considerable morbidity," affects over 100 million people in India [1]. Obstructive sleep apnea (OSA) is the medical term for an upper airway blockage that causes an obstruction of airflow while making a simultaneous breathing attempt. [2]. This indicates that severe OSA may result in grave health problems or even death because the patient is unable to get adequate oxygen and may even die while they are asleep.

The most common symptoms of OSA include morning headaches, dry mouth upon awakening, loud snoring, and difficulty focusing. [3]. Polysomnography (PSG), an overnight sleep study that analyses a range of physiological factors, including brain waves, oxygen levels, and breathing patterns, is the gold standard for diagnosing OSA [4].Three categories of sleep apnea are: Mild, moderate and severe. A six-fold increased risk of all-cause death was found to be independently linked to moderate-to-severe sleep apnea compared to no sleep apnea. There was no link found between mild sleep apnea and high mortality. However, PSG is expensive and time-consuming, and it is not always accessible to patients. If these instruments are hard to come by, waiting lists may be lengthy. The demand for straightforward and reliable diagnostic techniques grows as the number of patients with suspected OSAS rises [5].

In an age of unprecedented technological innovation and global health challenges, the intersection of sleep health and medical research has never been more vital. Machine learning (ML) has become integral in the field of healthcare, with a growing emphasis on using predictive models to enhance the screening process for Obstructive Sleep Apnea (OSA). This shift towards ML has resulted in improved sensitivity and accuracy, leading to more effective diagnoses and treatments. In a recent study by [6],
the utilization of ML and AI in healthcare has experienced significant growth. Various classification methods were employed to develop predictive models for OSA screening, including Decision Tree (DT), Logistic Regression (LR), Support Vector Machine (SVM), Naive Bayes (NB), Linear Discriminant Analysis (LDA), k-Nearest Neighbor (k-NN), and subspace discriminant (Ensemble) classifiers. [7] conducted an extensive investigation, utilizing a dataset consisting of 31 features, including demographic characteristics such as race, age, sex, BMI, and snoring habits, to determine the most accurate classifier for OSA screening. The results showcased that the k-NN classifier outperformed the other tested classifiers in terms of accuracy, highlighting its capability as a reliable tool for OSA diagnosis. To further enhance the effectiveness of the k-NN classifier, a wrapper feature selection method based on Binary Particle Swarm Optimization (BPSO) was employed. This technique aimed to identify the most relevant features associated with OSA, allowing for a more refined and accurate screening process.

In this paper, we delve into a detailed comparative analysis of five classification models using various feature selection methods. By examining the performance of Logistic Regression, Decision Tree, Random Forest, kNN, and SVM, we aim to uncover the most effective classifier model in combination with optimal feature selection techniques. To conduct this analysis, we utilized a publicly available dataset from Kaggle. This dataset was carefully curated to ensure its relevance and reliability. Our primary focus was to investigate the impact of feature selection on the classification models' performance. Two feature selection techniques, namely forward selection and backward elimination were applied to extract the most significant features for each model. To gauge the performance of these classification models accurately, we employed several evaluation metrics: accuracy, precision, recall and F1-score.

The proposed model can then be used for commercial purpose by using data collected from IoT devices (Wrist bands), applying the classification model, predicting the results of different treatment options, enhancing and assessing treatment, and personalize treatment with understanding of underlying mechanisms of OSA [2]. This result can be send to family members, consulting doctors or hospitals. They can then provide personalised treatment to the person. The device would be more economical, comfortable, and convenient than existing systems not only for patients but also for doctors.

2. Review Literature

2.1 Sleep Apnea

When a person has obstructive sleep apnea (OSA), their breathing is interrupted while they are asleep due to a contraction of their upper airway. This can occur hundreds of times a night and can cause a variety of health issues, such as stroke and cardiovascular disease. The collapse of the airway is due to an imbalance in the activity of the diaphragm and the upper airway muscles. Instead of creating productive airflow during inspiration, this leads the airway to collapse. Eventually, the body wakes up from sleep to resume breathing. There is an elevation in blood pressure and heart rate during each apnea. The rise in sympathetic tone, a measurement of the sympathetic nervous system's activity, is responsible for this. The sympathetic nervous system triggers the body's "fight or flight" response. Heart attacks, strokes, and high blood pressure are some of the cardiovascular issues associated with an increase in sympathetic tone. It is believed that sleep apnea poses an independent risk for these ailments [8] [9]. Polysomnography (PSG) test is used to diagnose OSA [4]. Other techniques, including overnight O_2 pulse oximetry or cardiorespiratory monitoring, are utilised to streamline the diagnosis of OSAS [5].
According to statistics, the prevalence of Sleep Apnea Syndrome (SAS) rises with population age, approximately 4% for males and 2% for females worldwide. For instance, 20% to 40% of elderly over 65 suffer from SAS [10]. The Apnea-Hypopnea Index (AHI) is typically used to gauge how severe a condition of apnea is [7]. [11] studied the Apnea-Hypopnea Index (AHI) and polysomnographic risk factors for predicting 5- to 8-year mortality in patients with OSA. They found that AHI and age were significant predictors of mortality. [12] conducted a study to investigate the relationship between age, comorbidities, and mortality in patients with severe OSA. The study found that age and comorbidities were both significant predictors of mortality. The risk of mortality increased with age and with the number of comorbidities.
In another study, Sleep apnea was identified by the respiratory disturbance index (RDI), which is the number of breathing pauses per hour of sleep.The participant's anticipated hours of sleep is divided by the total number of respiratory troubles to produce the Respiratory Disturbance Index (RDI), which is then calculated as an incident rate per hour. Three categories based on RDI are: no or subclinical (RDI < 5), mild (RDI 5 to < 15), and moderate-to-severe (RDI ≥ 15). The study also found that sleep apnea is an independent risk factor for all-cause mortality [13]. Obstructive Sleep Apnea (OSA) is a widespread medical illness with serious adverse effects; however, many people go undetected with OSA [14].

[15] reviewed OSA in adults. They discussed the pathophysiology, symptoms, diagnosis, and management of OSA. They also emphasised the importance of early diagnosis and prescription of OSA for improving patient outcomes. 47 million Indians with working age OSA have moderate-to-severe cases, out of an estimated 104 million cases. This is a significant public health issue in India that has implications for the disease's worldwide burden [1].
OSA's treatment options include Continuous Positive Airway Pressure (CPAP), Oral Appliances (OAs), Myofunctional Therapy (MT), behavioural modification, and surgical procedures.

