International Journal of Advances in Applied Sciences (IJAAS)
Vol. 15, No. 1, March 2026, pp. 29~41
ISSN: 2252-8814, DOI: 10.11591/ijaas.v15.i1.pp29-41

Enhancing sleep disorder diagnosis through ensemble ML models: a comprehensive study on insomnia and sleep apnea

Satyaprakash Swain[1,4], Binod Kumar Pattanayak[1], Mihir Narayan Mohanty[2], Amiya Kumar Sahoo[3], Suvendra Kumar Jayasingh[4]

[1]Department of Computer Science and Engineering, Institute of Technical Education and Research, SoA University, Bhubaneswar, India
[2]Department of Electronics and Communication Engineering, Institute of Technical Education and Research, SoA University, Bhubaneswar, India
[3]Department of Computer Science and Engineering, Aryan Institute of Engineering and Technology, Bhubaneswar, India
[4]Department of Computer Science and Engineering, Institute of Management and Information Technology, Cuttack, India

Article Info

Article history:

Received Nov 18, 2024
Revised Jun 17, 2025
Accepted Jul 14, 2025

Keywords:

Boosting model
Decision tree
Gradient boost
Naive Bayes
Principal component analysis
Random forest
Support vector machines

ABSTRACT

Sleep disorders are common and can significantly harm human health, with insomnia and sleep apnea being the most prevalent conditions. These disorders are often difficult to detect and treat accurately. Although machine learning (ML) techniques have shown promise in improving diagnostic precision and personalized treatment, most existing studies rely on single-source data or conventional ML models, which limit their robustness and generalizability across diverse populations. To address this research gap, this study integrates multi-modal data and ensemble learning techniques to enhance accuracy, interpretability, and real-time applicability in diagnosing insomnia and sleep apnea. A dataset of 400 samples was collected through manual methods and internet of things (IoT) devices from multiple sources. Statistical techniques were applied for data cleaning, followed by principal component analysis (PCA) to reduce dimensionality and improve training efficiency. Four base ML models: decision tree (DT), support vector machine (SVM), naive Bayes (NB), and random forest (RF) were initially trained and evaluated. Subsequently, a boosting-based ensemble model was implemented to further improve performance. The proposed gradient boosting model with RF as the base learner achieved the highest diagnostic accuracy of 96.01%. The results demonstrate that ensemble ML models combined with multi-modal data significantly enhance the accuracy of insomnia and sleep apnea diagnosis.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Suvendra Kumar Jayasingh
Department of Computer Science and Engineering, Institute of Management and Information Technology
Gourav Palace, Matha sahi, post-Tulsipur Cuttack-753008, Odisha, India
Email: sjayasingh@gmail.com

1. INTRODUCTION

Millions of individuals worldwide are impacted by the prevalent and serious problem of sleep problems. Among the many sleep-related issues, insomnia and sleep apnea are particularly common and provide substantial difficulties for proper diagnosis and efficient treatment. These sleep problems can have serious effects on one's physical health, mental health, to find general quality of life if they are not treated effectively. Fortunately, the development of machine learning (ML) presents a promising way to improve the accuracy of diagnostic procedures and, more importantly, to customize treatment plans to meet the specific needs of each person. This study provides a novel method that uses ML models to tackle the challenging problems of detecting insomnia and sleep apnea. With the use of common ML models like decision tree

(DT), support vector machines (SVM), naive Bayes (NB), and random forest (RF), we have proposed a comprehensive hybrid ML model after adding boosting methods that includes information from these four different ML base methods. These different models each have impressive diagnostic accuracies when used alone, ranging from 87.96 to 91.17%.

This study entered the field of boosting models like adaptive boosting (AdaBoost), gradient boosting, extreme gradient boosting (XGBoost), categorical boosting (CatBoost), and light gradient boosting machine (LightGBM), which are included in our comparison. With an astounding accuracy rate of 96%, the gradient boosting model stands out as the clear winner among them. This outstanding achievement highlights the enormous potential of ML to greatly improve the accuracy of diagnoses for sleep disorders. A carefully selected dataset of 400 data points, gathered through a combination of manual gathering and data supplied by internet of things (IoT) devices, serves as the basis for this study. This dataset contains a wide range of people from different occupations, such as farmers, solicitors, professors, and doctors. It includes a wide range of information, such as demographics, specific sleep patterns, thorough medical histories, and other important characteristics. The results of this study highlight the profound influence that ML models can have on the field of diagnosing sleep disorders. In addition to providing a significant improvement in diagnostic accuracy, the boosting model, particularly when paired with gradient boosting, has the potential to lay the groundwork for the creation of an advanced clinical decision support system.

By utilizing ensemble ML models designed to identify insomnia and sleep apnea, this study presents a novel method for diagnosing sleep disorders. The majority of current research on diagnosing sleep disorders uses single-source data or conventional ML models, which frequently lack generalisability and robustness across a range of demographics. To improve accuracy, interpretability, and real-time applicability in the diagnosis of insomnia and sleep apnea, a substantial research gap still exists in the integration of multi-modal data and ensemble learning techniques. In contrast to earlier research that mostly concentrated on individual disorders or used conventional diagnostic techniques, this work incorporates sophisticated ensemble techniques to increase the precision and dependability of sleep disorder diagnoses. The robustness of the detection process is improved by this work by combining different ML algorithms, which addresses the shortcomings of previous models that frequently ignore the complexity and overlap of symptoms of various sleep disorders. By providing a comprehensive, automated solution that not only expedites diagnosis but also lowers the possibility of human error, this study marks a significant advancement in the field and ultimately improves patient outcomes and streamlines clinical workflows. A system like this might provide medical personnel with individualized diagnostic insights, permitting customized therapy modalities and eventually enhancing outcomes for people suffering from sleep problems.

