International Journal of Artificial Intelligence and Computer Electronics
Volume: 02, Issue: 03, July-September 2026, pp:79-88
P-ISSN 3117-8529 | E-ISSN 3117-8537

Applying Machine Learning Algorithm For The Classification Of Sleep Disorders

Faiz Mohammed Khan[1], Dr. Syed Asadullah Hussaini[2]
[1]PG Scholar, Department of Computer Science & Engineering ISL Engineering College, Hyderabad, India.
[2]Associate Professor, Department of Computer Science & Engineering ISL Engineering College, Hyderabad, India.
Corresponding Author Email: owaiskhan2165@gmail.com

Article Received 18-05-2026, Revised 25-06-2026, Accepted 13-08-2026

ABSTRACT
Sleep disorders such as Insomnia and Sleep Apnea significantly affect health, well-being and daily functioning. The project develops a machine-learning based classification system that predicts sleep-disorder categories from demographic, sleep, lifestyle and vital-health information. The source thesis describes a dataset of 400 samples with 13 relevant features and three target classes: Insomnia, None and Sleep Apnea.
The proposed approach uses a Random Forest classifier. The model is selected because ensemble learning reduces the risk of overfitting, supports mixed data, requires comparatively modest computational resources and provides feature-importance information. The thesis reports an accuracy of 95% and evaluates the model using accuracy, precision, recall, F1-score, confusion matrix and ROC analysis.

Keywords: *Sleep Disorders, Insomnia, Sleep Apnea, Machine Learning, Random Forest, Classification, Demographic Features, Sleep Patterns, Lifestyle Factors, Vital Health Data, Feature Importance, ROC Analysis.*

1. Introduction

Sleep is an essential biological process that supports physical recovery, cognitive performance, emotional balance, and normal daily functioning. Disturbances in sleep duration, continuity, or quality can affect an individual's health and productivity and may develop into clinically significant sleep disorders. Among the commonly observed conditions are insomnia and obstructive sleep apnea, both of which can substantially influence an individual's quality of life. The increasing availability of health and lifestyle data has created opportunities to investigate computational approaches for identifying sleep-related conditions at an early stage.

Traditional sleep-disorder assessment generally depends on clinical consultation, questionnaires, physiological measurements, and, in some cases, specialized sleep studies. Although these approaches can provide valuable diagnostic information, they may require considerable time, specialized equipment, or clinical expertise. Machine learning offers an alternative data-driven approach in which relationships between demographic characteristics, lifestyle patterns, sleep behavior, and health measurements can be learned from previously observed cases. Such approaches have increasingly been investigated for automated sleep-stage and sleep-disorder classification [3].

Different machine-learning algorithms have been applied to sleep-related prediction problems. Support Vector Machines, Random Forests, decision trees, nearest-neighbor methods, neural networks, and ensemble approaches have all been examined for their ability to recognize patterns associated with abnormal sleep conditions. Comparative investigations indicate that algorithm performance can vary according to the characteristics of the dataset, selected features, and classification objective [17]. Consequently, selecting an appropriate model for a particular sleep-health dataset is an important component of developing a reliable classification system.

Random Forest is particularly attractive for classification tasks involving heterogeneous health and lifestyle information. The algorithm constructs multiple decision trees using randomized subsets of observations and features and combines their outputs to obtain the final prediction. This ensemble structure can reduce the sensitivity of the model to individual decision boundaries and can provide useful estimates of feature importance. These characteristics make Random Forest suitable for datasets containing a mixture of demographic, behavioral, sleep-related, and health variables [14].

Previous research has also demonstrated the application of machine learning to physiological signals for sleep-related assessment. Salari et al. investigated sleep-apnea detection using machine-learning algorithms based on ECG signals, demonstrating the potential of computational models to identify patterns associated with abnormal breathing during sleep [5]. Kim et al. similarly examined machine-learning prediction models for obstructive sleep apnea in Korean adults, indicating the usefulness of data-driven approaches for identifying individuals at risk of sleep-related disorders [15].

The emergence of publicly available sleep-health datasets has further encouraged research into automated classification. A dataset containing demographic, lifestyle, sleep, and health-related variables can provide a relatively accessible environment for evaluating machine-learning methods without requiring specialized physiological equipment. Hidayat investigated Random Forest-based classification using a sleep-health and lifestyle dataset, providing a direct methodological basis for applying tree-based ensemble learning to sleep-disorder categories [14].

Feature selection and representation are also important because health datasets can contain variables with different scales, distributions, and levels of relevance. Variables such as age, occupation, sleep duration, sleep quality, physical activity, stress level, heart rate, and daily habits may collectively contribute to the distinction between healthy and disordered sleep patterns. A model capable of identifying nonlinear interactions among these characteristics can therefore be advantageous over approaches that rely exclusively on individual variables.