2.2 Machine Learning

[2] studied the role of Artificial Intelligence (AI) in the treatment of OSA. They concluded that AI can improve the diagnosis and treatment of OSA by developing more accurate and efficient screening tools, personalizing treatment plans, and monitoring treatment response. A study achieved a high accuracy of 99.80% when detecting sleep apnea using RR interval (the time between two successive R waves of the QRS(Q,R, S waves- graphical defections) signal on the electrocardiogram (ECG)) data using a Long Short-Term Memory (LSTM) neural network [16]. For example, a study by [7] found that a k-nearest neighbor (kNN) classifier was able to achieve an accuracy of 99.80% in diagnosing OSA using RR interval data. Based on the electrocardiograph (ECG), [17] created a real-time algorithm for the diagnosis of sleep apnea. For the purpose of detecting OSA, support vector machine (SVM) and ensemble support vector machine approaches were used. The ensemble classification process obtained 95% accuracy, 95% sensitivity, and 95% specificity.
Heart Rate Variability, or HRV, is commonly acknowledged as a characteristic of OSA. For the purpose of detecting OSA by HRV, the paper by [18] proposed an ensemble learning based classifier – Random Forest and a machine learning based classifier – Support Vector Machine (SVM). A tenfold cross validation test was run, and the maximum accuracy for identifying OSA cases was 94%.
Another study examined how effectively clinical measurements, subjective questionnaires, anatomical scores, and associated comorbidities might be utilised to predict the severity of OSA using the SVM algorithm [19]. For assessing mild to moderate OSA, the SVM model demonstrated recall and precision scores of 0.80 and 0.93, respectively; however, the results were 0.93 and 0.81 for identifying severe OSA [20]. A Support Vector Machine (SVM)-based algorithm for OSA detection using ECG signal features was developed. They found that the SVM algorithm was able to achieve an accuracy of 93.2% in detecting OSA.
[6] reviewed the application of ML as a screening tool for OSA diagnosis. They found that ML algorithms can be used to develop accurate screening tools for OSA using a variety of physiological data, including demographic information, medical history, and sleep symptoms. Another study by [21] developed a deep learning model for classification of sleep stages using PSG signals. The model was able to achieve an accuracy of 98.2% in classifying sleep stages, which suggests that it could be used to develop a wearable device that could monitor sleep stages and identify OSA events.
A sample of adult Saudi Arabians was studied for OSA symptoms and risk [3]. They discovered that weariness, snoring, and daytime sleepiness were the most typical signs of OSA.

Additionally, they discovered that male sex and obesity constituted important risk factors for OSA. Using faciocervical measures obtained from a Chinese population, [22] carried out a cross-sectional study to create a Support Vector Machine (SVM) model for screening for moderate to severe OSA. With an accuracy rate of 85.7%, the model may find application in the creation of an OSA screening wearable.
OSA is a serious illness that can cause a variety of issues with health, including death. It is critical to treat OSA as soon as possible to lower the chance of severe consequences. Machine learning has the potential to modernise the diagnosis and treatment of OSA by developing more efficient and affordable methods for diagnosis and monitoring.

2.3 Feature Selection

SVM, Logistic Regression (LR), Linear Discriminant Analysis (LDA), and k-NN are a few of the classification methods used to select key features.
In a study, the optimal feature subset in a wrapper setting was chosen using a sequential forward selection. Using 10 features, the kNN algorithm achieved 88% accuracy, 85% sensitivity, and 90% specificity, whereas the NN method achieved 88% accuracy, 89% sensitivity, and 86% specificity [23].
[7] developed an ML algorithm for OSA diagnosis using feature selection- Binary Particle Swarm Optimization (BPSO), classification methods, and data grouping based on age, sex, and race. They found that kNN classifier was able to achieve an accuracy of 99.80% in diagnosing OSA.

2.4 Internet of Things

Connecting doctors and patients with smart devices when they are both travelling freely is the key to understanding how IoT may improve healthcare. Using big data technologies, cloud services can be used to upload patient data, and the transferred data can subsequently be examined. Internet-connected medical devices become the primary component of the healthcare system [24]. The "Apnea MedAssist Service" is an IoT device that was developed that harvests ECG data from the user and analyses the output from the categorization model-(SVM-) based system [20]. [25] developed an IoT-based E-Health and Sleep Apnea Monitoring System (IEHSAM). The IEHSAM system uses wearable sensors to collect physiological data, which is then analyzed using ML algorithms to detect OSA. A study by [10] developed a machine learning assisted wearable wireless device for OSA diagnosis that was able to achieve an accuracy of 95% using a combination of ECG, respiratory effort, and snoring data.
Additionally, medical body area networks (MBANs), which are networks of wearable devices on the patient's body, connect to a controller that is not attached via a wireless communication link. In order to diagnose the patient, this MBAN is used to monitor and evaluate the patient's physiological parameters in addition to other data. A threat intelligence model for a body area network (BAN) built on Internet of Things (IoT) sensors can be used to achieve this [26]. [24] discussed the use of the Internet of Things (IoT) to develop wireless body area networks (WBANs) for healthcare applications.

They noted that WBANs could be used to collect physiological data from patients with OSA, which could then be analyzed using ML algorithms to detect and monitor OSA.

3. Research Methodology

The software tool RapidMiner (https://rapidminer.com/) has been used to design the machine learning predictive models in this work.

Figure 1: The proposed methodology. The figure illustrates the step-by-step process of the proposed methodology, which involved five steps: data collection, data preprocessing, feature selection, classification and evaluation.

4. Dataset

The data analyzed in the present research is taken from the Kaggle website. The Sleep Health and Lifestyle Dataset comprises 374 rows and 13 columns. It includes the features such as gender, age(in years), occupation, sleep duration(in hours), quality of sleep (scale: 1-10), physical activity level (minutes/day), stress levels(scale: 1-10), BMI category(Underweight, Normal, Overweight), blood pressure, heart rate(beats per minute), daily steps, and Sleep disorder(None, Insomnia, Sleep Apnea).

Table 1: Descriptive Statistics of Data

	Age	*Sleep Duration*	*Quality of Sleep*	*Physical Activity Level*	*Stress Level*	*Blood Pressure Systolic*	*Blood Pressure Diastolic*	*Daily Steps*
Mean	42.184	7.132	7.313	59.171	5.385	128.553	84.650	6816.845
Median	43.000	7.200	7.000	60.000	5.000	130.000	85.000	7000.000
Mode	43.000	7.200	8.000	60.000	3.000	130.000	80.000	8000.000
Standard Deviation	8.673	0.796	1.197	20.831	1.775	7.748	6.162	1617.916
Minimum	27.000	5.800	4.000	30.000	3.000	115.000	75.000	3000.000
Maximum	59.000	8.500	9.000	90.000	8.000	142.000	95.000	10000.000

5. Preprocessing

5.1 Nominal to Binomial

The dependent variable – "Sleep disorder" is a polynomial label attribute. It contains polynomial data i.e. it has more than 2 categories of data, namely, Sleep Apnea, Insomnia, and None. Other attributes such as BMI category is also polynomial with categories – Normal, Obese, Overweight. The modeling operators such as Logistic Regression (LR) and Support Vector Machine (SVM), used in the study cannot handle polynomial attributes. So we have used "Nominal to Binomial" operator to convert the polynomial data to binomial data.

5.2 Nominal to Numerical

Each operator has particular capabilities for dataset handling. Some learners such as Support Vector Machine (SVM) can only handle numerical attributes and cannot learn from nominal attributes. To transform the data, we have used "Nominal to Numerical" to transform the nominal data to numerical data.

6. Feature Ranking

Following method of feature ranking have been used in the study to understand the importance of each feature in determining the value of the target variable – "Sleep Disorder".