This study has reviewed a number of works related to healthcare applications using ML or deep learning (DL) models. They have taken a different parameterized dataset for their prediction work. But here we have taken a dataset that contains 13 parameters, which are most relevant to sleeping disorders. Also, this work has implemented the new ensemble boosting models to increase the accuracy levels for good predictions of disease.

2. LITERATURE REVIEW

Carmes *et al*. [1] introduced a semi-automatic approach for sleep apnea diagnosis utilizing three-channel polysomnography (PSG) data. This method employs an ElasticNet classifier to achieve a remarkable 97.9% accuracy in sleep apnea classification. Notably, it minimizes the time-consuming manual analysis of PSG recordings, effectively mitigating inter-scorer variability. With potential as a viable sleep apnea screening tool in clinical settings, this method enhances efficiency and reliability, showcasing its value in advancing diagnostic accuracy and streamlining the assessment process for sleep specialists.

Choyon *et al*. [2] propose an inventive solution by integrating IoT and ML to bolster health monitoring and combat the pandemic more efficiently. It highlights the shortcomings of existing monitoring systems and introduces an IoT-based method for real-time health data collection, facilitating prompt detection of COVID-19 severity. The envisioned system seeks to deliver holistic healthcare, remote communication, and emergency support, offering a pragmatic approach to alleviate the repercussions of the pandemic.

Korkalainen *et al*. [3] proposed a model by utilizing DL and diverse epoch durations to uncover neglected sleep-wake transitions. The examination of 446 obstructive sleep apnea (OSA) patients indicates that traditional approaches may underestimate sleep fragmentation in severe OSA. Shorter epoch-to-epoch durations reveal heightened wakefulness and reduced rapid eye movement/non-REM 3 (REM/N3), emphasizing the necessity for a thorough analysis of sleep architecture when evaluating sleep disorders.

Shahin *et al*. [4] present a dual-phase automated method to identify insomnia from overnight electroencephalogram (EEG) recordings, addressing the deficiency of prompt insomnia detection tools.

The approach utilizes deep neural network models, training both sleep stage and epoch-level insomnia detection models. Subject-level features extracted from these models facilitate the ultimate binary classification of subjects into control or insomniac categories. Assessment on 115 participants demonstrates encouraging outcomes, including an F1-score of 0.88, 84% sensitivity, and 91% specificity, underscoring its potential clinical applicability.

Kuo and Chen [5] introduce a novel short-time insomnia detection system utilizing refined composite multi-scale entropy (RCMSE) analysis on single-channel sleep electrooculography (EOG). With an SVM classifier, the proposed system demonstrated high accuracy (89.31%), sensitivity (96.63%), and F1-score (90.04%) when tested on 32 subjects, half healthy and half with insomnia. RCMSE emerges as a valuable feature for short-duration insomnia detection, and the single-channel sleep EOG enhances homecare applicability for potential integration into portable PSG systems.

Islam *et al*. [6] delve into the influence of technological advancements on human health, specifically the surge in insomnia linked to virtual engagements and sedentary habits. Acknowledging the drawbacks of costly and time-intensive medical tests, the authors suggest an intelligent ML model to predict chronic insomnia. Utilizing seven classifiers, the logistic regression model demonstrates exceptional accuracy at 98%, presenting a compelling and effective method for detecting insomnia in individuals, particularly in settings with limited resources.

Zulfiker *et al*. [7] tackle the widespread health issue of insomnia, a prevalent sleep disorder associated with mental health challenges and substance abuse. Introducing an ML approach, the multilayer stacking model predicts insomnia based on socio-demographic factors, achieving a notable 88.60% accuracy. Leveraging principal component analysis (PCA) for feature reduction, the proposed ensemble model surpasses other leading classifiers like AdaBoost and gradient boost, showcasing superior performance across multiple efficacy metrics. This emphasizes its potential as an effective tool for predicting insomnia.

Alghwiri *et al*. [8] in their cross-sectional study with 1,600 university students, the investigation focuses on sleep disturbances, employing logistic regression, and advanced ML techniques. The study discloses a 70% prevalence of poor sleep quality, identifying the RF model as the most accurate predictor with 74% accuracy and 95% specificity. Factors associated with better sleep quality include age and tea consumption, while risk factors consist of electronics usage, headaches, systemic diseases, and neck pain. These results provide valuable insights for enhancing student well-being and crafting tailored interventions.

Jayasing *et al.* [9] proposed a model for weather forecasting, highlighting that its nonlinear nature is impacted by climate change and the butterfly effect. The research introduces hybrid soft computing models, merging SVM, multi-layer perception, and fuzzy logic, aimed at improving the accuracy of weather predictions for Delhi. Assessment metrics such as mean absolute error (MAE), root mean squared error (RMSE), relative absolute error (RAE), and root relative squared error (RRSE) substantiate the effectiveness of these models in enhancing forecasting precision.

Swain *et al*. [10] in their study compare different ML models for tackling the complex task of illness prognosis, specifically focusing on individuals aged 20 and above, particularly those with HbA1c levels surpassing 6.5%, indicative of diabetic diseases. Employing IoT, the research assesses risk factors akin to heart diseases, highlighting the significance of ML in diagnosis and prevention. The incorporation of advanced technologies such as IoT, cloud applications, and artificial neural networks holds potential for improving healthcare strategies in addressing heart diseases.

Mondal *et al*. [11] focus on the increasing risk of coronary heart disease (CHD) in adults, utilizing ML for timely detection. With a dataset of 4,240 instances and 15 features, diverse ML models were examined, highlighting RF and gradient boost classifiers for their accuracy. The ultimate ensemble model, combining boosting models, demonstrated an outstanding 92.26% accuracy, highlighting its efficacy in early CHD risk prediction with minimal false negatives.