Machine-learning software frameworks have simplified the development and evaluation of such classification systems. Pedregosa et al. introduced Scikit-learn as a Python-based machine-learning library containing implementations of widely used algorithms and evaluation procedures [2]. Its practical capabilities make it suitable for preprocessing, model development, cross-validation, prediction, and performance measurement in health-related classification studies.

The present research focuses on a dataset containing **400 observations and 13 relevant features**, with three target categories: **Insomnia, None, and Sleep Apnea**. A Random Forest classifier is employed to learn the relationship between the available demographic, sleep-related, lifestyle, and vital-health variables and the corresponding sleep-disorder category. The model is evaluated using accuracy, precision, recall, F1-score, confusion-matrix analysis, and ROC-based assessment.

The principal objective of the study is to determine whether a Random Forest-based classification framework can reliably distinguish among the three sleep-health categories using routinely available structured information. In addition to measuring overall predictive performance, the study examines the classification behavior of individual categories and the contribution of input variables to the final prediction. Such an approach can provide a practical foundation for data-driven preliminary sleep-health assessment while recognizing that machine-learning predictions should complement rather than replace professional clinical evaluation.

2. Literature Review

2.1 Machine Learning for Sleep Classification

Satapathy et al. investigated the performance of machine-learning algorithms using feature sets designed for automated sleep staging. Their research demonstrates that computational classification can extract meaningful patterns from sleep-related measurements and distinguish different sleep conditions or stages. The study is relevant to the present work because it establishes the broader applicability of supervised learning to automated sleep analysis and highlights the importance of selecting suitable features for classification performance [3].

Alickovic and Subasi examined an ensemble Support Vector Machine approach for automatic sleep-stage classification. Their study demonstrated the potential of combining machine-learning techniques to improve automated recognition of sleep stages. The use of an ensemble strategy is particularly relevant to the current research because Random Forest follows the same general principle of combining multiple individual learners to obtain a more stable classification decision [8].

Van Der Donckt et al. investigated traditional machine-learning approaches for sleep-related classification and challenged the assumption that increasingly complex models are always necessary for biomedical prediction. Their findings highlight the continuing value of conventional machine-learning algorithms when the available features are informative and appropriately prepared. This observation supports the selection of Random Forest for structured sleep-health data rather than assuming that a deep neural architecture is automatically superior [12].

2.2 Sleep Apnea Detection Using Machine Learning

Salari et al. investigated the detection of sleep apnea using machine-learning algorithms based on ECG signals. Their study demonstrates how physiological information can be converted into predictive features for identifying sleep apnea. Although the present research uses structured demographic, lifestyle, and health variables rather than ECG signals, the underlying objective is similar: identifying patterns associated with sleep-related disorders through supervised computational learning [5].

Kim et al. developed machine-learning-based prediction models for obstructive sleep apnea among Korean adults. Their research demonstrates that demographic and health-related characteristics can be useful for estimating the likelihood of sleep apnea. The study provides support for incorporating routinely measurable patient characteristics into predictive models and reinforces the potential of machine learning for preliminary identification of individuals who may require further sleep-related assessment [15].

Tripathi et al. investigated an ensemble computational-intelligence approach for insomnia and sleep-stage detection using sleep ECG signals. Their work illustrates the value of ensemble learning in processing physiological information and distinguishing different sleep-related states. The findings are relevant to the present study because Random Forest also combines multiple decision structures rather than relying on a single classifier, potentially improving robustness in heterogeneous datasets [8].

2.3 Random Forest and Comparative Classification

Hidayat applied Random Forest to the classification of sleep disorders using a sleep-health and lifestyle dataset. The study is directly related to the present research because it considers structured information describing sleep and lifestyle conditions and uses Random Forest as the primary classification mechanism. Its approach demonstrates the suitability of ensemble decision-tree methods for identifying sleep-disorder categories from non-invasive, structured health information [14].

Bansal, Goyal, and Choudhary conducted a comparative analysis involving K-nearest neighbor, genetic algorithms, Support Vector Machines, decision trees, and Long Short-Term Memory models. Their investigation illustrates that algorithm selection can have a substantial effect on predictive outcomes and that different learning techniques have different strengths depending on the characteristics of the problem. This supports the need to evaluate Random Forest empirically rather than selecting an algorithm solely on the basis of theoretical complexity [17].

Pedregosa et al. developed Scikit-learn, a Python machine-learning library that provides implementations of numerous classification and evaluation algorithms. The framework has become widely useful for experimental machine-learning workflows because it supports preprocessing, model training, prediction, validation, and performance evaluation within a common programming environment. Its availability facilitates reproducible implementation of classification experiments such as the Random Forest model used in this study [2].

2.4 Feature Learning and Sequential Health Data

Ordóñez and Roggen investigated deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. Although their application is focused on activity recognition rather than sleep disorders, the study demonstrates how physiological and behavioral observations can be transformed into machine-learning representations for human-state classification. This provides methodological support for using behavioral and lifestyle variables as informative inputs in health-related prediction systems [10].

