Sleep Disorder Detection Using Deep Learning and Genetic Algorithm Optimization

1st **Suresh Munnangi**
Dept. of CSE
Narasaraopeta Engineering College
Narasaraopeta, India
sureshmunnangi55@gmail.com

2nd **Sarath Chandra Beesabattuni**
Dept. of CSE
Narasaraopeta Engineering College
Narasaraopeta, India
beesabattunisarathchandra@gmail.com

3rd **Sai Manikanta Konakanchi**
Dept. of CSE
Narasaraopeta Engineering College
Narasaraopeta, India
saimanikanta723@gmail.com

4th **Sairam Konakanchi**
Dept. of CSE
Narasaraopeta Engineering College
Narasaraopeta, India
konakanchisairam029@gmail.com

5th **Charles Babu Gundepalli**
Dept. of CSBS
GRIET, Bachupally
Hyderabad, Telangana, India
charles1624@grietcollege.com

6th **Meeniga Vijaya Lakshmi**
Dept. of ECE
GNITS (for Women)
Shaikpet, Hyderabad, India
m.vijayalakshmi@gnits.ac.in

7th **K.V. Narasimha Reddy**
Dept. of CSE
Narasaraopeta Engineering College
Narasaraopeta, India
narasimhareddynec03@gmail.com

***Abstract*—Understanding and categorizing sleep problems is crucial for enhancing general health and quality of life since conditions like sleep apnea and insomnia can significantly impair day-to-day functioning. Because medical professionals' traditional diagnosis methods can be inconsistent and time-consuming, automated solutions are a desirable substitute. In this study, the classification of sleep disorders using deep learning and traditional machine learning approaches is compared. The Sleep Health and Lifestyle Dataset, 400 records and 13 characteristics defining different lifestyle, health, and sleep aspects, is made freely accessible for use in this study. Encoding categorical data, transforming blood pressure measurements into numerical form, normalizing feature values, and dividing the data into training and testing sets were all steps in the preprocessing procedure. A Random Forest Classifier, an XGBoost Classifier, and a Keras-based Artificial Neural Network (ANN) with Dense and Dropout layers are among the models that were put to the test. Confusion matrices, classification reports, and accuracy scores were used to gauge the model's performance. The ANN outperformed the other models and had the greatest classification accuracy among the evaluated methods; Random Forest and XGBoost also showed promising results. These results demonstrate the usefulness of deep learning architectures in developing precise, scalable systems for the early identification of sleep problems, assisting medical professionals in establishing diagnoses.**

***Index Terms*—Sleep disorder detection, Artificial Neural Network, Random Forest model, XGBoost algorithm, healthcare data analysis, feature encoding and scaling.**

I. INTRODUCTION

Sleep problems like not being able to sleep or having trouble breathing at night can really mess up a person's life. They make it hard to think, get work done, and stay healthy [1]. To treat them well, doctors need to know what's wrong quickly and correctly. But the old way of doing things—where a doctor looks at a patient's sleep patterns by hand—is slow and can easily have mistakes [2]. That's where machine learning comes in.Smart computer programs can play an important role in detecting sleep disorders. They are able to analyze patient data automatically, which reduces the workload for doctors and improves accuracy [3]. Because these systems can work in real time, doctors can monitor patients and provide timely support, whether the patient is at home or in a clinic [3].

These programs learn from data collected during sleep studies, such as brain signals or records of daily activities [4], [5]. Integrating such technology into devices like smartwatches or mobile health applications could make sleep disorder detection more accessible and less costly, removing the need for specialized sleep laboratories [6].

In one study, Alshammari used a publicly available dataset [7], [8] to compare several machine learning models, including Artificial Neural Networks (ANN) [9], Random Forests (RF) [10], and others [11]–[13]. To improve their performance, the researchers applied a Genetic Algorithm (GA) [14], which fine-tuned the model parameters. The ANN model achieved the highest accuracy of 92.92

The research process included key steps: (1) preparing the dataset by splitting it into training and testing groups, (2) converting text-based information into numerical form for analysis [3], and (3) using the GA to optimize model performance in terms of accuracy, recall, and F1-score [15].

The results demonstrated that optimization methods such as GA can significantly improve the accuracy of machine learning models when applied to healthcare classification tasks. Despite these contributions, the study also reported certain limitations [1]. Only 400 samples from a single source make up the dataset, which would restrict how broadly the findings can be applied. Furthermore, even though the GA-optimized ANN obtained great accuracy, the study was more concerned with model comparison than with implementation in embedded or real-time healthcare systems. This provides room for more study on model interpretability, scalability, and integration with practical diagnostic processes.