Kumar and Reddy [12] delve into different cardiac diseases, highlighting heart failure (HF), coronary artery disease (CAD), and cardiovascular disease (CVD). The proposed diagnostic system utilizes supervised learning, specifically the gradient boosting technique, to accurately detect HF, utilizing the Cleveland dataset. Achieving an outstanding 97.10% accuracy, the model proves effective in automated HF diagnosis, outperforming alternative methods. The emphasis on gradient boosting signifies a significant advancement in heart disease detection techniques.

Das *et al*. [13] addressing the urgent need for early heart disease detection, this study utilizes ML with five boosting methods: AdaBoost, gradient boosting, XGBoost, CatBoost, and LightGBM on the University of California, Irvine (UCI) Cleveland HD dataset. Comparative analysis demonstrates that AdaBoost and XGBoost excel with the highest accuracy (≈93%), precision (0.94), recall (0.93), and F1-scores (0.93). These results surpass recent ensemble techniques and stacking classifiers applied to the same dataset.

Sen and Verma [14] advocates for advanced diagnostic methods using ML. The proposal introduces a soft voting meta-classifier composed of CatBoost, LightGBM, Gaussian naive Bayes (GNB), RF, and

XGBoost, surpassing traditional approaches. Conducted on a combined UCI heart disease and Statlog dataset, the experiment yields remarkable results, achieving a 91.85% accuracy and a 0.9344 area under the curve score (AUC). These outcomes signify enhanced diagnostic efficacy in predicting heart disease, showcasing the potential of the proposed model.

Afreen *et al*. [15] utilize ML, employing CatBoost and LightGBM models to predict and categorize liver disease, addressing the increasing prevalence of patients with diverse complications. Implementing a gradient boosting-based classifier with feature selection through preprocessing enhances model accuracy. CatBoost achieves the highest accuracy at 86.8%, while LightGBM attains 82.6%, demonstrating their efficacy in early identification of liver diseases for improved patient outcomes.

3. RESEARCH DESIGN

3.1. Machine learning

A crucial area of artificial intelligence (AI) is ML [16], which enables computers to make judgments and learn from data on their own. It is a fast-developing discipline with a wide range of applications, all of which are basically based on data analysis to identify trends and linkages. ML encompasses various approaches, including:

- Supervised learning, where algorithms use labeled data for learning.
- Unsupervised learning for uncovering hidden patterns in unlabeled data.
- Reinforcement learning, applied to interactive settings.

ML is used in speech and image recognition, recommendation systems, autonomous vehicles, healthcare, and robotics, advancing computer vision, language understanding, and robotics. Data quality, model interpretability, and ethical issues are challenges. Nevertheless, ML continues to reshape industries, reaffirming its pivotal position in contemporary technology and AI, and having a significant impact on how we live our lives.

3.2. Machine learning architecture

Data collection and preprocessing are the first steps in the ML architecture process, as shown in Figure 1. Next, depending on the task at hand, a suitable ML model is chosen. For the purposes of model training and evaluation, the data is divided into training and validation sets. When hyperparameter tuning is successful in optimizing model performance, the model is used to make predictions about the real world. To automate decision-making and pattern identification across a variety of applications, ML is a powerful tool. Constant monitoring and interpretability ensure the model remains effective and understandable.

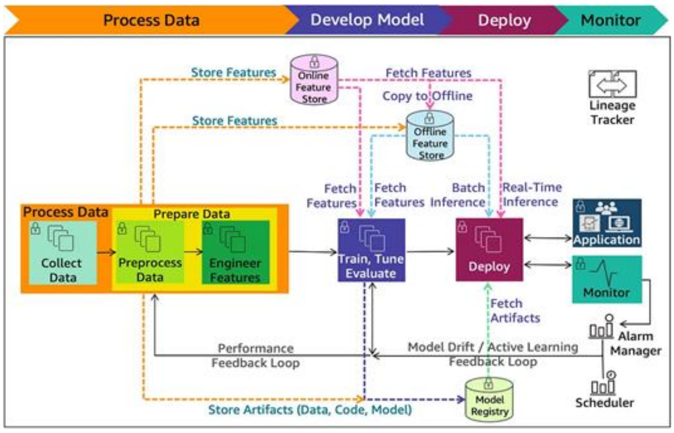

Figure 1. Architecture of ML

3.3. Ensemble learning

A ML technique termed ensemble learning [17] combines the predictions of various separate models (commonly referred to as "base models" or "weak learners") in order to enhance overall predictive performance. The theory behind ensemble learning is that by combining the advantages of various models, it can lessen the drawbacks of each model separately and generate predictions that are more reliable and accurate. There are several ensemble learning models available, such as bagging, boosting, voting, and stacking. This work has implemented a voting and boosting model with the best estimator, which are describe as follows.

4. METHODOLOGY
4.1. Flow of work
Figure 2 shows the proposed work model to achieve the target. The raw dataset is preprocessed by using different statistical methods like mean, mode, and median to reduce empty, null, or irrelevant data items in the dataset. Then PCA techniques is used to reduce the dimension of the dataset. 70% data in the dataset is used for machine training, 20% is for testing, and 10% is used for validation work. Four base-level ML voting models were used for training purpose and five type boosting models were used to enhance the accuracy results of the base voting models.