Li et al. investigated adversarial learning for semi-supervised pediatric sleep staging using a single EEG channel. Their work demonstrates the potential of advanced learning strategies to extract useful information from limited physiological observations. The study also reflects the continuing development of automated sleep analysis, where computational models are increasingly being explored as tools for reducing dependence on manual interpretation of physiological data [20].

2.5 Evaluation of Machine-Learning Models

Powers examined commonly used classification measures, including precision, recall, and F-measure, while also discussing ROC-related evaluation concepts. These measures are particularly important for health-related classification because overall accuracy alone may not adequately describe how a model performs for individual categories. In the present study, precision, recall, F1-score, confusion-matrix analysis, and ROC evaluation are therefore considered alongside overall accuracy to provide a more complete assessment of the Random Forest classifier [24].

2.6 Research Gap and Motivation

The reviewed literature demonstrates that machine learning has been applied extensively to sleep staging, sleep apnea detection, insomnia analysis, and related biomedical classification problems. However, many existing studies depend on physiological signals such as ECG or EEG, specialized measurements, or narrowly defined sleep-stage datasets [5], [20]. There is comparatively greater practical value in investigating whether commonly available demographic, lifestyle, sleep-pattern, and vital-health information can support the classification of multiple sleep-disorder categories within a single framework.

The present study addresses this requirement by applying a Random Forest classifier to a structured dataset containing 400 samples, 13 relevant features,

International Journal of Artificial Intelligence and Computer Electronics
Volume: 02, Issue: 03, July-September 2026, pp:79-88
P-ISSN 3117-8529 | E-ISSN 3117-8537

and three target categories: Insomnia, None, and Sleep Apnea. The approach emphasizes an interpretable ensemble model capable of handling heterogeneous variables while providing feature-importance information. The model is evaluated through multiple complementary metrics rather than relying solely on accuracy, thereby providing a broader assessment of its classification capability.
Overall, the literature indicates that both conventional and advanced machine-learning methods can provide useful support for automated sleep-health analysis. The proposed work builds upon this foundation by examining a practical Random Forest-based classification framework using structured sleep-health and lifestyle information. The resulting model is intended as a data-driven assessment tool that can identify patterns associated with different sleep-disorder categories and potentially support early screening and further clinical investigation.

3. Dataset And Methodology

The proposed sleep-disorder classification system is developed using demographic, occupational, lifestyle, sleep-related, and health-related attributes. The dataset contains information required to distinguish among three target categories: Insomnia, None, and Sleep Apnea. Person ID serves only as an identifier for individual records. Gender represents the participant as male or female, while Age records the individual's age in years. Occupation identifies the person's profession. Sleep Duration represents the number of hours of sleep obtained per day, and Quality of Sleep records a subjective sleep-quality assessment on a scale ranging from 1 to 10. Physical Activity represents the number of minutes of physical activity performed per day, while Stress Level records a subjective stress score from 1 to 10. The dataset also incorporates health-related characteristics that may contribute to sleep-disorder classification. BMI Category represents weight-related categories such as underweight, normal, and overweight. Blood Pressure contains systolic and diastolic measurements, whereas Heart Rate represents the resting heart rate in beats per minute. Daily Steps records the number of steps completed by an individual each day. The target variable is Sleep Disorder, which categorizes each observation as None, Insomnia, or Sleep Apnea. These attributes collectively provide the input information required by the Random Forest classification model.

Data Processing Methodology

The methodology follows a sequential workflow beginning with data collection and exploratory analysis. The collected information is examined to understand the characteristics of the dataset and identify relevant patterns. Feature selection or engineering is subsequently performed to prepare suitable inputs for the prediction model. The processed dataset is divided into training and testing subsets, after which the Random Forest classifier is trained using the training data. The trained model is then evaluated using appropriate classification measures before being integrated into the prediction application.
The workflow can be represented as Data Collection → Exploratory Data Analysis → Feature Selection/Engineering → Train/Test Split → Random Forest Training → Model Evaluation → Deployment and Prediction → User Interface → Testing and Monitoring. The thesis specifies an 80:20 training-to-testing split as an example. Cross-validation is also considered for examining model stability and reducing the possibility of overfitting. Model performance is assessed using accuracy, precision, recall, F1-score, classification report, and confusion matrix analysis.

Random Forest Algorithm

Random Forest is an ensemble-learning technique that generates a collection of decision trees and combines their predictions to produce a classification result. Instead of relying on a single tree, the algorithm creates multiple trees using bootstrap samples obtained from the training dataset. At each decision point, only a randomly selected subset of features is considered for determining the next split. Consequently, the individual trees learn from different combinations of observations and features.
The Random Forest procedure begins with the preparation of input features and their corresponding target labels. Bootstrap sampling is then applied to create multiple training subsets. A random feature subset is considered at each tree split, followed by independent training of the individual decision trees. After training, a new observation is passed through all available trees. The predictions are aggregated using majority voting, and the class with the highest vote count is returned as the final output.