6.1 Information Gain: The weight of attributes is calculated by Weight by Information Gain operator, with respect to the class attribute by using the information gain. Higher the weight of an attribute, more is the relevance.

Information Gain = Entropy(parent) – [average entropy(children)]
(1)

where:

Entropy(parent) is a measure of the uncertainty or impurity of the node.

Average entropy(children) is the average entropy of the child nodes, which is the weighted sum of the entropies of each child node, where the weight is the proportion of examples that fall into that child node.

Shannon entropy: This is the most general type of entropy and is used in information theory. It is defined as the average amount of information required to encode a random variable. The formula for Shannon entropy is:

$H(X) = -\sum p(x) \log2(p(x))$
(2)

where:
$H(X)$ = Entropy of the random variable X
$p(x)$ = Probability of event x
$\log2(x)$ = Base-2 logarithm of x

In other words, Information Gain is the reduction in entropy that is achieved by splitting a set of examples into smaller subsets. The higher the Information Gain, the more information is gained by splitting the set, and the more useful the split is for building a decision tree.

Table 2: Features ranked by weight by Information Gain

Attributes	**Weights**
Blood Pressure	0.825
Occupation	0.710
BMI Category	0.548
Sleep Duration	0.248
Age	0.238
Physical Activity Level	0.220
Daily Steps	0.208
Heart Rate	0.168
Quality of Sleep	0.156
Stress Level	0.116
Gender	0.113

6.2 Chi Squared Statistics*:* Using the chi-squared statistic, the Weight by Chi Squared Statistic operator determines the weight of attributes in relation to the class attribute. An attribute is taken into consideration more frequently the higher its weight.
The chi-squared statistic is calculated as follows:

$$\chi^2 = \Sigma (O - E)^2 / E \quad (3)$$

where:
O is the observed frequency for a particular category
E is the expected frequency for that category
Σ is the summation symbol

Observed frequency is the actual number of times a particular event or outcome occurs in an experiment or real-life situation.
Expected frequency is the theoretical number of times an event or outcome is expected to occur based on a specific hypothesis or assumption. It is calculated using the formula:

$$\text{Expected frequency} = \text{Probability of event} * \text{Total number of trials} \quad (4)$$

Table 3: Features ranked by weight by Chi Squared statistics

Attributes	**Weights**
Blood Pressure	453.153
Occupation	421.363
Daily Steps	384.364
Sleep Duration	354.682
Age	340.544
Physical Activity Level	309.329
Heart Rate	249.995
BMI Category	246.968
Stress Level	240.199
Quality of Sleep	156.960
Gender	54.306

6.3 Relief*:* Relief's fundamental principle is to assess feature quality by measuring how well its values distinguish examples of similar and different classes that are situated close to one another. The training examples are iterated over by the Relief algorithm. It finds the closest neighbour in the same class (referred to as the "nearest hit") and the closest neighbour in the opposite class (referred to as the "nearest miss") for each sample. It then changes the weight of each attribute according to the difference in value between the target instance and the nearest hit and the nearest miss.

$$W\_i = W\_i - \text{diff}(X[i], H[i]) + \text{diff}(X[i], M[i]) \quad (5)$$

where:
W_i = weight of attribute i
X[i] = Measure of attribute i for the target example
H[i] = Measure of attribute i for the nearest hit
M[i] = Measure of attribute i for the nearest miss
diff(x, y) = Difference between x and y

Table 4: Features ranked by weight by Relief

Attributes	**Weights**
Blood Pressure	4.104
BMI Category	3.647
Occupation	3.034
Physical Activity Level	1.342
Sleep Duration	1.206
Stress Level	1.188
Gender	1.089
Quality of Sleep	1.058
Age	1.052
Heart Rate	0.957
Daily Steps	0.739

The features have been sorted in descending order of their respective weights.

7. Classification Models for decision making

For the comparison analysis in the study, the following classification models were used.

7.1 Logistic Regression

For classification problems where the objective is to predict the probability that an instance belongs to a specific class or not, one supervised machine learning approach that is utilised is called logistic regression. This type of statistical method examines how a collection of independent variables and a set of dependent binary variables relate to one another.
The idea of the threshold value, which indicates the likelihood of either 0 or 1, is used in logistic regression. The anticipated values are converted to probabilities using the sigmoid function. It converts any real number between 0 and 1 into another value. The logistic regression's result must lie between 0 and 1, and as it cannot be greater than this, it takes the shape of a "S" curve, called the Sigmoid function or the logistic function.