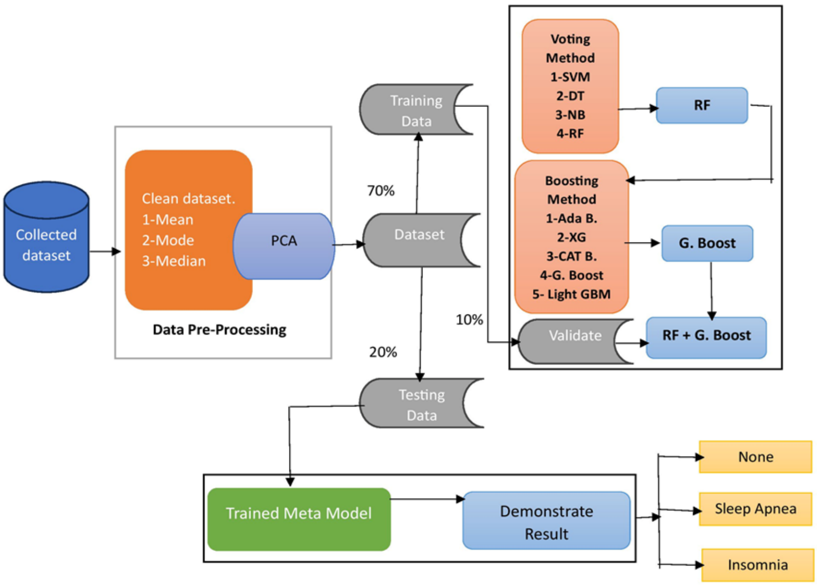

Figure 2. Proposed model of work

4.2. Dataset
4.2.1. Data collection
A real time data is gathered from 400 patients to create a complete dataset. This work dataset contains a wide range of patient-related data that was gathered using both modern IoT devices and manual recording techniques. The dataset includes crucial demographic information like gender and age, as well as important health-related factors like sleep duration and quality, physical activity levels, stress levels, body mass index (BMI) category, blood pressure, and heart rate. The incorporation of IoT devices improved the accuracy and breadth of our research findings by enabling us to acquire precise and real-time measurements. With the use of this comprehensive dataset, which serves as the basis for our research, we were able to examine and evaluate numerous facets of patient health and well-being and offer insightful contributions to the fields of medicine and healthcare.

4.2.2. Dataset cleaning
There is a chance of irrelevant data in the raw dataset, which may cause wrong machine training. Different statistical methods like mean, mode, and median are used to clean the dataset. These techniques help to reduce values like empty cells, garbage values, and non-related values.

4.2.3. Principal component analysis
A dimensionality reduction technique used in both data analysis and ML is PCA [18]. Its main goal is to preserve the most important information while reducing the complexity present in high-dimensional datasets. Some common significances of PCA are applied over our collected dataset, such as: dimensionality reduction, visualization, noise reduction, feature engineering, multi-collinearity mitigation, data compression, and anomaly detection.

A useful method with many applications in numerous study domains is PCA. It is a crucial tool for deriving valuable insights from complicated datasets because of its capacity to reduce dimensionality, increase visualization, and improve data quality. It enables them to detect hidden patterns, filter out background noise, and optimize subsequent steps. Figure 3 displays a graphic depiction of our obtained dataset after PCA was applied to it.

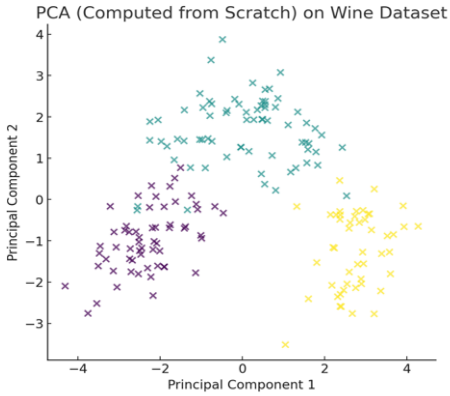

Figure 3. Graphical representation of PCA over collected dataset

5. SIMULATION AND RESULT ANALYSIS
5.1. Base model

Base model (voting) classifiers [19]–[23] are ML models that anticipate an output based on the class that has the highest possibility of being the output, as shown in Figure 4. These are gathered expertise by training on a collection of various models. It simply averages the outcomes of each classifier that was submitted into the voting classifier in order to forecast the output class based on the highest majority of votes. Instead of creating individual specialized models and evaluating their correctness, the idea is to develop a single model that learns from multiple models and predicts output based on the aggregate majority of voting for each output class [3], [24], [25]. The final prediction "y" is determined by counting the votes from each base model as in (1).

$$y = argmax\,(\sum_{i=1}^{N} 1\;[hi(x) - C]) \tag{1}$$

Let *N* is number of base models, *hi(x)* be the prediction made by the ith base model for input *x*, and *y* be the final ensemble prediction.

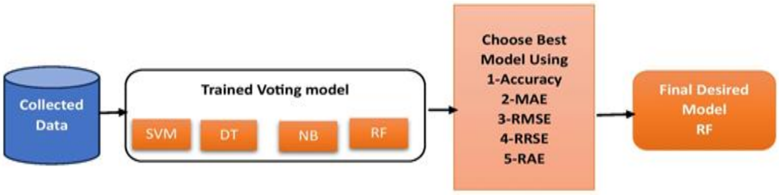

Figure 4. Voting model made using 4 different ml models

Here, four ML models like DT, SVM, NB, and RF are selected for ensemble learning voting model creation:

i) DT: a tree-like structure called a decision tree is employed for classification and regression applications. In order to decide or predict something, it divides the data into branches depending on feature values. Every node and branch in the tree indicates different features and options, respectively. DTs are renowned for being easy to understand and comprehend.

ii) SVM: a potent classifier that locates a hyperplane in high-dimensional space to divide various classes. It is efficient for both linear and non-linear classification jobs since it seeks to maximise the margin between classes. SVM is renowned for its dependability and adaptability.

iii) NB: based on Bayes' theorem, NB is a probabilistic classification algorithm. Calculations are made simpler by the assumption that characteristics are independent. NB is well-renowned for being quick and effective and is frequently used in text classification and spam detection.

iv) RF: a group of DTs is called a random. To lessen overfitting and increase prediction accuracy, it mixes the predictions of various trees. Since RF is so resilient and adaptable, it can be used for a variety of classification and regression applications. It is renowned for handling intricate data relationships well.