Comparison Between ANN and Random Forest

The proposed Random Forest model is considered as a practical alternative to an Artificial Neural Network for the tabular sleep-health dataset. ANN-based systems can involve comparatively greater computational complexity and hardware requirements, particularly when extensive training and parameter adjustment are required. Their internal decision processes can also be more difficult to interpret. Random Forest, in contrast, has comparatively lower computational requirements and can operate efficiently on standard CPU systems.
Another advantage of Random Forest is its ensemble structure. By combining predictions from several decision trees, the model reduces dependence on a

International Journal of Artificial Intelligence and Computer Electronics
Volume: 02, Issue: 03, July-September 2026, pp:79-88
P-ISSN 3117-8529 | E-ISSN 3117-8537

single tree and can lower the risk of overfitting. The availability of feature-importance information also provides an additional mechanism for interpreting the contribution of input variables. Furthermore, Random Forest generally requires comparatively simpler tuning than an ANN-based architecture. These characteristics support its selection for the proposed sleep-disorder classification task.

4. System Design

The system design establishes the software structure required to implement the sleep-disorder prediction application. The design translates the project requirements into defined components, interactions, data-processing stages, and execution elements. UML-based and data-flow representations are used to describe different aspects of the system. The Use Case Diagram represents the relationship between the user and system functions, while the Class Diagram defines classes, attributes, and methods. The Object Diagram illustrates object-level relationships.

State and Activity Diagrams are used to represent system states, workflows, and control activities. Sequence and Collaboration Diagrams describe interactions and message exchange among system components. Component and Deployment Diagrams represent software components and their execution environment. In addition, the Data Flow Diagram describes how information moves through the application during collection, processing, storage, and distribution.

System Architecture

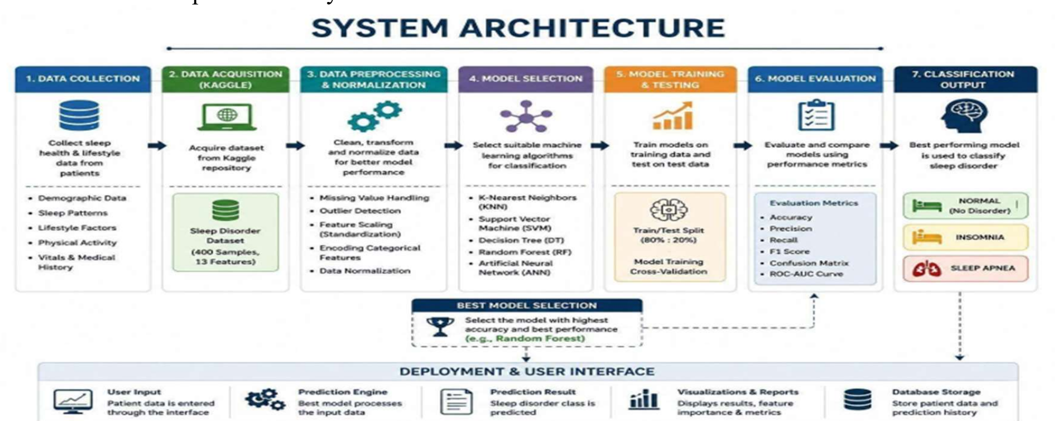
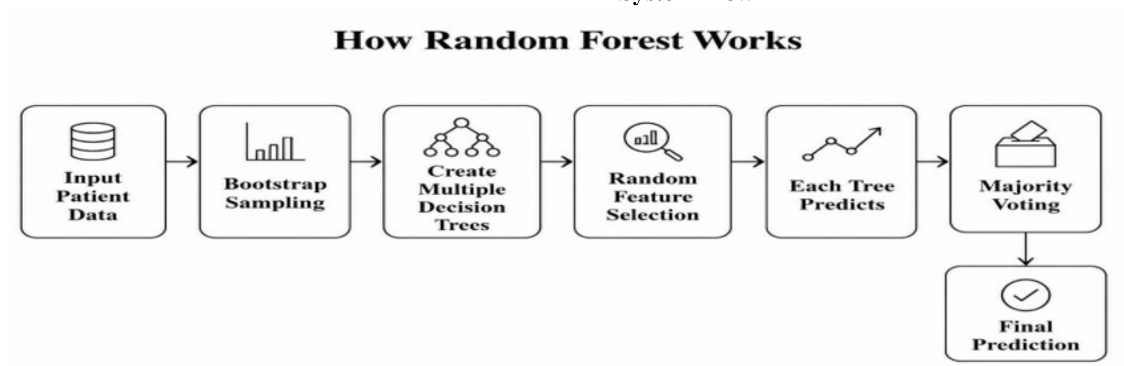


Fig ;1 System Architecture

The proposed architecture provides a web-based interface through which users submit demographic, health, lifestyle, and sleep-related information. The required inputs include gender, age, occupation, sleep duration, quality of sleep, physical activity, stress level, BMI category, blood pressure, heart rate, and daily steps. Once the information is entered, it is prepared according to the requirements of the trained model and supplied to the Random Forest classifier. The classifier processes the input using multiple decision trees. Each tree produces an individual prediction, and majority voting is applied to combine these predictions. The resulting class is then displayed to the user through the web interface. The possible outcomes are **Insomnia, None, and Sleep Apnea**. This architecture is intended to provide rapid classification and data-driven decision support.