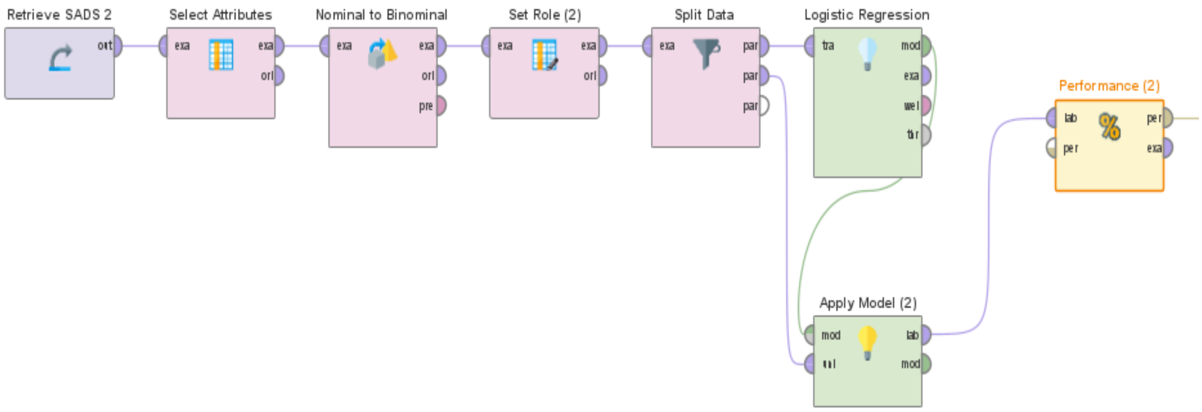

Figure 2 : Logistic Regression using selected attributes

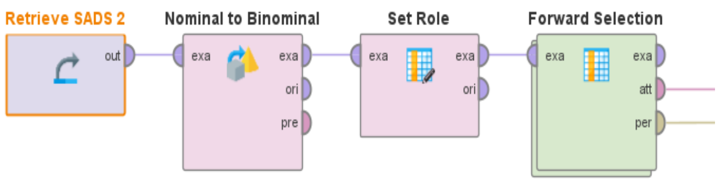

Figure 3: Logistic Regression using Forward Selection

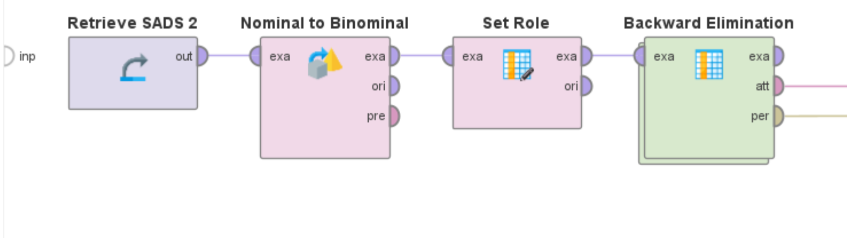

Figure 4: Logistic Regression using Backward Elimination

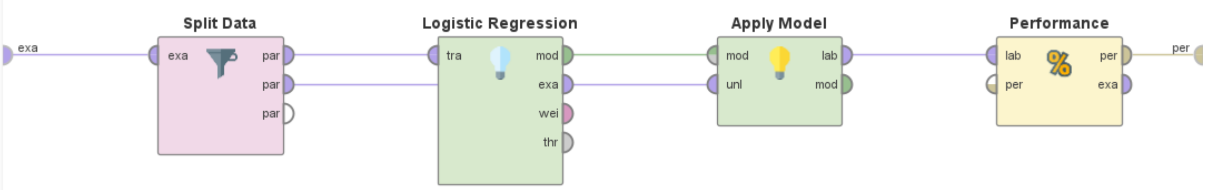

Figure 5: Inside Forward Selection and Backward Elimination of Logistic Regression Workflow

7.2 Decision Tree

A decision tree algorithm is a supervised machine learning algorithm used for both classification and regression problems. It works by constructing a tree-like model of decisions and their potential consequences, where each internal node is a feature in the dataset, each branch is a decision rule, and each leaf node is the outcome.

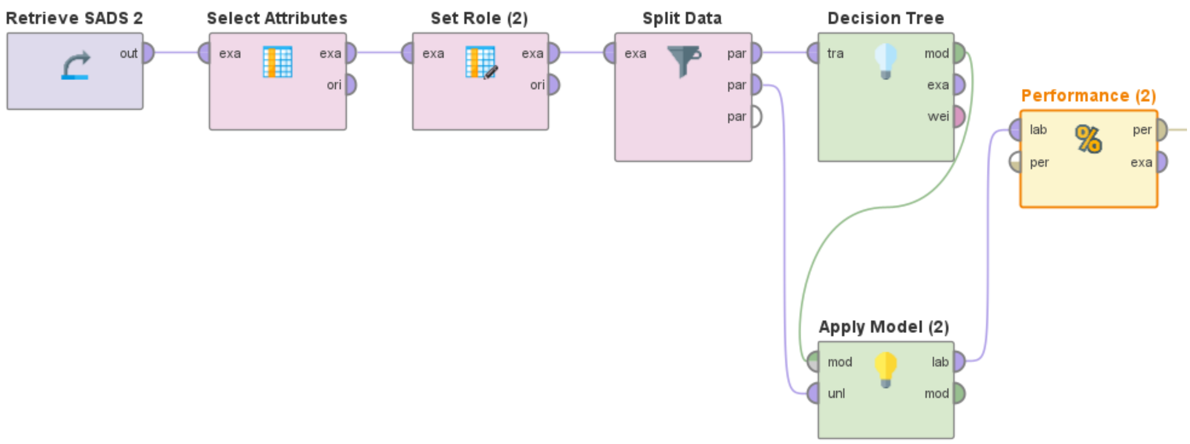

Figure 6: Decision Tree using selected attributes

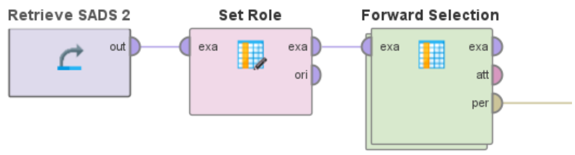

Figure 7: Decision Tree using Forward Selection

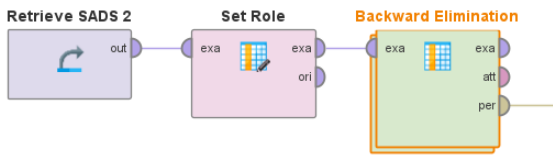

Figure 8: Decision Tree using Backward Elimination

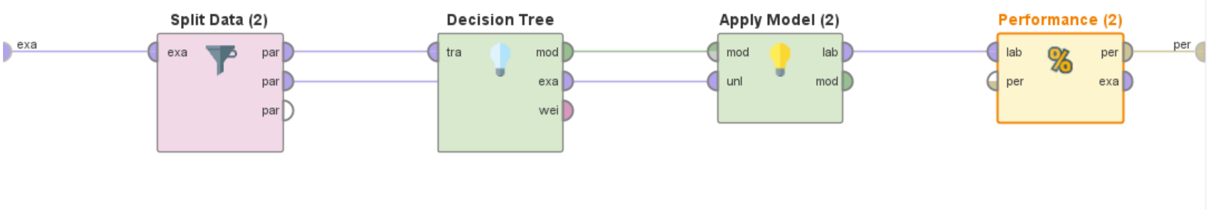

Figure 9: Inside Forward Selection and Backward Elimination of Decision Tree Workflow

7.3 Random Forest

Within the category of supervised learning techniques comes the well-known machine learning algorithm Random Forest. It is applicable to machine learning issues involving both classification and regression. The foundation of this approach lies in the notion of ensemble learning, which involves merging several classifiers to address intricate issues and enhance the model's functionality.

It is a classifier that contains a number of decision trees and takes the average to improve the predictive accuracy of that dataset.