Here, the main dataset, which contains 400 samples, is divided into 4 equal numbers of instances, and apply voting model is applied on each instance. Based on the taken dataset and its Python simulation, it is clearly seen that RF gives the best accuracy among the other three models, which is shown in Table 1 and its graphical representation in Figure 5. Then, the comparisons between four models are provided by taking the parameters like MAE, RMSE, RAE, and RRSE for better understanding.

Table 1. Accuracy table of voting model of DT, SVM, NB, and RF over 4 equally divided datasets

Dataset instance	SVM	DT	NB	RF
1-100	90.2	92.39	88.04	89.13
100-200	88.09	86.9	85.7	90.47
200-300	88	85	85	92
300-400	91.83	91.83	90.81	92.85

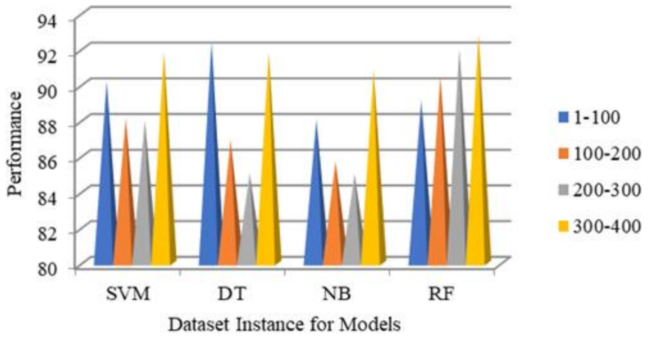

Figure 5. Graphical representation of accuracy of DT, SVM, NB, and RF

5.1.1. Mean absolute error

The average size of the errors between the expected and actual values is measured statistically as MAE. The performance of ML models is evaluated using the MAE, a popular measure of forecast error in time series research. It is a particularly helpful indicator for assessing continuous value prediction models. The MAE value presented in Table 2 and the graphic representation in Figure 6 when evaluating the performance of our model. The formula for MAE is (2).

$$MAE = (1/n)\,\Sigma(i = 1\ to\ n)\,|y_i - \hat{y}_i| \quad (2)$$

Where n is sample size, y_i is actual value at observation i, and $\hat{y}_i$ is the predicted value at observation i.

Table 2. MAE value of voting model of DT, SVM, NB, and RF

Dataset instance	SVM	DT	NB	RF
1-100	0.06	0.1	0.08	0.09
100-200	0.07	0.14	0.09	0.09
200-300	0.08	0.19	0.09	0.07
300-400	0.05	0.15	0.06	0.06

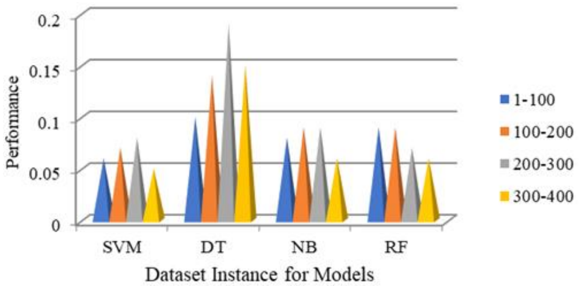

Figure 6. Graphical representation of MAE data over DT, SVM, NB, and RF

5.1.2. Root mean squared error

The statistical measurement of the discrepancy between expected and actual values is called RMSE. It is calculated by using the MSE square root as in (3). In model performance evaluation, the RMSE value is shown in Table 3, and a graphical representation is shown in Figure 7.

$$RMSE = sqrt(MSE) \tag{3}$$

Table 3. RMSE value of voting model of DT, SVM, NB, and RF

Dataset instance	SVM	DT	NB	RF
1-100	0.25	0.2	0.27	0.23
100-200	0.28	0.2	0.29	0.25
200-300	0.28	0.29	0.3	0.21
300-400	0.23	0.24	0.24	0.2

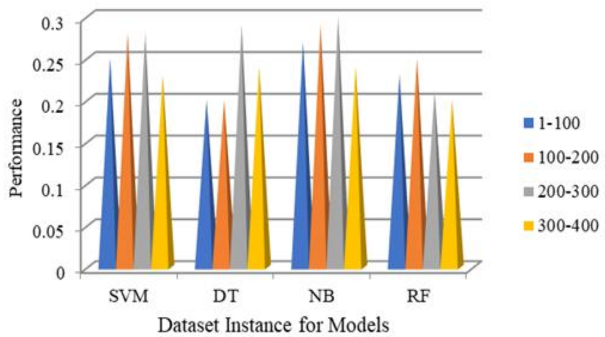

Figure 7. Graphical representation of RMSE data over DT, SVM, NB, and RF models

5.1.3. Relative absolute error

A statistical indicator of a prediction's accuracy in relation to the actual value is the RAE. By dividing the absolute error by the true value, it is calculated. The formula for RAE is (4). For our model performance evaluation, the RAE value is shown in Table 4, and a graphical representation is shown in Figure 8.

$$RAE = |(y_i - \hat{y}_i)|/y_i \tag{4}$$

Where y_i is the actual value at observation i, $\hat{y}_i$ is the predicted value at observation i.

