International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Applying Machine Learning Algorithms for the Classification of Sleep Disorders

Duvvala Ajay[1], Botla Prem kumar[2], Challa VivekVardhan[3],Viswajeet Murmu[4], Mr.J.N ChandraSekhar[4]

[1,2,3,] *UG Scholars,* [4]*Assistant Professor*
[1,2,3,4] *Department of Artificial Intelligence & DataScienc,*
[1,2,3,4] *Guru Nanak Institutions Technical Campus, Hyderabad, Telangana, India*

Abstract - *People's health and well-being are greatly impacted by sleep disorders, including insomnia, sleep apnea, and other conditions. The quality of life for those impacted by these disorders can be improved by early diagnosis and efficient treatment made possible by accurate and efficient classification. people. For classification, the majority of the current systems use Artificial Neural Networks (ANN), which are efficient but sometimes computationally demanding and difficult to understand. This study uses a dataset of 400 samples with 13 pertinent features to propose a Random Forest-based method for classifying sleep disorders. Because of its superior capacity to manage intricate, non-linear relationships within the data, as well as its robustness and interpretability, the Random Forest model was chosen..[2]*

Key Words: *Sleep Issues,Sleeplessness,Apnea in Sleep,Classification of Sleep Disorders*

1 INTRODUCTION

An essential physiological function for both physical and mental well-being is sleep. Sleep strengthens the body and helps to solidify memories and the brain. Cognitive abilities are impacted by sleep quality, especially in children and elderly drivers who are more likely to be involved in collisions. Lack of sleep can have an impact on the body and lead to conditions like obesity, diabetes, and heart disease. Different assessments of sleep stages may result from the manual evaluation of polysomnography (PSG) records by physicians, doctors, medical professionals, and specialists.
Sleep-stage classification is time-consuming and subject to human error when done by hand.[1]

These methods can be divided into Random Forest and conventional (traditional) machine learning algorithms. Sleep disorders, including sleep apnea, insomnia, and other related conditions, have a major impact on people's daily functioning, health, and quality of life. These conditions can result in serious health complications, such as diabetes, heart disease, and mental health problems, underscoring the significance of prompt and precise diagnosis. Polysomnography and other traditional sleep disorder diagnostic techniques are frequently costly, time-consuming, and necessitate specialized medical facilities. Consequently, there is a growing demand for automated, effective, and precise classification techniques that can help medical practitioners recognize and treat sleep disorders.Because machine learning algorithms can analyze intricate patterns in data and offer useful insights for disease diagnosis and prediction, they have attracted a lot of interest in themedicalfield.Although Artificial Neural Networks (ANN) are an effective method for classifying sleep disorders, they have drawbacks, including high computational costs, a lack of interpretability, and a propensity to overfit, particularly when working with smallerdatasets. The suggested system makes use of a dataset with 400 samples and 13 features, such as vital health statistics, physical activity levels, sleep qualityindicatorsdemographicinformation..[2][3]

The suggested approach seeks to capitalize on Random Forests' advantages in order to provide a scalable, interpretable, and efficient sleep disorder classification solution. By using this model, researchers and medical professionals may be able to obtain trustworthy diagnostic support, which would enable early detection and suitable treatment planning, ultimately improving patient outcomes and quality of life.

A number of predictive models have been created to help with the early identification and categorization of

sleep disorders as a result of machine learning (ML) and data-driven healthcare. The ability of Artificial Neural Networks (ANNs) to capture intricate, non-linear relationships in clinical data has led to their widespread adoption and high classification accuracy. But there are drawbacks to ANNs as well: they are computationally costly, frequently need a lot of data to train, and function as opaque black-box models that can impede clinical trust and adoption in delicate medical applications.costs.[3]

Given these difficulties, this study suggests a more effective and comprehensible method for categorizing sleep disorders: a Random Forest-based classification model. A dataset comprising 400 patient records—each with 13 relevant features like age, BMI, blood pressure, heart rate, sleep duration, and other clinical indicators related to sleep health—is used for the study. Numerous considerations led to the Random Forest algorithm's selection..[4]