In the present work, we build upon the foundations laid by prior studies and propose a streamlined yet effective classification pipeline. we implemented preprocessing techniques including label encoding, standardization, feature extraction, and dataset partitioning. We employed three models—ANN with Dense and Dropout layers [1], Random Forest Classifier [2], and XGBoost Classifier [3]—evaluating them using accuracy, classification reports, and confusion matrices.

The results from our experiments demonstrate that the ANN model continues to deliver superior performance compared to the other tested algorithms, while Random Forest and XGBoost also produce competitive results. By focusing on model efficiency and adaptability.

Section 2 provides a comprehensive review of related work in sleep disorder detection using machine learning and deep learning approaches. Section 3 details the proposed methodology, including data preprocessing, feature engineering, and model architectures. Section 4 presents the experimental setup, datasets, evaluation metrics, and performance outcomes. Section 5 analyzes the experimental findings, highlighting key insights and practical implications. Section 6 concludes by summarizing the contributions and suggesting possible paths of inquiry for further research.

II. Literature Review

Zan and Yildiz [4] proposed *FullSleepNet*, a fully convolutional multi-task CNN for both arousal detection and sleep stage classification from single-channel EEG. The architecture integrates convolutional layers for feature extraction, recurrent modules to capture temporal dependencies, attention mechanisms for improved focus, and segmentation layers for final predictions. The model achieved an AUPRC of 0.70 for arousal detection and classification accuracies of 0.88 and 0.83 on two benchmark datasets, demonstrating robustness and adaptability for clinical deployment.

Mohammadi and Mohammadi [5] introduced *SleepLiteCNN*, apnea subtype classification using single-lead ECG data at 1-second resolution. The approach targets embedded systems and wearable devices, employing model quantization and FPGA deployment to minimize power consumption. The proposed model achieved more than 95% accuracy with a macro-F1 score of 92% while consuming only 1.8 μJ per inference.

Papillon et al. [6] investigated sleep position classification using pressure-sensitive mats placed beneath the mattress, enabling a non-intrusive monitoring approach. with pre-trained Vision Transformer models (ViT-MAE and ViTPose). This strategy outperformed conventional methods such as Temporal Convolutional Networks (TCNs) and classical machine learning classifiers (SVM, RF, XGBoost), offering potential for integration into smart-home healthcare systems.

Kazemi et al. [7] developed a multitask explainable 1D Vision Transformer for simultaneous sleep stage and apnea detection using multimodal physiological signals, including photoplethysmography (PPG), respiratory flow, and respiratory effort. The model achieved an accuracy of 78% (κ=0.66) for sleep staging and 74% (κ=0.58) for apnea detection. Additionally, the attention weights provided interpretability, revealing the relative importance of each input modality in the decision-making process.

Monowar et al. [8] proposed a multi-layered ensemble learning approach that integrates thresholding, predictive scoring, and Softmax-to-vector conversion, followed by voting and stacking methods. To address dataset imbalance. The ensemble achieved 96.88% accuracy on balanced data and 99.5% under 8-fold cross-validation, outperforming individual models and showing strong potential for large-scale clinical applications.

III. Proposed Methodology

A. Experimental Setup

The implementation and evaluation were conducted in a Python-based development environment, utilizing both a local system and Google Colab for GPU-accelerated processing. The software configuration included Python 3.10.OpenCV for image handling, for statistical analysis and metric evaluation.

- **CPU**: i5-12450H (12th Gen)
- **Ram**: 8gb
- **gpu**: NVIDIA Tesla T4 (16 GB VRAM, accessed via Google Colab)
- **Platform**: Windows 11, 64-bit, x64 architecture

B. Datasets

Sleep Health and Lifestyle Dataset [9], which is openly accessible on the Kaggle platform, is the dataset utilized in this work. It includes 400 records and 13 variables that reflect lifestyle factors, clinical measures related to sleep problems and quality, and demographic information. "Gender," "Age," "Occupation," "Sleep Duration," "Quality of Sleep," "Physical Activity Level," "Stress Level," "BMI Category," "Blood Pressure," "Heart Rate," and "Daily Steps" are some of the key elements. There are three classifications for the target variable *Sleep Disorder*: *None*, *Sleep Apnea*, and *Insomnia*.

Limitation: It is important to note that the dataset size of 400 records is relatively limited, which may affect the generalizability of our findings and model performance on broader populations. This constraint should be considered when interpreting the results.