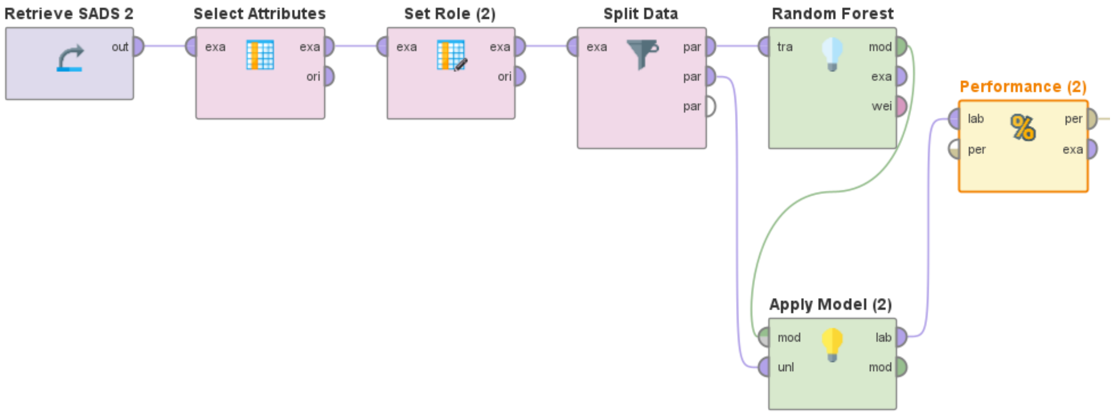

Figure 10: Random Forest using selected attributes

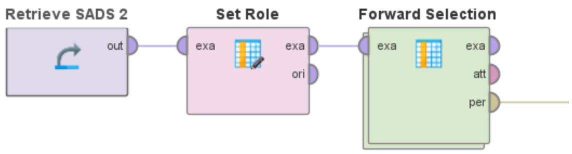

Figure 11: Random Forest using Forward Selection

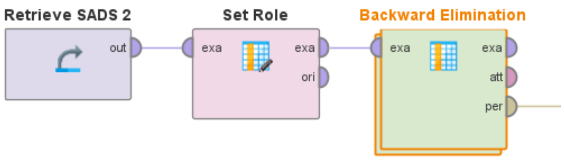

Figure 12: Random Forest using Backward Elimination

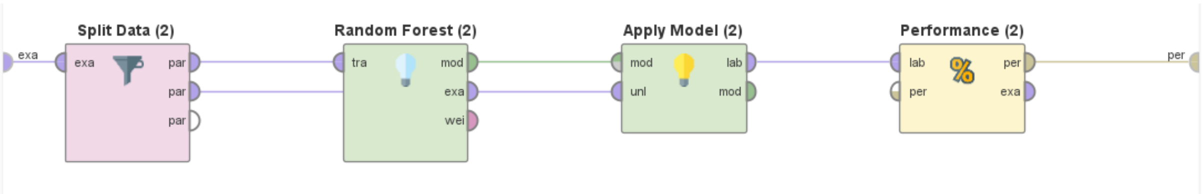

Figure 13: Inside Forward Selection and Backward Elimination of Random Forest Workflow

7.4 k- Nearest Neighbour

The basis of the k-Nearest Neighbour algorithm is the comparison of an unknown example with the k training Examples, which are the unknown example's closest neighbours. Finding the k nearest training datapoints is the first stage in applying the k-Nearest Neighbour algorithm to a new Example. The n Attributes in the training ExampleSet define "closeness" as a distance in the n-dimensional space.

The metric used to determine the separation between the training Examples and the unknown Example, is the Euclidean distance. The unknown example is classified in the second phase by the k-Nearest Neighbour algorithm based on the majority vote of the neighbours that were located.

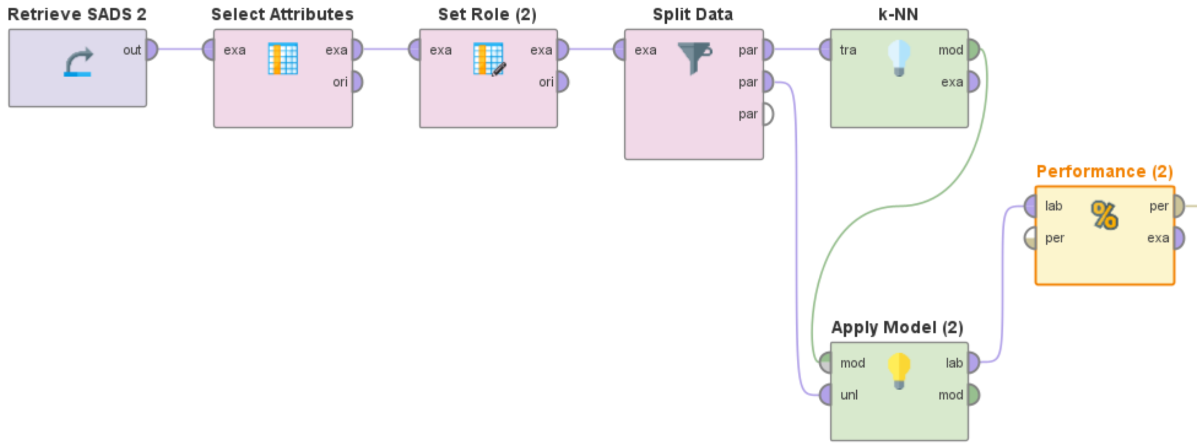

Figure 14: k-Nearest Neighbour using selected attributes

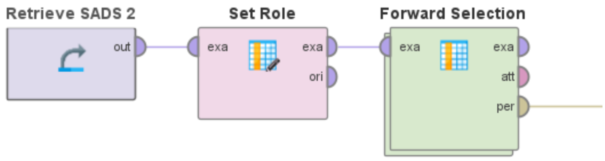

Figure 15: k-Nearest Neighbour using Forward Selection

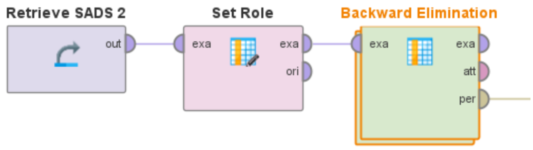

Figure 15: k-Nearest Neighbour using Backward Elimination

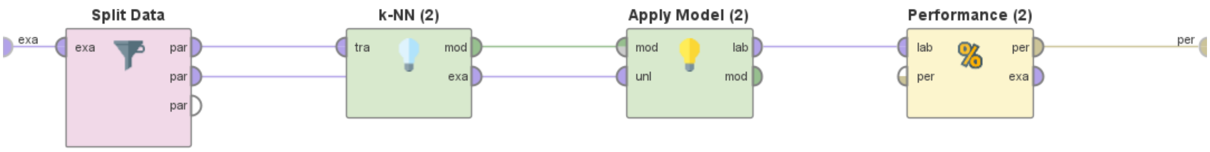

Figure 16: Inside Forward Selection and Backward Elimination of k-Nearest Neighbour Workflow

7.5 Support Vector Machine

SVM is a supervised machine learning approach that is used for regression as well as classification. The SVM algorithm's primary goal is to locate the best hyperplane in an N-dimensional space that may be used to divide data points into various feature space classes. The hyperplane attempts to maintain the largest possible buffer between the nearest points of various classes. The number of features determines the dimension of hyperplane.