System Flow



Fig; 2 System Flow

International Journal of Artificial Intelligence and Computer Electronics
Volume: 02, Issue: 03, July-September 2026, pp:79-88
P-ISSN 3117-8529 | E-ISSN 3117-8537

The operational flow begins with the **User**, who accesses the system through **Login/Authentication**. After successful access, the user provides the required **Health Details**. These details undergo **Preprocessing** before being supplied to the **Random Forest Model**. The model generates predictions through its individual decision trees, and **Majority Voting** determines the final category. The resulting **Predicted Class** is then presented through **Result Visualization**.

Development Tools And Implementation

Python and Supporting Libraries

Python is used as the primary programming language for implementing the machine-learning model and application. Its readability, interactive execution capability, object-oriented programming support, and broad collection of libraries make it suitable for the development of the proposed system. The implementation uses **NumPy, Pandas, Matplotlib, Scikit-learn, and Django**.

NumPy is used for numerical arrays and computational operations. Pandas supports data analysis and DataFrame manipulation. Matplotlib is used to generate plots and graphical representations of results. Scikit-learn provides machine-learning functionality and model-evaluation methods. Django serves as the web application framework through which the prediction interface is implemented.

System Implementation

The implemented application loads a previously serialized Random Forest model and scaler. The Django application includes separate views for the home page, user registration, login, input collection, and prediction. During prediction, the application receives the information entered by the user and constructs the feature vector required by the trained model. The saved scaler is applied to the input before the processed values are passed to the Random Forest classifier.

The source implementation maps the model outputs to the three sleep-disorder categories. Output **0** corresponds to **Insomnia**, output **1** corresponds to **None**, and another output corresponds to **Sleep Apnea**. The predicted class is subsequently returned to the web page and displayed to the user.

Hardware and Software Requirements

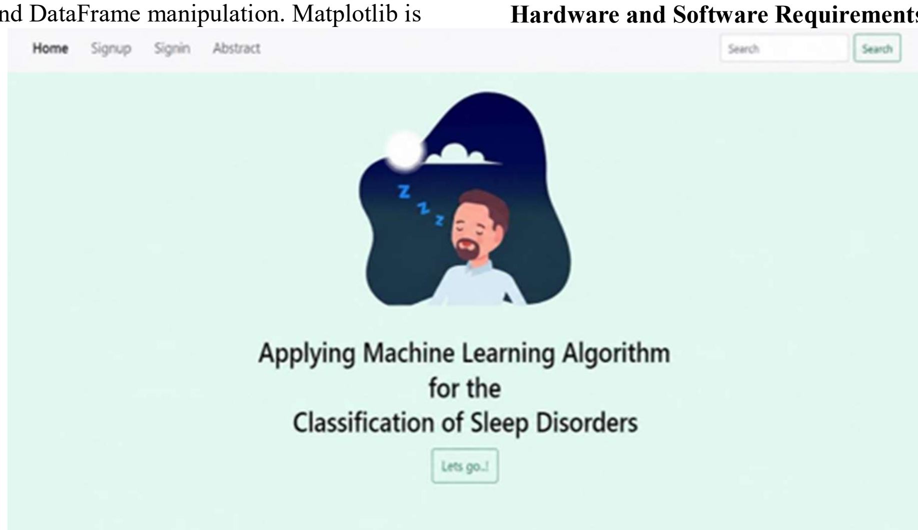


Fig. 3 Home screen of the implemented application

The hardware requirements specified in the thesis consist of a dual-core processor, 4 GB RAM, and a 250 GB hard disk. The software environment includes a Windows operating system, Python, and the development environment identified in the report. These resources provide the computational environment required for developing and executing the proposed machine-learning application.

Application Home and Registration Interface

The home page functions as the primary entry point to the Sleep Disorder Classification System. It provides navigation options including **Home, Signup, Signin, and Abstract**. A **“Let's go..!”** option is provided for accessing the sleep-disorder prediction functionality. The interface also displays the project title and an illustration associated with sleep.

The registration module allows new users to create an account before accessing the prediction functionality. The registration form collects basic information such as username, email address, and password. After submission, the application verifies the provided information and stores the registration details in the database. Once registration is completed, the user can sign in and access the prediction system. This module provides authenticated and personalized access to the application.