Blood pressure data were divided into systolic and diastolic components, numerical characteristics were normalized to provide equal scaling across models, and categorical variables were label-encoded for preprocessing.

C. Preprocessing

To get the Sleep Health and Lifestyle Dataset ready for training and assessing machine learning models, data preprocessing was done. *Label Encoding* [10] was used to convert categorical variables such *Gender*, *Occupation*, *Sleep Disorder*, and *BMI Category* into numerical form. By giving each category a distinct integer, this step makes it possible for algorithms to efficiently process these characteristics. The researchers got their data ready for the computer models. First, they handled the Sleep Disorder information, which was what the models were trying to predict. Because computers only work with numbers, the researchers changed the word labels for sleep disorders, such as "None," "Sleep Apnea," and "Insomnia," into simple numerical codes like 0, 1, and 2 [3]. They also had to deal with the Blood Pressure column. Since it had two numbers combined, like "120/80," they split it into two new columns: one for Systolic Blood Pressure and another for Diastolic Blood Pressure. After splitting them, they turned these new values into numbers that the models could use. To keep the dataset clean and avoid repeated information, the original "Blood Pressure" column was then deleted [3]. After that, they standardized the numerical data using the StandardScaler method. This process rescaled the values so that all features had a mean of zero and a standard deviation of one. This step is extremely important for models that are affected by the size of the numbers, such as the neural network they used [3]. Finally, they split the dataset into two parts using a train-test split technique. Seventy percent of the data was used for the model to learn from (the training set), and the remaining 30A trustworthy indicator of generalization performance is provided by this separation, which guarantees that model assessment is done on unseen data.

D. Model Architectures and Functional Roles

1) Artificial Neural Network (ANN): An Artificial Neural Network (ANN) is a learning model inspired by how the human brain processes information. It is made up of layers of interconnected nodes, often called neurons, that transform input data into meaningful outputs [1]. In this study, the ANN was built using Keras *Sequential* API, with multiple fully connected (*Dense*) layers and *Dropout* layers to reduce overfitting. The basic operation of a neuron is to take a weighted sum of its inputs and apply an activation function:

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \tag{1}$$

The first one, the Artificial Neural Network, is kind of like a learning brain. It takes in a bunch of information, processes it, and then learns from its mistakes to get better and better at guessing the right answer [9].

2) Random Forest Classifier: The other one, the Random Forest, is more like a team. It's actually a group of a bunch of smaller computer programs. Each one gets a random piece of the information. When it's time to make a final guess, they all "vote" on what they think the answer is. The final result is just whatever most of the team members agreed on [10]. is trained on a random sample of the data, and at each split, a subset of features is chosen to reduce similarity between trees. The final prediction is obtained through majority voting:

$$\hat{y} = \text{mode}\left\{h_1(x), h_2(x), \ldots, h_T(x)\right\} \tag{2}$$

where $h_t(x)$ is the prediction of the t-th tree and T is the number of trees. This method is effective in handling both numerical and categorical features, and its ensemble nature makes it less prone to overfitting.

3) Extreme Gradient Boosting (XGBoost): XGBoost is an advanced gradient boosting method that builds trees sequentially, with each new tree focusing on correcting the mistakes of the previous ones [3]. Its objective function combines both the loss term and a regularization term to control complexity:

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{3}$$

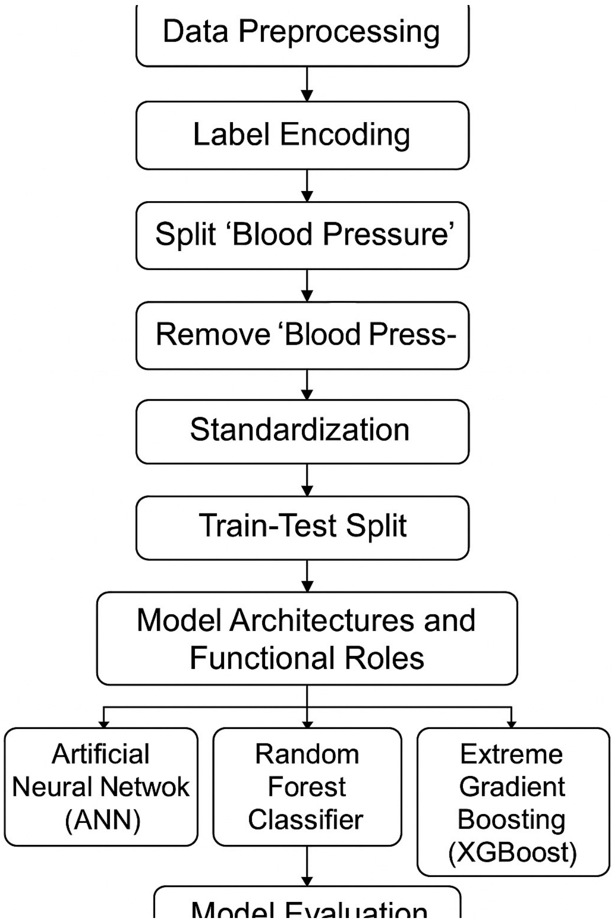

Fig. 1: Workflow for Sleep Health and Lifestyle classification from preprocessing to model evaluation.

with

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \|w\|^2 \tag{4}$$

where T is the number of leaves, w the leaf weights, and γ, λ are regularization parameters. XGBoost is widely used because it can handle missing values, supports parallel training, and delivers strong performance on structured datasets.

The Sleep Health and Lifestyle dataset undergoes thorough preparation before the workflow in Fig. 1 starts. To guarantee algorithm compatibility and repeatability, categorical attributes—*Gender*, *Occupation*, *Sleep Disorder*, and *BMI Category*—are mapped to integers using label encoding. Fine-grained analysis of cardiovascular signals is made possible by the breakdown of the composite *Blood Pressure* field into its physiological components, *Systolic* and *Diastolic*. *StandardScaler* is used to normalize all numerical characteristics in order to stabilize optimization and prevent high-variance qualities from dominating. Lastly, a 70/30 stratified train-test split maintains class proportions while offering an objective generalization estimate.

Three complementary approaches are used in the modeling process: a Random Forest (RF) to take advantage of ensemble bagging and feature sub-sampling, a Keras-based Artificial Neural Network (ANN) for non-linear feature interactions, and XGBoost to capture residual patterns using gradient-boosted trees with regularization. The pipeline is able to balance interpretability, resilience to noise, and expressive power because of the variety of inductive biases. The last assessment step, where metrics calculated on the held-out test set quantify comparative performance and direct the selection of the most dependable model for sleep-disorder screening, is also highlighted in the diagram.

E. *Performance Evaluation Metrics*

Evaluating the performance of classification models requires more than just looking at how many predictions are correct. In this work, three common evaluation techniques were used: *accuracy score*, *classification report*, and *confusion matrix*. Together, they provide both an overall picture and class-level details of how well the models perform.

1) Accuracy Score: Accuracy represents the ratio of correctly predicted cases to the total number of cases [9]. It can be expressed as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

where TP stands for true positives, TN for true negatives, FP for false positives, and FN for false negatives. Although accuracy is simple and easy to interpret, it may not always give a fair view when the dataset is imbalanced, as it could favor the majority class.

2) Classification Report: The classification report provides a more detailed view by reporting precision, recall, and the F1-score for each class [9]. Precision reflects how many of the predicted positives are actually correct:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6}$$

Recall (or sensitivity) measures how many of the actual positives are correctly identified:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7}$$

The F1-score combines these two aspects into a single value by taking their harmonic mean:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

This balance makes the F1-score especially useful when classes are unevenly distributed or when both false positives and false negatives carry significant consequences.

F. Experimental Results

To strengthen the reliability of our accuracy estimates given the limited dataset size, we employed 5-fold cross-validation in addition to the standard train-test split. The accuracy scores obtained for each model are summarized in Table I. Among the three classifiers, the ANN achieved the highest performance, indicating its ability to capture complex patterns in lifestyle and physiological attributes. Random Forest and XGBoost also produced competitive results, showing that ensemble approaches remain effective for structured health datasets.

TABLE I: Accuracy of Implemented Models (5-Fold Cross-Validation)

Model	Test Accuracy (%)	Cross-Val Accuracy (%)
Artificial Neural Network (ANN)	94.17	92.85 ± 1.23
Random Forest Classifier	91.67	90.42 ± 1.56
XGBoost Classifier	92.50	91.18 ± 1.34

In addition to accuracy, the classification reports revealed high precision, recall, and F1-scores across all three classes (*None*, *Sleep Apnea*, *Insomnia*) for the ANN, indicating balanced performance without significant bias toward any particular class. Confusion matrix analysis showed that most misclassifications occurred between *Sleep Apnea* and *Insomnia*, which may be due to overlapping symptoms reflected in the dataset features.