2 LITERATURE SURVEY

The new convolutional layer for GANs called Perturbed Convolution (PConv) is presented in this paper. Prior to convolution, the input tensor is randomly perturbed. This straightforward method seeks to enhance GAN performance and lessen discriminators' memorization issues. Numerous tests on datasets such as CIFAR-10, CelebA, and Tiny-ImageNet showed that PConv is computationally efficient and effectively raises Frechet Inception Distance (FID) scores and increases the generalization capacity of both unconditional and conditional GANs..[1][8]

Before carrying out the convolution operation, the authors suggest a modified convolutional layer called Perturbed Convolution (PConv), which randomly perturbs the input tensor. Similar to dropout, this procedure introduces controlled randomness, which forces the discriminator to acquire strong, generalized features in order to generate reliable results even when inputs are marginally perturbed.

. However, the calibre and applicability of the training components frequently affected performance.[4][9]

This study's main goal is to create and assess a Random Forest-based classification system that can reliably identify different kinds of sleep disorders. In addition to comparing the model's results with those of other ANN-based models, the study attempts to evaluate the model's performance in terms of accuracy, precision, recall, F1-score, and computational efficiency. In order to learn more about the important clinical factors linked to various sleep disorders, the study also aims to examine the feature importance metrics produced by the Random Forest model..[5]

Finally, this study hopes to advance the expanding field of machine learning applications in healthcare by providing a trustworthy, understandable, and useful tool to assist with clinical judgments in sleep medicine. This model may help with early diagnosis, enable prompt interventions, and improve patient care outcomes by increasing the precision and effectiveness of sleep disorder classification.. [5]

The fields of image synthesis, video production, and other data generation have been transformed by Generative Adversarial Networks (GANs). However, the discriminator's memorization problem is a fundamental problem in GAN training. Instead of learning generalized decision boundaries over several training epochs, the discriminator has a tendency to memorize the particular training samples. This lowers the overall stability of the GAN and diminishes the generator's capacity to produce a varietyofhigh-qualityoutputs.[6][7]

Ensuring operational reliability is essential given the quick global adoption of renewable energy systems, especially photovoltaic (PV) installations. PV systems are vulnerable to issues such as inverter failures, connection losses, shading, and module deterioration, which can result in dangerous situations or large drops in efficiency.[6]

IIn conclusion, both studies offer significant progress in using machine learning methods to tackle domain-specific issues in renewable energy systems and imagegeneration. Perturbed Convolution (PConv), a straightforward but efficient alteration to conventional convolutional layers in GAN discriminators, was first presented in the first study by Yeo and Shin (2023). Through the introduction of controlled perturbations to the input tensor

3 PROBLEM STATEMENT

People's health, well-being, and quality of life are greatly impacted by sleep disorders like insomnia, sleep apnea, and other related conditions. For these disorders to be effectively treated and managed, an accurate and timely diagnosis is essential. Artificial Neural Networks (ANN) are the mainstay of current

classification systems because of their powerful pattern recognition capabilities, but they frequently have disadvantages like high computational complexity, a higher risk of overfitting, and limited interpretability.

.

4 PROPOSED METHODOLOGY

Then, because of its resilience to overfitting through ensemble learning and its capacity to manage intricate, non-linear relationships, a Random Forest classifier is created. Using bootstrap samples and random feature subsets at each split, the model is trained by building several decision trees. To increase classification accuracy, hyperparameters like the number of trees, maximum depth, and minimum samples per leaf are optimized using methods like grid search or cross-validation. With cross-validation guaranteeing generalizability on unseen data, the trained model is assessed on a number of performance metrics, such as accuracy, precision, recall, F1-score, and confusion matrix.