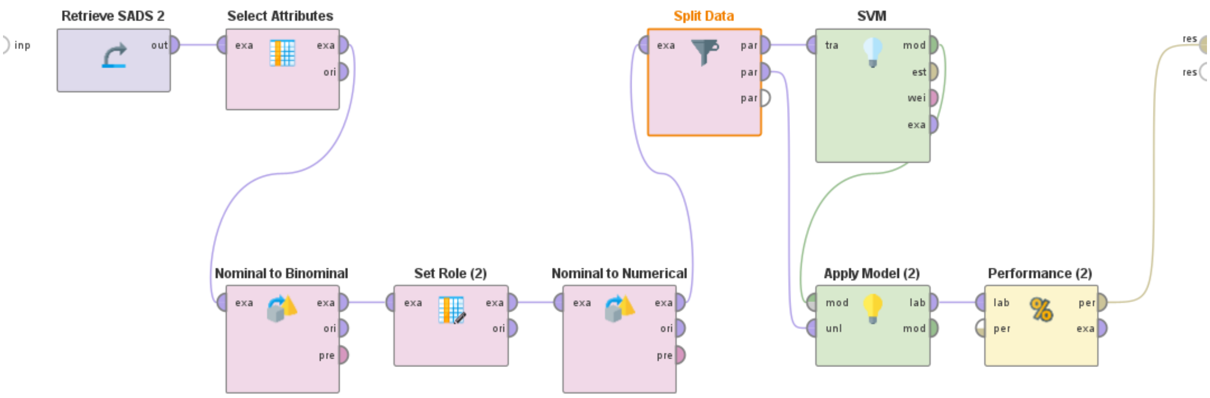

Figure 17: Support Vector Machine using selected attributes

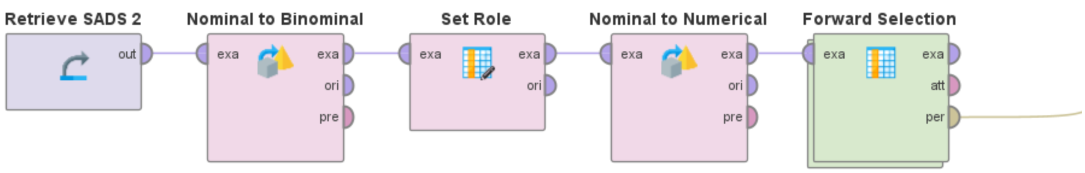

Figure 18: Support Vector Machine using Forward Selection

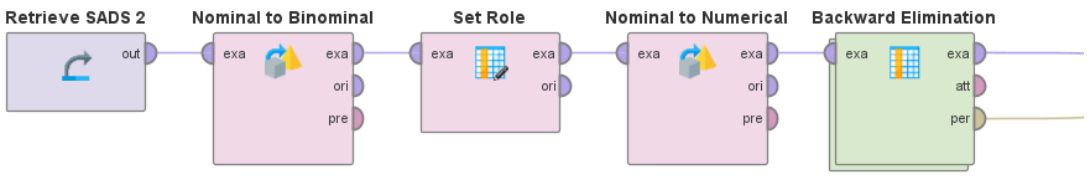

Figure 19: Support Vector Machine using Backward Elimination

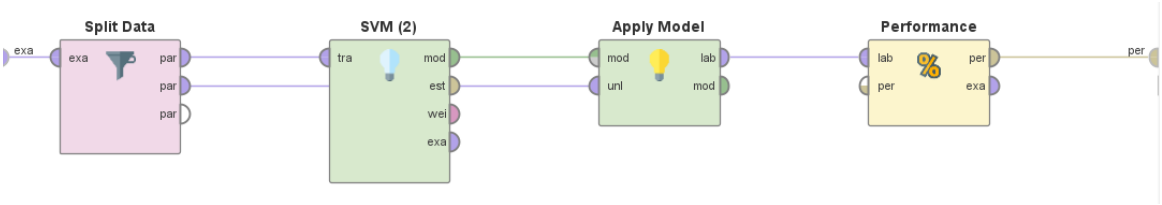

Figure 20: Inside Forward Selection and Backward Elimination of Support Vector Machine Workflow

8. Metrics

8.1 Classification accuracy

Classification accuracy (CA) gives the fraction of samples in the test set which have been correctly classified.

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \quad (6)$$

8.2 Precision

Precision gives the ratio between true positives that are correctly predicted out of total positives predicted.

$$\text{Precision} = \frac{\text{Number of true positive predictions}}{\text{Total number of positives predictions}} \quad (7)$$

8.3 Sensitivity (a.k.a. recall)

Sensitivity gives the ratio between true positives that are correctly predicted out of total actual positives.

$$\text{Sensitivity} = \frac{\text{Number of true positive predictions}}{\text{Total number of actual positives}} \quad (8)$$

8.4 F1 score

The F1 score is defined as the harmonic mean between Precision and Recall

$$\text{F1 Score} = \frac{2 \text{ X Precision X Recall}}{\text{Precision+Recall}} \quad (9)$$

9. Feature Selection

Feature selection is a technique for choosing the most important features from a dataset to build a machine learning model. This can improve the performance of the model by reducing overfitting and enhancing the model's interpretation.

9.1 Forward Selection

Forward selection begins with an empty model and adds variables one by one. In each forward step, it adds the one variable that enhances the individual improvement to your model.

9.2 Backward Elimination

This technique used in machine learning to remove features that do not have a substantial effect on the dependent variable or prediction of output. The process starts with all possible predictors and then uses lm() to compute the model. The summary() function is used to find each predictor's significance level. The predictor with the least significance has the largest p-value.

10. RESULTS

10.1 After applying feature selection techniques (forward selection and backward elimination)

Table 5: Accuracy of models using Forward Selection and Backward Elimination

		Feature Selection	
Model	**No Feature Selection(All features)**	**Forward Selection**	**Backward Elimination**
Logistic Regression	100%	97.32%	100%
Decision Tree	92.78%	95.54%	85.71%
Random Forest	87.50%	92.86%	82.14%
k-Nearest Neighbour	81.25%	93.75%	77.68%
Support Vector Machine	95.54%	100%	100%

In general, feature selection improves the accuracy of all machine learning models, except for the logistic regression model. Forward selection is more effective than backward elimination at improving accuracy, except for the SVM model. The k-nearest neighbor model is the most sensitive to feature selection, while the logistic regression model is the least sensitive. The improvement in accuracy from feature selection is relatively small for most models, except for the k-nearest neighbor model. This suggests that the features in the dataset are generally relevant to the target variable (sleep apnea).

The SVM model has the highest accuracy (95.54%) without feature selection. Logistic Regression has the highest accuracy (97.32%) with Forward feature selection. This suggests that the SVM model is a great choice for this dataset with all the features. Overall, the results in the table suggest that feature selection is a useful technique for improving the accurateness of machine learning models except for SVM, especially for complex models that are prone to overfitting. Forward selection is a particularly effective feature selection technique.

Table 6: Precision of models using Forward Selection and Backward Elimination

		Feature Selection	
Model	**No Feature Selection(All features)**	**Forward Selection**	**Backward Elimination**
Logistic Regression	100%	94.23%	100%
Decision Tree	91.75%	95.54%	83.23%
Random Forest	85.73%	91.28%	81.22%
k-Nearest Neighbour	80.14%	92.89%	74.70%
Support Vector Machine	91.73%	100%	100%

The Logistic Regression model has a better precision (94.23%) with forward feature selection, since 100% precision is not considered as the optimal result. This implies that the Logistic Regression model is sensitive to the presence of irrelevant features. The Decision Tree model has a good precision (91.75%) without feature selection, and its precision (95.54%) increases slightly when feature selection is used.

The Random Forest model has a lower precision than Decision Tree model, but its precision increases after forward feature selection. The k-Nearest Neighbour model has the lowest precision of the models tested. This suggests that the model is not as well-suited for detecting sleep apnea. The SVM model has a better precision (94.23%) without feature selection, since 100% precision after feature selection is not considered as the optimal result. This suggests that the Decision Tree model is a good all-around model for detection of sleep apnea in terms of precision.

Table 7: Recall of models using Forward Selection and Backward Elimination

		Feature Selection	
Model	**No Feature Selection(All features)**	**Forward Selection**	**Backward Elimination**
Logistic Regression	100%	98.315%	100%
Decision Tree	90.87%	94.54%	86.25%
Random Forest	83.48%	91.24%	73.84%
k-Nearest Neighbour	75.23%	92.68%	74.15%
Support Vector Machine	95.58%	100%	100%

Based on the results, Decision Tree model has decent recall and F1-score with feature selection, suggesting that it is better at identifying true positives and has a good overall performance. The Random Forest model has its recall and F1-score are lower than the Decision Tree and Logistic Regression models.