International Journal of Artificial Intelligence and Computer Electronics
Volume: 02, Issue: 03, July-September 2026, pp:79-88
P-ISSN 3117-8529 | E-ISSN 3117-8537

5. Results

Health-Condition Input Interface

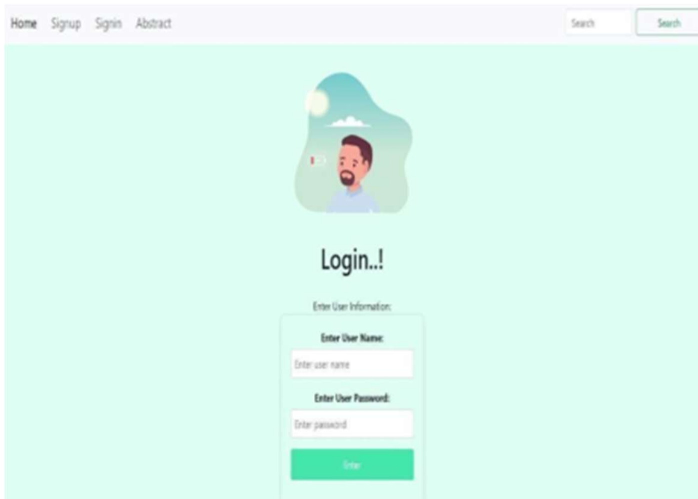


Fig. 4 Health-condition input screen

The health-condition input screen provides fields for entering the variables required by the trained Random Forest model. Users provide the relevant demographic, lifestyle, sleep, and health information through the web interface. After all required values have been entered, the **Predict** option initiates the classification process. The submitted information is processed and supplied to the Random Forest model to obtain the predicted sleep-disorder category.

Prediction and Visualization

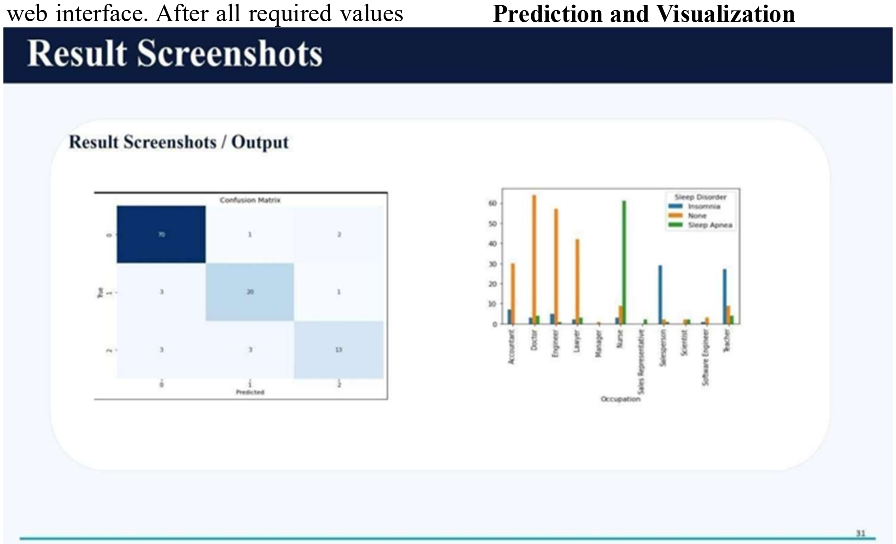


Fig. 5 Prediction-result visualization and confusion matrix

The prediction interface displays the classification generated by the trained model. According to the thesis, the Random Forest classifier achieves an overall **95% accuracy**. The result presentation includes a confusion matrix and graphical visualization of the predicted sleep-disorder categories. The confusion matrix shows that the majority of observations are positioned on the correct classification categories, while a relatively small number of observations appear as classification errors in the off-diagonal positions.

The visualization provides an additional representation of the classification results and supports interpretation of the model's behavior across the three target categories. The combination of numerical performance measures and graphical outputs provides a more comprehensive view of the effectiveness of the proposed classifier.

International Journal of Artificial Intelligence and Computer Electronics
Volume: 02, Issue: 03, July-September 2026, pp:79-88
P-ISSN 3117-8529 | E-ISSN 3117-8537