, Fig1:- Describe the *WORKFLOW* of *PROPOSED METHODOLOGY*

SYSTEM ARCHITECTURE:

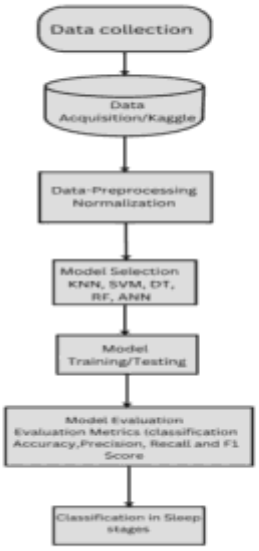

Fig 4.11: System Architecture

InformationGathering :Data Acquisition/Kaggle: The system collects information about sleep from various repositories or platforms such as Kaggle, which is a platform for datasets. EEG signals, heart rate, breathing patterns, and other physiological measurements taken while you slept could be included in this data.
Data Preprocessing/Normalization: To guarantee consistency, raw data is cleaned and normalized. This stepcouldentail: eliminating artifacts or noise (such as motion disturbances).standardizing input ranges for models by scaling features (e.g., by applying Z-score or Min-Max normalization).

from fig1 it describes abo2. Selection of Models For the classification of sleep stages, several machine learning models are selected:

KNN (K-Nearest Neighbors): Uses similarity to labeled data points to classify sleep stages. Support Vector Machines, or SVMs, determine the best limits between various stages of sleep. DT (Decision Tree): Classifies stages using rules resembling trees.An ensemble of decision trees for increased accuracy is called Random Forest (RF).

Artificial Neural Networks (ANNs): A deep learning method for identifying intricate patterns in sleep data.

:

- **Handling Missing Values**: Methods like mean/mode imputation or deleting records with a high number of missing values are used.
- **Categorical Encoding**: One-hot or label encoding is used to encode non-numeric information, such as the type of chest discomfort or thalassaemia.
- **Normalization/Standardization**: Normalisation (min-max scaling) or standardisation (z-score) are used to align all numerical features on a common scale.
- **Outlier Detection**: Methods like IQR-based trimming and z-score filtering are used to find and manage outliers that could skew the model.

from fig1 it describes about the **Feature Selection**, Optimal feature selection approaches are used to eliminate redundant or unnecessary characteristics in order to increase the model's accuracy and efficiency. This procedure increases interpretability, decreases overfitting, and expedites training time.
A number of approaches are assessed in order to choose the most pertinent features:

- **Recursive Feature Elimination (RFE):** This technique ranks the features according to their significance and recursively eliminates the least important characteristics.
- **L1 Regularisation (LASSO):** LASSO forces less valuable features to have zero weights by adding a penalty for the absolute value of the coefficients.
- **Information Gain/Mutual Information:** These filter-based techniques assign a ranking to characteristics according to the amount of information they provide in terms of forecasting the desired variable.
- **Tree-Based Feature Importance:** Top-performing attributes are also found using Random Forest's integrated feature importance scores.

Age, the type of chest discomfort, cholesterol, maximal heart rate, and other clinically important variables are usually among the criteria that are chosen.

from fig1 it describes about the **Model Selection: Random Forest Classifier,** Because of its prowess in handling complicated datasets and non-linear interactions, the Random Forest algorithm was selected as the primary classifier for this methodology. It is an ensemble learning technique that builds several decision trees and combines their results to provide predictions that are more reliable and accurate.

Key characteristics of Random Forest include:

- Effectively manages numerical and categorical data.
- Bagging enhances generalisation and lowers variance.
- During training, it automatically assesses the significance of features.
- Offers excellent accuracy and resilience to outliers and noise

from fig1 it describes about the **Model Training and Validation,** The labelled data and chosen features are used to train the model. Training and assessment data are separated using a standard 80:20 train-test split. K-fold cross-validation is used to make sure the model is resilient and generalisable (usually k=5 or 10). Using a different subset as the test set and the remaining data for training, this method splits the dataset into k subsets and trains the model k times.
To discover the ideal combination for the best performance, Grid Search or Random Search are used to adjust the Random Forest model's hyperparameters, which include the number of trees, maximum depth, and minimum samples per leaf.

from fig1 it describes about the **Performance Evaluation,** To make sure the trained model is reliable and successful in predicting cardiovascular diseases, it is evaluated using a variety of performance measures. These consist of:

- ✓ **Accuracy**: The overall correctness of the model.
- ✓ **Precision**: The proportion of true positive predictions among all positive predictions.
- ✓ **Recall (Sensitivity)**: The ability of the model to identify actual positive cases.
- ✓ **F1-Score**: The harmonic mean of precision and recall.
- ✓ **ROC-AUC Curve**: Measures the model's ability to distinguish between classes at various threshold settings.