The k-Nearest Neighbour has the lowest recall and F1-score whereas Support Vector Machine has highest recall and F1-score without feature selection. The Logistic Regression model is also a good option, as it has a high recall and F1-score with forward feature selection.

Table 8: F1-Score of models using Forward Selection and Backward Elimination

		Feature Selection	
Model	**No Feature Selection(All features)**	**Forward Selection**	**Backward Elimination**
Logistic Regression	1	0.962292	1
Decision Tree	0.913079	0.950374	0.847131
Random Forest	0.8459	0.9126	0.773544
k-Nearest Neighbour	0.776074	0.927849	0.74424
Support Vector Machine	0.936154	1	1

With regard to the dataset, Logistic Regression, Decision tree, Random Forest and k-Nearest Neighbour are giving the best result with feature selection, while SVM is giving the best result without any feature selection technique. Model with 100% accuracy is not considered as optimal model.

10.2 After selecting features

Following features were selected based on feature selection techniques- Information gain, Chi Squared Statistics and Relief: Blood Pressure, BMI Category and Occupation

Based on the selected features, the models' performance have been evaluated using accuracy, precision, recall and F1-score.

Table 9: Accuracy of models using selected attributes

Model	**All Features**	**With Selected Features**
Logistic Regression	100%	100%
Decision Tree	92.78%	84.82%
Random Forest	87.50%	83.93%
k-Nearest Neighbour	81.25%	83.93%
Support Vector Machine	95.54%	95.80%

The table shows that the SVM model has the highest accuracy with and without feature selection. This suggests that with the selected features, the model is the most accurate at detecting sleep apnea. The Random Forest model has a lower accuracy than the Decision Tree and Logistic Regression models. The k-Nearest Neighbour model has the lowest accuracy of the models tested.

Table 10: Precision of models using selected attributes

Model	All Features	With Selected Features
Logistic Regression	100%	100%
Decision Tree	91.75%	80.39%
Random Forest	85.73%	80.17%
k-Nearest Neighbour	80.14%	79.51%
Support Vector Machine	91.73%	91.66%

Based on the precision values, the Support Vector Machine model has the highest precision. This means that they are more likely to accurately identify patients with sleep apnea.The precision of Decision Tree model decreases slightly when feature selection is used. The Random Forest and k-Nearest Neighbour models have the lowest precision values.

Table 11: Recall of models using selected attributes

Model	All Features	With Selected Features
Logistic Regression	100%	100%
Decision Tree	90.87%	79.14%
Random Forest	83.48%	77.69%
k-Nearest Neighbour	75.23%	77.69%
Support Vector Machine	95.58%	97.34%

Table 12: F1-Score of models using selected attributes

Model	All Features	With Selected Features
Logistic Regression	1	1
Decision Tree	0.913079	0.797601
Random Forest	0.8459	0.789105
k-Nearest Neighbour	0.776074	0.785895
Support Vector Machine	0.936154	0.944146

Based on the recall and F1-score values, the SVM model is the best overall model for detecting sleep apnea. It has the highest recall and F1-score, both with and without feature selection. This means that it has a good balance of precision and recall, making it a reliable and effective tool for detecting sleep apnea. The Decision Tree model is a good all-around model, but its F1-score decreases slightly when feature selection is used. The Random Forest and k-Nearest Neighbour models have the lowest recall and F1-score, suggesting that they are not as well-suited for detecting sleep apnea.

11. Conclusion

The purpose of the study was to develop the best machine learning algorithm to detect OSA and to find the features that have high relevance and assist in improving detection of OSA. The most relevant features based on the feature selection techniques are Blood pressure, BMI Category and occupation.

In terms of accuracy, precision, recall, and F1-score, SVM is providing the most accurate result, along with feature selection. Model has an accuracy of 95.54%, precision of 91.73%, recall of 95.58% and F1-score of 0.936154. Decision Tree is giving the best result along with forward selection and backward elimination.

The study's findings imply that feature selection may be a helpful method for raising machine learning models' accuracy in detecting sleep apnea. With and without feature selection, the best accurate model was the Support Vector Machine model. The Decision Tree models is also good choice, while the Logistic Regression, Random Forest and k-Nearest Neighbour models were not as effective.

12. Discussion

The results of this study are consistent with previous research on feature selection for machine learning models. Feature selection can help to improve the accuracy of models by removing irrelevant or redundant features. This can make the models more efficient and less prone to overfitting. In this study, we used different feature selection techniques: forward selection, backward elimination, information gain, chi squared statistics and relief. All the techniques were effective at improving the accuracy of the models. However, forward selection was the most effective technique for most of the models. The results of this study also suggest that the SVM model is a good choice for detecting sleep apnea. The model was the most accurate model, both with and without feature selection. Additionally, the model had a high precision and recall, both with and without feature selection.

Overall, the study's findings point to feature selection as a possible approach for increasing the precision with which machine learning models detect sleep apnea. The study's most accurate model was the Support Vector Machine model, thus making appropriate for supervised feature selection. For unsupervised feature selection, the Decision Tree model is also a useful general-purpose model for detecting sleep apnea.

While the research on ML for OSA diagnosis is promising, there are still some challenges that need to be addressed before ML-powered tools can be widely adopted in clinical practice. One challenge is that the accuracy of ML models can vary depending on the quality and quantity of data that they are trained on. Future research should focus on developing ML models that are trained on large and diverse datasets of OSA patients. This will help to improve the accuracy and robustness of the models. Additionally, researchers should work to develop ML models that are more interpretable and easier for clinicians to use.

13. Future Scope

According to the study's findings, machine learning may prove to be an effective method to detect sleep apnea. However, more research is needed to develop and validate more accurate and reliable models. One area of future research is to develop models that can use data from wearable devices to detect sleep apnea. Wearable devices are becoming increasingly popular, and they can collect a variety of data about people's sleep habits. This data could be used to develop models that are more accurate and convenient than models that require polysomnography.

Another area of future research is to develop models that can predict the severity of sleep apnea. This would be useful for identifying people who are at risk of serious health problems from sleep apnea and for guiding treatment decisions. Finally, more research is needed to understand how feature selection may be used to improve the accuracy of machine learning models for detecting sleep apnea. This could lead to the development of models that are more accurate and reliable than current models.

An ML-powered software platform can be developed to help doctors detect and treat OSA. Partnership with healthcare providers can offer ML-powered OSA screening and treatment planning services. This could be done on a fee-per-service basis or through a subscription model. More research is needed to develop and validate ML algorithms for OSA diagnosis using data from wearable devices and electronic health records, and polysomnography data. Additionally, more research is needed to develop ML-powered tools that can be used to personalize OSA treatment.

14. References

[1] T. M. Suri, T. Ghosh, S. Mittal, V. Hadda, K. Madan and A. Mohan, "Systematic review and meta-analysis of the prevalence of obstructive sleep apnea in Indian adults," *Sleep Medicine Reviews,* 2023.