Table 4. RAE value of voting model DT, SVM, NB, and RF

Dataset instance	SVM	DT	NB	RF
1-100	36.61	58.2	48	51.51
100-200	50.6	95.5	60.46	61.8
200-300	21.4	52.53	26.19	21.14
300-400	14.4	42.05	16.25	17.47

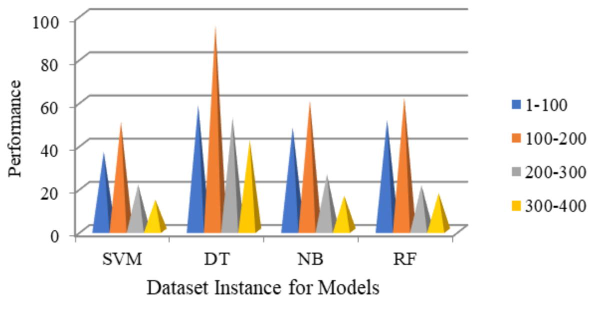

Figure 8. Graphical representation of RAE value over DT, SVM, NB, and RF models

5.1.4. Root relative squared error

A statistical indicator of a prediction's accuracy in relation to the mean value of the actual data is the RRSE. It is calculated by dividing the actual data's mean value by the RMSE. The formula for RRSE is (5). In our model performance evaluation, the RRSE value shown in Table 5 and its graphical representation are shown in Figure 9.

$$RRSE = RMSE/mean\ (actual\ values) \tag{5}$$

Table 5. RRSE value of voting model over DT, SVM, NB, and RF

Dataset instance	SVM	DT	NB	RF
1-100	87.7	73.7	95.44	82.42
100-200	103.7	102.66	109.31	92.62
200-300	65.59	69.38	71.81	49.48
300-400	53.87	57.51	56.7	47.36

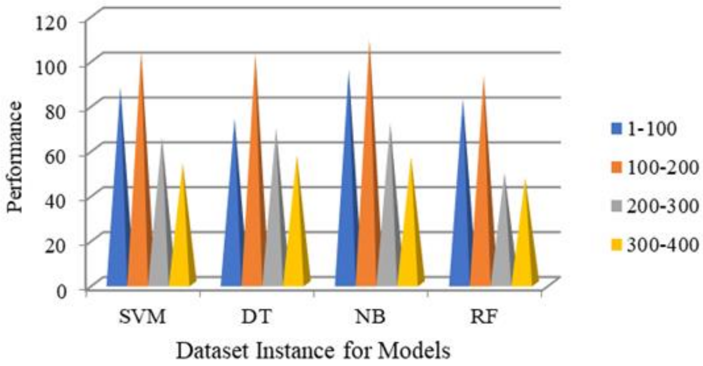

Figure 9. Graphical representation of RRSE value over DT, SVM, NB, and RF models

5.2. Boosting model

Boosting is a well-known ensemble ML technique that creates a robust learner by combining predictions from several weak learners, frequently DTs or shallow models. Its main goal is to increase model accuracy by giving cases that were previously erroneously classified more weight. The approach is developed in iterative steps, with a strong emphasis on correcting the mistakes made by earlier models, as shown in Figure 10.

Its characteristics include adaptability, sequential building, the elimination of bias, and vulnerability to potential over-fitting in variance finding. To reduce the risk of over-fitting, methods like early stopping and restricting the depth of poor learners are frequently used. Here, we observed that RF delivers the best prediction accuracy when compared to other models by using DT, SVM, and NB. In order to compare, we use various boosting models, such as AdaBoost, gradient boosting, XGBoost, LightGBM, and CatBoost for prediction over the dataset. After applying several boosting models over RF, we found the prediction accuracy, which is stored in Table 6, and its graphical representation is shown in Figure 11. Table 7 highlights the innovations and enhancements made in this research and offers a clear comparison between this work and the existing related works.

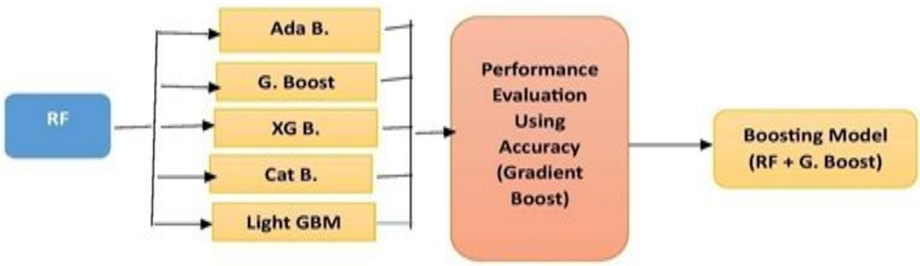

Figure 10. Boosting model architecture

Table 6. Accuracy table after applying boosting models over RF

Models	Accuracy (%)
AdaBoost	90.56
Gradient boost	96.01
XGBoost	92.47
CatBoost	90.56
LightGBM	85.20

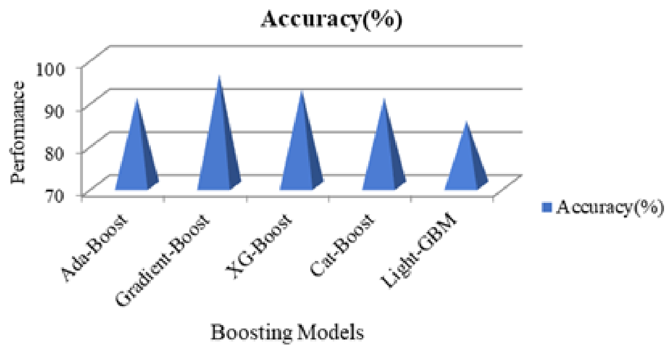

Figure 11. Graphical representation accuracy of boosting models

Table 7. Comparisons between this work with existing works

Study	Health condition	Methodology	Dataset	Accuracy (%)	Notable findings
Korkalainen *et al.* [3]	OSA	DL and diverse epoch durations	446 OSA patients	N/A	Revealed underestimation of sleep fragmentation in severe OSA and emphasizing the need for detailed sleep architecture analysis.
Kuo and Chen [5]	Insomnia	SVM and RCMSE analysis	32 subjects	89.31	Utilizes single-channel sleep EOG for short-duration insomnia detection and suitable for portable PSG systems.
Zulfiker *et al.* [7]	Insomnia	Multilayer stacking model and PCA	Socio-demographic dataset	88.60	Ensemble model surpasses leading classifiers and providing superior performance in predicting insomnia based on socio-demographic factors.
Alghwiri *et al.* [8]	Sleep quality	Logistic regression and RF	1,600 university students	74	Identified key factors affecting sleep quality, with RF showing the highest prediction accuracy.
This study	Sleep disorder	Statistical methods for data cleaning, PCA, Base ML models, and ML boosting models	Real-time dataset of 400 patients	96.01	Novel integration of various ML techniques for enhanced diagnostic accuracy on sleep disorder prediction, with a focus on early and reliable detection.