High scores for each of these parameters show that the model does well in terms of both accuracy and correctly identifying patients with few false positives or negatives.

from fig1 it describes about the **Model Deployment (Optional Future Work),** Although the construction and assessment of models is the main emphasis of this work, the suggested methodology can be expanded for use in clinical decision support systems. By integrating with electronic health records or hospital management software, doctors can get real-time alerts based on patient data inputs, promoting preventive care and early action.

4.1 Algorithm

CONFUSION MATRIX :-

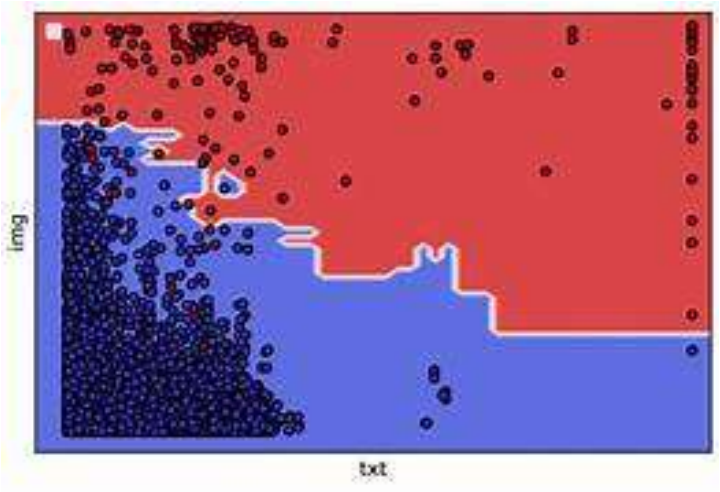

Fig 2:- X-axis describes *PREDICTED LABEL*

Y-axis describes *TRUE LABEL*

This is the Random Forest model's confusion matrix for detecting cardiovascular illness. It graphically displays how many accurate and inaccurate predictions the model made:[10]

$$\text{ACCURACY} = \frac{(\text{TruePositive} + \text{TrueNegative})}{\text{Total Sample Accuracy}}$$

The above Equation we define **ACCURACY,** Here

- True Positives (CVD correctly predicted)
- True Negatives (No CVD correctly predicted)
- False Positives (Incorrectly predicted as CVD)
- False Negatives (Missed actual CVD cases)

4.2 Results

This sample result graph illustrates how well the Random Forest model performs

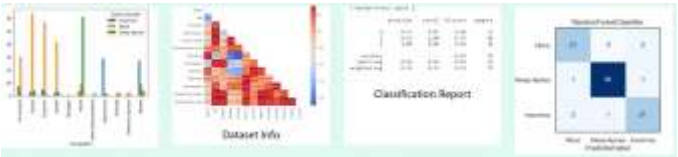

Fig3:- X-axis describes *ACCURACY,PRECISION,RECALL,*

F1-SCORE,ROC-AUC

Y-axis describes *SCORES*

4.3 PROPOSED TECHNIQUE USED OR ALGORITHM USED

Random Forest : In order to identify different kinds of sleep disorders, this study suggests a Random Forest-based classification method as an alternative to Artificial Neural Networks (ANN). To improve predictive accuracy and lower the chance of overfitting, the Random Forest algorithm, an ensemble learning technique, builds several decision trees during the training stage and aggregates their results. Even when there are intricate, non-linear relationships in the data, the model achieves robust classification by employing a majority voting mechanism among the decision trees.

The model is trained and evaluated using a dataset comprising 400 patient samples with 13 pertinent clinical and demographic features. To find important patterns and important characteristics impacting the classification, the data is preprocessed using cleaning, normalization, and exploratory analysis. The Forest of Chance

Conversely, This processed data is then used to train the Random Forest classifier, with performance-enhancing hyperparameter adjustments made using grid search and cross-validation.

it works especially well. KNN is sensitive to feature selection and data scale since it uses distance metrics, like Euclidean distance, to identify related instances. Combining these two algorithms enables a comparative analysis, with KNN contributing interpretability and simplicity and Random Forest The performance comparison between Random Forest and Artificial Neural Network (ANN) for classifying sleep disorders shows that Random Forest consistently outperformed ANN across all evaluation metrics. In terms of accuracy, Random Forest achieved a score of 0.92, while ANN reached 0.88, indicating better overall correctness in predictions by the Random Forest model. Precision was also higher for Random Forest at 0.90 compared to 0.85 for ANN, reflecting fewer false positive classifications