[2] H. . L. Brennan and . S. . D. Kirby, "The role of artificial intelligence in the treatment of obstructive sleep apnea," *Journal of Otolaryngology - Head & Neck Surgery,* 2023.

[3] H. ALRUWAILI, . A. AHMED, . A. FATANI, . K. AL-OTAIBI, . S. AL-JAHDALI, Y. ALI, . A. AL-HARBI, S. BAHAROON, . M. KHAN and H. AL-JAHDALI, "Symptoms and risk for obstructive sleep apnea among sample of Saudi Arabian adults," *Sleep and Biological Rhythms,* p. 332–341, 2015.

[4] V. . K. Kapur, D. . H. Auckley, S. Chowdhuri, . D. C. Kuhlmann, R. Mehra, K. Ramar and . C. . G. Harrod, "Clinical Practice Guideline for Diagnostic Testing for Adult Obstructive Sleep Apnea: An American Academy of Sleep Medicine Clinical Practice Guideline," *Journal of Clinical Sleep Medicine,* p. 479–504, 2017.

[5] C. Mencar, C. Gallo, M. Mantero, P. Tarsia, G. . E. Carpagnano, M. . P. F. Barbaro and D. Lacedonia, "Application of machine learning to predict obstructive sleep apnea syndrome severity," *Heakth Informatics Journal,* 2019.

[6] I. Aiyer, L. Shaik, A. Sheta and S. Surani, "Review of Application of Machine Learning as a Screening Tool for Diagnosis of Obstructive Sleep Apnea," *Medicina,* 2022.

[7] A. Sheta, T. Thaher, S. R. Surani, H. Turabieh , M. Braik, J. Too, N. Abu-El-Rub, M. Mafarjah, H. Chantar and S. Subramanian , "Diagnosis of Obstructive Sleep Apnea Using Feature Selection, Classification Methods, and Data Grouping Based Age, Sex, and Race," *Diagnostics,* 2023.

[8] C. GUILLEMINAULT, S. CONNOLLY, R. WINKLE, K. MELVIN and A. TILKIAN, "CYCLICAL VARIATION OF THE HEART RATE IN SLEEP APNOEA SYNDROME: Mechanisms, and Usefulness of 24 h Electrocardiography as a Screening Technique," *Lancet,* pp. 126-131, 1984.

[9] T. Penzel, G. Amend, K. Meinzer, J. H. Peter and P. von Wichert, "MESAM: A Heart Rate and Snoring Recorder for Detection of Obstructive Sleep Apnea," *Sleep,* pp. 175-182, 1990.

[10] S. Wang, W. Xuan, D. Chen, Y. Gu, . F. Liu, J. Chen, S. Xia, S. Dong and . J. Luo, "Machine Learning Assisted Wearable Wireless Device for Sleep Apnea Syndrome Diagnosis," *Biosensors,* 2023.

[11] Y. Varol, Z. . Z. Uçar, B. . O. Arslan and I. Karasu , "Apnea-hypopnea index and the polysomnographic risk factors for predicting 5- to 8-year mortality in patients with OSA," *Sleep and Breathing,* 2023.

[12] D. Scrutinio, P. Guida, M. Aliani, G. Castellana, P. Guido and M. Carone, "Age and comorbidities are crucial predictors of mortality in severe obstructive sleep apnoea syndrome," *European Journal of Internal Medicine,* pp. 71-76, 2021.

[13] N. S. Marshall, K. K. H. Wong, P. Y. Liu, S. R. J. Cullen, M. W. Knuiman and R. R. Grunstein, "Sleep Apnea as an Independent Risk Factor for All-Cause Mortality: The Busselton Health Study," *Sleep,* p. 1079–1085, 2008.

[14] D. . M. Hiestand, . P. Britz , M. Goldman and B. Phillips, "Prevalence of Symptoms and Risk of Sleep Apnea in the US Population," *Chest,* pp. 780-786, 2006.

[15] S. . C. Veasey and . I. . M. Rosen, "Obstructive sleep apnea in adults," *New England Journal of Medicine,* p. 1442–1449, 2019.

[16] O. Faust, . R. Barika, A. Shenfield, E. . J. Ciaccio and U. R. Acharya, "Accurate detection of sleep apnea with long short-term memory network based on RR interval signals," *Knowledge-Based Systems,* 2021.

[17] V. Kalaivani, "Automated recognition of obstructive sleep apnea using ensemble support vector machine classifier," *International Journal of Biomedical Engineering and Technology,* pp. 274-289, 2020.

[18] B. B. Rekha and A. Kandaswamy , "Ensemble classification approach for screening of obstructive sleep apnoea using ECG," *International Journal of Biomedical Engineering and Technology,* pp. 139-150, 2018.

[19] A. Maniaci, P. . M. Riela, G. Iannella, J. R. Lechien, I. . L. Mantia, M. . D. Vincentiis, G. Cammaroto, C. Calvo-Henriquez, M. D. Luca, C. C. Estomba, A. M. Saibene, I. Pollicina, G. Stilo, P. . D. Mauro, A. Cannavicci, R. Lugo, . G. Magliulo, A. Greco, A. Pace, G. Meccariello, . S. Cocuzza and C. Vicini, "Machine Learning Identification

of Obstructive Sleep Apnea Severity through the Patient Clinical Features: A Retrospective Study," *Life*, 2023.

[20] L. Almazaydeh, K. Elleithy and M. Faezipour, "Obstructive Sleep Apnea Detection Using SVM-Based Classification of ECG Signal Features," *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* 2012.

[21] O. Yildirim, U. . B. Baloglu and . U. . R. Acharya , "A Deep Learning Model for Automated Sleep Stages Classification Using PSG Signals," *International Journal of Environmental Research and Public Health*, 2019.

[22] L. Zhang, Y. . R. Yan, S. Q. Li, H. . P. Li, Y. N. Lin, N. Li, X. W. Sun, Y. . J. Ding, C. . X. Li and Q. Y. Li, "Moderate to severe OSA screening based on support vector machine of the Chinese population faciocervical measurements dataset: a cross sectional study," *BMJ Open,* 2021.

[23] M. O. Mendez, M. A. Bianchi, . M. Matteucci, S. Cerutti and . T. Penzel, "Sleep Apnea Screening by Autoregressive Models From a Single ECG Lead," *IEEE Transactions on Biomedical Engineering,* pp. 2838-50, 2009.

[24] G. Elhayatmy, N. Dey and A. . S. Ashour, "Internet of Things Based Wireless Body Area Network in Healtcare," *Internet of Things and Big Data Analytics Toward Next-Generation Intelligence,* pp. 3-20, 2018.

[25] M. Chowdhury, G. K. Dey and M. R. Karim, "IEHSAM: IoT based E-Health and Sleep Apnea Monitoring System," *2021 Emerging Trends in Industry 4.0 (ETI 4.0),* pp. 1-7, 2021.

[26] D. Sharma, . T. K. Singhal and . D. Singh, "Threat Intelligence Model to Secure IoT based Body Area Network and Prosthetic Sensors," *ECS Transactions,* 2022.