6. CONCLUSION

In order to tackle the difficult task of detecting sleeping disorders, this study has effectively applied ensemble learning algorithms, specifically voting and boosting, on the data dataset gathered by IoT sensors and manually. The strategy of this study involved combining four different base models into a voting ensemble: DT, SVM, NB, and RF. Using a dataset made up of 400 patient records, these base models demonstrated efficiency in improving the predictability of sleeping disorders. The boosting approach, which built on the advantages of the base models and added five boosting-learners: AdaBoost, gradient boosting, XGBoost, LightGBM, and CatBoost, was responsible for the real breakthrough. Gradient boosting excelled among them, attaining a remarkable accuracy rate of 96.01%, demonstrating the capability of boosting to enhance model performance. Future research has a number of potential directions, to be sure. Model robustness may be improved by enlarging the dataset and adding new characteristics, like sleep patterns and lifestyle variables. The accuracy of predictions may be further improved by investigating cutting-edge DL architectures to replace the neural network component.

FUNDING INFORMATION

This research constitutes purely academic work conducted by the authors without external financial support from any funding agency, grant organization, foundation, or corporate sponsor. All resources utilized, including computational infrastructure, dataset access, and research time, were provided through institutional facilities and personal contributions.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Satyaprakash Swain	✓	✓	✓		✓	✓	✓	✓	✓					
Binod Kumar Pattanayak	✓	✓			✓	✓		✓		✓	✓	✓	✓	
Mihir Narayan Mohanty	✓	✓			✓	✓		✓		✓	✓	✓	✓	
Amiya Kumar Sahoo	✓	✓	✓	✓		✓	✓			✓	✓		✓	
Suvendra Kumar Jayasingh	✓	✓	✓	✓			✓			✓	✓	✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest. This research was conducted using publicly available UCI Cleveland heart disease and Statlog datasets without any external funding or competing financial interests that could influence the reported findings.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

This article does not contain any studies with human participants or animals performed by any of the authors.

DATA AVAILABILITY

The data that support the findings of this study are available on request from the corresponding author [SKJ]. The data, which contain information that could compromise the privacy of research participants, are not publicly available due to certain restrictions.

REFERENCES

[1] K. Carmes, L. Kempfner, H. B. D. Sorensen, and P. Jennum, "Novel method for detection of sleep apnea using respiration signals," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2014, pp. 258–261, doi: 10.1109/EMBC.2014.6943578.

[2] M. M. S. Choyon, M. Rahman, M. M. Kabir, and M. F. Mridha, "IoT based health monitoring automated predictive system to confront COVID-19," in *2020 IEEE 17th International Conference on Smart Communities: Improving Quality of Life Using ICT, IoT and AI (HONET)*, Dec. 2020, pp. 189–193, doi: 10.1109/HONET50430.2020.9322811.

[3] H. Korkalainen *et al.*, "Detailed assessment of sleep architecture with deep learning and shorter epoch-to-epoch duration reveals sleep fragmentation of patients with obstructive sleep apnea," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 7, pp. 2567–2574, Jul. 2021, doi: 10.1109/JBHI.2020.3043507.

[4] M. Shahin, L. Mulaffer, T. Penzel, and B. Ahmed, "A two stage approach for the automatic detection of insomnia," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Jul. 2018, pp. 466–469, doi: 10.1109/EMBC.2018.8512360.

[5] C.-E. Kuo and G.-T. Chen, "A short-time insomnia detection system based on sleep EOG with RCMSE analysis," *IEEE Access*, vol. 8, pp. 69763–69773, 2020, doi: 10.1109/ACCESS.2020.2986397.

[6] M. M. Islam, A. K. M. Masum, S. Abujar, and S. A. Hossain, "Prediction of chronic Insomnia using machine learning techniques," in *2020 11th International Conference on Computing, Communication and Networking Technologies*, Jul. 2020, pp. 1–7, doi: 10.1109/ICCCNT49239.2020.9225570.

[7] M. S. Zulfiker, N. Kabir, A. A. Biswas, and P. Chakraborty, "Predicting insomnia using multilayer stacked ensemble model," in *Advances in Computing and Data Sciences (ICACDS 2021)*, Springer, Cham, 2021, pp. 338–350, doi: 10.1007/978-3-030-81462-5_31.

[8] A. A. Alghwiri, F. Almomani, A. A. Alghwiri, and S. L. Whitney, "Predictors of sleep quality among university students: the use of advanced machine learning techniques," *Sleep and Breathing*, vol. 25, no. 2, pp. 1119–1126, Jun. 2021, doi: 10.1007/s11325-020-02150-w.

[9] S. K. Jayasingh, J. K. Mantri, and S. Pradhan, "Weather prediction using hybrid soft computing models," in *Intelligent Systems*, Springer, Singapore, 2021, pp. 35–47, doi: 10.1007/978-981-33-6081-5_4.

[10] S. Swain *et al.*, "Application of IoT framework for prediction of heart disease using machine learning," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 10s, pp. 168–176, Oct. 2023, doi: 10.17762/ijritcc.v11i10s.7616.

[11] S. Mondal, R. Maity, Y. R. Singh, S. Ghosh, and A. Nag, "Early prediction of coronary heart disease using boosting-based voting ensemble learning," in *2022 IEEE Bombay Section Signature Conference*, 2022, pp. 1–5, doi: 10.1109/IBSSC56953.2022.10037445.