G. Discussion

The results suggest that deep learning, particularly ANN, provides a slight edge over traditional ensemble methods for sleep disorder prediction. The ANN's superior accuracy can be attributed to its ability to model non-linear feature interactions, especially between lifestyle indicators and physiological measures such as blood pressure. However, Random Forest and XGBoost achieved strong and stable performance with lower training complexity, making them appealing options in scenarios where interpretability and computational efficiency are more critical.

The cross-validation results shown in Table I demonstrate consistent performance across different data splits, with the ANN maintaining its superior performance. The small standard deviations in cross-validation accuracy indicate stable model performance despite the limited dataset size.

The three computer models all worked well, with none of them showing a bias toward one specific sleep problem. Looking at the results more closely, the models mostly had a hard time telling the difference between Sleep Apnea and Insomnia. This makes sense, as these two disorders often have very similar symptoms in real life. This shows how difficult it is to separate them just by using information about a person's lifestyle and what they report about their own health.

Overall, the study proved that combining different types of computer models is a good idea. While the ANN (Artificial Neural Network) was the most accurate, other models like Random Forest and XGBoost are also very good. They are still reliable choices that are easier to understand and work with, depending on what a doctor or researcher needs.

Limitation Acknowledgment: While our models achieved promising results, the limited dataset size of 400 records remains a constraint. The performance reported should be interpreted with caution, as larger and more diverse datasets would be needed to fully validate the generalizability of these models across different populations and clinical settings.

IV. Closing Remarks and Future Perspectives

A recent study investigated how to better identify sleep disorders using machine learning models by utilizing lifestyle and sleep data. The researchers meticulously prepared the dataset by standardizing numerical values, dividing it into training and testing sets, and transforming categorical data and measurements like blood pressure into numerical form in order to guarantee accurate results. Three models—an Artificial Neural Network (ANN), XGBoost, and Random Forest—were assessed. With an accuracy of 94.17, the ANN outperformed the others, closely followed by Random Forest (91.67) and XGBoost (92.50). The ability of the ANN to identify intricate patterns without overfitting, which facilitated better generalization to unknown data, contributed to its exceptional performance. Random Forest and XGBoost, on the other hand, also showed solid and consistent outcomes, which makes them viable choices in situations with constrained computational resources.

To address the limitation of dataset size, future work should focus on collecting larger and more diverse datasets. Additionally, techniques such as bootstrapping and advanced cross-validation strategies could be employed to provide more robust performance estimates. Data augmentation methods specific to healthcare data may also help mitigate the challenges posed by limited sample sizes.

In order to further increase prediction accuracy, the researchers intend to add more physiological signals, such as electroencephalogram (EEG) and polysomnography (PSG) data, and broaden the dataset to include a larger and more varied population in subsequent work. In order to give doctors useful information, the incorporation of model interpretability approaches like SHAP or LIME will be investigated. Additionally, the process will be modified for use in real-time healthcare systems, such as wearable technology and telemedicine platforms, allowing for early diagnosis of sleep problems and ongoing monitoring in real-world clinical settings.

REFERENCES

[1] Y. You *et al.*, "Automatic sleep stage classification: A light and efficient deep neural network model based on time, frequency and fractional fourier transform domain features," *Artificial Intelligence in Medicine*, vol. 127, p. 102279, 2022.

[2] P. Tripathi *et al.*, "Ensemble computational intelligence for insomnia sleep stage detection via the sleep ecg signal," *IEEE Access*, vol. 10, pp. 108 710–108 721, 2022.

[3] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.

[4] Y. Zan and O. Yildiz, "Fullsleepnet: A fully convolutional multi-task network for arousal detection and sleep stage classification," *arXiv preprint arXiv:2406.01834*, 2024.

[5] M. Mohammadi and H. Mohammadi, "Sleeplitecnn: Energy-efficient sleep apnea subtype classification using single-lead ecg," *arXiv preprint arXiv:2508.02718*, 2025.

[6] J. Papillon *et al.*, "Sleep position classification using transfer learning with vision transformers," *arXiv preprint arXiv:2505.08111*, 2025.

[7] S. Kazemi *et al.*, "Multimodal vision transformer for sleep stage and apnea detection," *arXiv preprint arXiv:2502.17486*, 2025.

[8] M. Monowar *et al.*, "Multi-layered ensemble learning with data balancing for sleep disorder detection," *Frontiers in Artificial Intelligence*, vol. 7, p. 1506770, 2024.

[9] UOM190346A, "Sleep health and lifestyle dataset," https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset, 2023.

[10] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.