Model Evaluation, Testing And Future Enhancement

ROC Analysis

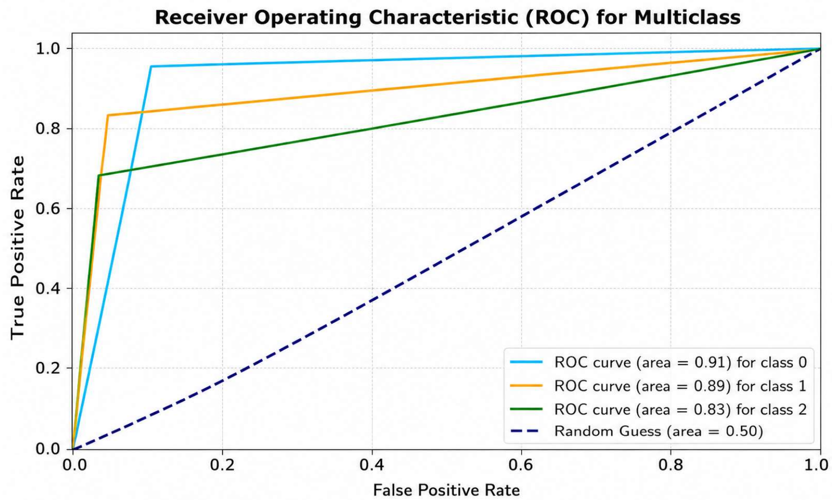


Fig. 6 ROC analysis

Receiver Operating Characteristic analysis is used to examine the discrimination capability of the multiclass classification model. The thesis reports ROC AUC values of **0.91, 0.89, and 0.83** for the three sleep-disorder classes. A ROC curve positioned closer to the upper-left region represents stronger discriminatory performance. The reported AUC values therefore provide additional evidence for evaluating the model's ability to distinguish among the target categories.

Software Testing

Software testing is conducted to identify implementation errors and determine whether the developed application satisfies its intended requirements. The testing strategy described in the report includes unit testing, functional testing, system testing, performance testing, integration testing, and acceptance testing. Unit testing examines individual components, whereas functional testing verifies system behavior under valid and invalid input conditions.

System testing evaluates the complete application after integration, while integration testing examines the interaction between different modules. Performance testing considers application response time and execution behavior. Acceptance testing provides user-oriented validation of the developed system. Collectively, these testing levels are intended to verify the functionality and reliability of the implemented application.

Future Enhancements

Future development can focus on increasing the size and diversity of the dataset by incorporating information from multiple hospitals and sleep clinics. Such expansion can support improved generalization of the model. The system can also be extended to include additional sleep disorders, including **Narcolepsy, Restless Legs Syndrome (RLS), and Circadian Rhythm disorders**.

Further research can compare Random Forest with alternative algorithms such as **XGBoost, LightGBM, CatBoost, SVM, LSTM, and CNN**. Automated hyperparameter optimization can be incorporated to identify suitable model configurations. Integration with wearable-device information and development of a mobile application can extend the accessibility of the system. Cloud deployment, EHR integration, and Explainable AI approaches such as **SHAP and LIME** can also be considered.

Additional improvements may include enhanced dashboards, improved security, multilingual support, and continuous learning capabilities. Finally, evaluation using larger real-world clinical datasets and clinical studies is necessary to examine the performance of the system across broader populations.

6. Conclusion

The proposed work develops a machine-learning-based system for classifying sleep disorders using demographic, health, lifestyle, and sleep-related information. Random Forest is employed as a practical alternative to an ANN-based approach for the tabular dataset considered in the project. The ensemble mechanism combines multiple decision trees and determines the final class through majority voting. In addition, feature-importance information provides greater interpretability than a prediction process based solely on an opaque model structure.

The thesis reports that the Random Forest model achieved **95% classification accuracy** across the three categories of **Insomnia, None, and Sleep Apnea**. The developed web application allows users to enter the required information and obtain a predicted sleep-disorder category. The reported results demonstrate the potential of the proposed approach as a rapid, data-driven decision-support system for preliminary sleep-disorder classification. Nevertheless, broader deployment requires further evaluation using larger and more diverse datasets. Expansion to additional sleep-disorder categories, comparison with other machine-learning and deep-learning methods, integration of wearable information, and validation through real-world clinical datasets are identified as important future directions. These developments can strengthen the applicability and generalizability of the proposed sleep-disorder classification framework.

References

[1]. Sannidhanam, A. H. (2026). LLM-Driven Workflow Optimization: Intelligent Routing in Asynchronous Distributed Systems. *International Journal of AI and Machine Learning, 1*(2), 23–33.

[2]. F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, Nov. 2011.

[3]. S. Satapathy et al., “Performance analysis of machine learning algorithms on automated sleep staging feature sets,” 2021.

[4]. Dr. Abdul Bari, Dr. Imtiyaz Khan, Dr. Rafath Samrin, Dr. Akhil Khare, “*VPC & Public Cloud Optimal Performance in Cloud Environment*,” *Educational Administration: Theory and Practice*, ISSN No.: 2148-2403, vol. 30, issue 6, June 2024.

[5]. N. Salari et al., “Detection of sleep apnea using machine learning algorithms based on ECG signals,” *Expert Systems with Applications*, 2022.