[12] K. L. Kumar and B. E. Reddy, "Heart disease detection system using gradient boosting technique," in *2021 International Conference on Computing Sciences*, Dec. 2021, pp. 228–233, doi: 10.1109/ICCS54944.2021.00052.

[13] S. S. G. K. Das, K. D. Mahato, C. Azad, and U. Kumar, "Heart disease prediction using different boosting models," in *2023 International Conference on Advanced & Global Engineering Challenges*, Jun. 2023, pp. 131–136, doi: 10.1109/AGEC57922.2023.00036.

[14] K. Sen and B. Verma, "Heart disease prediction using a soft voting ensemble of gradient boosting models, random forest, and Gaussian naive Bayes," in *2023 4th International Conference for Emerging Technology*, IEEE, May 2023, pp. 1–7, doi: 10.1109/INCET57972.2023.10170399.

[15] N. Afreen, R. Patel, M. Ahmed, and M. Sameer, "A novel machine learning approach using boosting algorithm for liver disease classification," in *2021 5th International Conference on Information Systems and Computer Networks*, Oct. 2021, pp. 1–5, doi: 10.1109/ISCON52037.2021.9702488.

[16] K. S. Bohnsack, M. Kaden, J. Abel, and T. Villmann, "Alignment-free sequence comparison: a systematic survey from a machine learning perspective," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 20, no. 1, pp. 119–135, Jan. 2023, doi: 10.1109/TCBB.2022.3140873.

[17] N. Ramteke and P. Maidamwar, "Cardiac patient data classification using ensemble machine learning technique," in *2023 14th International Conference on Computing Communication and Networking Technologies*, Jul. 2023, pp. 1–6, doi: 10.1109/ICCCNT56998.2023.10307702.

[18] F. R. On, R. Jailani, S. L. Hassan, and N. M. Tahir, "Analysis of sparse PCA using high dimensional data," in *2016 IEEE 12th International Colloquium on Signal Processing & Its Applications*, Mar. 2016, pp. 340–345, doi: 10.1109/CSPA.2016.7515857.

[19] P. Jindal, P. Saini, S. Jain, and M. Chandraprabha, "Design and development of cardiovascular disease prediction system using voting classifier," in *2022 4th International Conference on Advances in Computing, Communication Control and Networking*, Dec. 2022, pp. 876–881, doi: 10.1109/ICAC3N56670.2022.10074223.

[20] J. Filosa, P. M. Omland, K. Hagen, K. Langsrud, M. Engstrøm, and T. Sand, "Validation of trøndelag apnea score proxy for obstructive sleep apnea in the general population of Norway: the HUNT study," *Sleep Disorders*, vol. 2024, no. 1, Jan. 2024, doi: 10.1155/2024/1242505.

[21] T. Marjot, D. W. Ray, F. R. Williams, J. W. Tomlinson, and M. J. Armstrong, "Sleep and liver disease: a bidirectional relationship," *The Lancet Gastroenterology & Hepatology*, vol. 6, no. 10, pp. 850–863, Oct. 2021, doi: 10.1016/S2468-1253(21)00169-2.

[22] C. Formentin, M. Garrido, and S. Montagnese, "Assessment and management of sleep disturbance in cirrhosis," *Current Hepatology Reports*, vol. 17, no. 1, pp. 52–69, Mar. 2018, doi: 10.1007/s11901-018-0390-1.

[23] A. E. Abdullah *et al.*, "Symptoms of daytime sleepiness and sleep apnea in liver cirrhosis patients," *Annals of Hepatology*, vol. 16, no. 4, pp. 591–598, Jul. 2017, doi: 10.5604/01.3001.0010.0304.

[24] S. Saeda, K. Fujiwara, Y. Sumi, and H. Kadotani, "Heart rate variability-based model for estimating the severity of orthostatic hypotension in patients with REM sleep behavior disorder," in *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Jul. 2024, pp. 1–4. doi: 10.1109/EMBC53108.2024.10781845.

[25] N. Mandas *et al.*, "Characterization of sleep structure and autonomic dysfunction in REM sleep behavior disorder," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 5, pp. 859–866, 2024, doi: 10.1109/OJEMB.2024.3397550.

BIOGRAPHIES OF AUTHORS

Dr. Satyaprakash Swain is currently working as an assistant professor of Computer Science and Engineering at IMIT, Cuttack, under BPUT, Odisha, India. His research interests include WSN, natural language processing, and applications of machine learning techniques and IoT for real-world problems. He can be contacted at email: satyaimit@gmail.com.

Prof. Binod Kumar Pattanayak is a professor in the Department of Computer Science and Engineering, Institute of Technical Education and Research, Siksha 'O' Anusandhan University, Bhubaneswar, India, since 1999. His research interests include ad-hoc networks, compiler design, software engineering, computer architecture, and intelligent systems. He can be contacted at email: binodpattanayak@soa.ac.in.

Prof. Mihir Narayan Mohanty is a professor in the Department of ECE, ITER, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha. His research area includes signal processing and machine learning. He can be contacted at email: mihirmohanty@soa.ac.in.

Prof. Amiya Kumar Sahoo is currently working as a professor in the Department of Computer Science and Engineering, Aryan Institute of Engineering and Technology, Bhubaneswar, Odisha, India since 2010. His research interests include text mining, AI and ML, cloud computing, and IoT. He can be contacted at email: dramiya79@gmail.com.

Prof. Suvendra Kumar Jayasingh is presently working as an associate professor at Institute of Management and Information Technology, Cuttack, a Constituent College of BPUT, Odisha, India. His main research work focuses on machine learning, data mining, database design, computational intelligence, and soft computing. He can be contacted at email: sjayasingh@gmail.com.