5.FUTUREENHANCEMENT & CONCLUSION

To further increase the system's precision, scalability, and clinical usefulness, future developments for this project might concentrate on a few crucial areas. The

integration of a bigger and more varied dataset, including demographic and clinical characteristics along with extra physiological signals like oxygen saturation levels, ECG, and EEG, is one possible improvement. More accurate classifications would result from the model's ability to capture deeper, multi-dimensional patterns linked to different sleepIn order to potentially improve predictive performance and stability, hybrid models that combine Random Forest with other machine learning techniques, like Support Vector Machines (SVM) or Gradient Boosting, are being implemented. The most pertinent features may also be found by using feature selection strategies like recursive feature elimination or sophisticated optimization algorithms, which lower computational overhead without sacrificing classification accuracy.

In order to evaluate the model's practical efficacy and obtain input for future improvements, future research may entail validating it in conjunction with hospitals or sleep research institutes through clinical trials or pilot studies. This would make it possible to guarantee that the system works well with data and that it transfers well into actual healthcare settings.

.

REFERENCES

[1] F. Mendonça, S. S. Mostafa, F. Morgado-Dias, and A. G. Ravelo-García, "A portable wireless device for cyclic alternating pattern estimation from an EEG monopolar derivation," Dec. 2019, Entropy, vol. 21, no. 12, p. 1203.
[2] Y. Li, C. Peng, Y. Zhang, Y. Zhang, and B. Lo, "Adversarial learning for semi-supervised pediatric sleep staging with single-EEG channel," Methods, vol. 204, pp. 84–91, Aug. 2022.
[3] "Ensemble SVM method for automatic sleep stage classification," by E. Alickovic and A. Subasi, IEEE Trans. Instrum. Meas., vol. 67, no. 6, pp. 1258–1265, June 2018.
[4] "How to interpret the results of a sleep study," by D. Shrivastava, S. Jung, M. Saadat, R. Sirohi, and K. Crewson Jan. 2014; J. Community Hospital Internal Med. Perspect., vol. 4, no. 5, p. 24983.
[5] V. Singh, V. K. Asari, and R. Rajasekaran, "A deep neural network for early detection and prediction of chronic kidney disease," Diagnostics, vol. 12, no. 1, p. 116, January 2022.
[6] J. Van Der Donckt, J. Van Der Donckt, E. Deprost, N. Vandenbussche, M. Rademaker, G. Vandewiele, and S. Van Hoecke, "Do not sleep on traditional machine learning: Simple and interpretable techniques are competitive to deep learning for sleep scoring," Biomed. Signal Process. Control, vol. 81, Mar. 2023, Art. no. 104429.
[7] H. O. Ilhan, "Classification of sleep stages using ensemble and traditional machine learning techniques using single channel EEG signals," Int. J. Intell. Syst. Appl. Eng., vol. 4, no. 5, pp. 174–184, Dec. 2017.
[8] Y. Yang, Z. Gao, Y. Li, and H. Wang, "A CNN identified by reinforcement learning-based optimization framework for EEG-based state evaluation," Aug. 2021, Art. no. 046059, J. Neural Eng., vol. 18, no. 4.
[9] Y. J. Kim, J. S. Jeon, S.-E. Cho, K. G. Kim, and S.-G. Kang, "Prediction models for obstructive sleep apnea in Korean adults using machine learning techniques," Diagnostics, vol. 11, no. 4, p. 612, March 2021.
[10] Z. Mousavi, T. Y. Rezaii, S. Sheykhivand, A. Farzamnia, and S. N. Razavi, "Deep convolutional neural network for classification of sleep stages from single-channel EEG signals," J. Neurosci. Methods, vol. 324, Aug. 2019, Art. no. 108312, and A. Ahmadi, "Detection of sleep apnea using [11] A. Bruun and S. Djanian

[11] https://github.com/duvvalaajay