[6]. Performance Analysis of Wireless Sensor Networks for Smart Monitoring Applications. (2016). *International Journal of Humanities and Information Technology, 1*(04), 10–29. doi:10.21590/ijhit.01.04.04

[7]. A. H. Sannidhanam. (2024). Guardrails and Safety Mechanisms for LLM-Powered Enterprise Applications. *International Journal of Engineering Science & Humanities, 14*(3), 273–285.

[8]. E. Alickovic and A. Subasi, “Ensemble SVM method for automatic sleep stage classification,” *IEEE Transactions on Instrumentation and Measurement*, 2018.

[9]. M. Q. Hatem, “Skin lesion classification system using a K-nearest neighbor algorithm,” *Visual Computing for Industry, Biomedicine, and Art*, vol. 5, no. 1, pp. 1–10, Dec. 2022. *Sleep Health and Lifestyle Dataset*, 2023. Available through Kaggle.

[10]. F. Ordóñez and D. Roggen, “Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition,” *Sensors*, vol. 16, no. 1, p. 115, Jan. 2016.

[11]. Distributed Data Processing Frameworks for Large-Scale Healthcare Analytics. (2019). *International Journal of Advance Industrial Engineering, 7*(04), 1–10. doi:10.14741/ijaie/v.7.4.01

[12]. J. Van Der Donckt et al., “Do not sleep on traditional machine learning,” *Biomedical Signal Processing and Control*, 2023.

[13]. Mrs. Misbah Kousar, Dr. Sanjay Kumar, Dr. Mohammed Abdul Bari**,** “*Design of a Decentralized Authentication and Off-Chain Data Management Protocol for VANETs Using Blockchain*,” *Communications on Applied Nonlinear Analysis*, ISSN: 1074-133X, vol. 32, no. 2, 2025.

[14]. I. A. Hidayat, “Classification of sleep disorders using random forest on sleep health and lifestyle dataset,” *J. Dinda: Data Science, Information Technology, and Data Analytics*, vol. 3, no. 2, pp. 71–76, Aug. 2023.

[15]. Y. J. Kim et al., “Prediction models for obstructive sleep apnea in Korean adults using machine learning techniques,” *Diagnostics*, 2021.

[16]. Sannidhanam, A. H. (2025). Autonomous AI Agents for Cloud Infrastructure Operations. *Journal of Contemporary Science and Technology Management, 1*(01), 83–100.

[17]. M. Bansal, A. Goyal, and A. Choudhary, “A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short-term memory algorithms in machine learning,” *Decision Analytics Journal*, vol. 3, Jun. 2022, Art. no. 100071. *Sleep Health and Lifestyle Dataset*, 2023.

[18]. L. K. Suresh Kumar and Mohammed Abdul Bari, “Zero-Day Attack Detection in Multi-Tenant Cloud Environments Using Variational Autoencoders,” *International Journal of Applied Mathematics*, ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version), vol. 38, no. 3s, 2025.

[19]. Sannidhanam, A. H. (2020). Scalable Web Services Architecture for Healthcare Data Processing. *International Journal of Recent Advances in Science and Technology, 7*(01), 1–14.

[20]. Y. Li et al., "Adversarial learning for semi-supervised pediatric sleep staging with single-EEG channel," *Methods*, 2022.

[21]. S. Akbar, A. Ahmad, M. Hayat, A. U. Rehman, S. Khan, and F. Ali, "IAtbP-Hyb-EnC: Prediction of antitubercular peptides via heterogeneous feature representation and genetic algorithm based ensemble learning model," *Computers in Biology and Medicine*, vol. 137, Oct. 2021, Art. no. 104778.

[22]. Event Streaming Architectures for High-Volume Transaction Processing. (2022). *International Journal of Advance Industrial Engineering, 10*(01), 1–8. doi:10.14741/

[23]. Afsha Nishat, Dr. Mohd Abdul Bari**,** Dr. Guddi Singh, "*Mobile Ad Hoc Network Reactive Routing Protocol to Mitigate Misbehavior Node*," *International Journal Of Intelligent Systems And Applications In Engineering*, IJUSEA, ISSN: 2147-6799, Nov. 2023.

[24]. D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," 2020, arXiv:2010.16061.

[25]. Sannidhanam, A. H. (2020). Scalable Web Services Architecture for Healthcare Data Processing. *International Journal of Recent Advances in Science and Technology, 7*(01), 1–14.

[26]. I. A. Hidayat, "Classification of sleep disorders using random forest on sleep health and lifestyle dataset," 2023.

[27]. M. S. Uddin and M. D. Zainlabuddin, "Secure video processing and watermark embedding using Shamir secret rule," *International Journal of Artificial Intelligence, Computing and Electronics*, vol. 2, no. 1, pp. 25–31, Mar. 2026.

[28]. Anjani Haritha Sannidhanam. (2023). Self-Healing Distributed Systems: AI-Driven Failure Prediction and Automated Recovery. *International Journal of Research & Technology, 11*(4), 168–174.

[29]. F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, 2011.