IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

USE OF MACHINE LEARNING FOR IMPROVING SLEEP DISORDER CLASSIFICATION

[1]Lokesh Pramod Meshram, [2]Varad Krishna Marawar, [3]Hanuman Gajanan Limbalkar, [4]Kaif Mulla

Department Of Computer Engineering PCET'S PCCOE Pune

PCCOE PUNE, Pune, India

Abstract: The categorization of sleep disorders has proven to be of main importance in the advancement of human quality of life, due to the fact that sleep apnea and other sleep disorders affect the general health of a person. Sleep stage analysis is inherently difficult and vulnerable to human errors, meaning that great need lies in the development of strong machine learning algorithms for an effective classification of sleep disorders. The paper discusses a comparative study of conventional machine learning algorithms with for classifying sleep disorders using the publicly available Sleep Health and Lifestyle Dataset that contains 400 rows and 13 columns representing various features related to sleep and daily activities. Several parameters of different MLAs are optimized in the paper using a genetic algorithm to improve classification performance. The performance of the k-nearest neighbors (KNN), support vector machine (SVM), decision tree (DT), random forest (RF), and logistic regression algorithms were tested alongside XGBoost for an all-around check on their correctness. Experimental results show that classification accuracy of with PCA showed 94.44%, 93.06%, 94.44%, 94.44%, 93.06%, and 94.44%, on applying PCA showed result 95.83%, 94.44%, 94.44%, 94.44%, 94.44%, and 94.44% respectively. The Logistic Regression achieved the highest classification accuracy of 95.83%, and its precision, recall and F1-score values on the testing data were 94.50%, 94.44% and 94.44%, respectively. The findings of the study increase the efficiency of the proposed algorithms in classifying sleep disorder. It contributes to the efforts that have been made in enhancing the accuracy of diagnostics. This integrated review declares the immense potential of machine learning in creating high-quality, automatic sleep health solutions, which serve as doorsteps to future innovation in sleep disorder management.

Index Terms - Sleep disorders, Machine learning, Classification, Sleep stages, Diagnosis, Prediction, Sleep apnea, Sleep quality, Sleep patterns.

I. Introduction

Sleep is the most fundamental biological process through which an organism maintains both body and mental well-being. During this process, the body renovates through cellular repair, tissue growth and regulation of various hormonal processes. These functions are essential for maintaining energy levels, immuno-functionality and brain

functions. Despite the fact that sleep is one of the most overlooked health care in modern society, millions of people worldwide are experiencing disrupted sleep patterns or continuous sleep disorders. The conditions greatly impact the quality of life and result in a major public health challenge.

Sleep disorders can be broadly classified into several conditions: insomnia, obstructive sleep apnea (OSA), narcolepsy, restless leg syndrome (RLS) and circadian rhythm disorders. Each of these disorders affects sleep quality in a different manner, but they all contribute to negative health outcomes if left untreated. For instance, insomnia is a lack of ability to fall asleep or remain asleep, leading to continuous fatigue, low productivity, difficulty in concentrating and brain fog. Obstructive sleep apnea is marked by frequent occurrence of interruptions in breathing during sleep and may cause significant heart diseases like high blood pressure and strokes. Narcolepsy and other oversleeping disorders severely disrupt the individual's ability to remain awake during the day and seriously impact the activities of the day and the person's safety. Sleep disorders are not only of significance to health outcomes but also affect societies at large by causing accidents, poor performance at the workplace and increased healthcare costs.

Accurate diagnosis and timely treatment of sleep disorders are necessary in resolving the corresponding issues. The current preferred diagnostic standard is PSG, an detailed examination of sleep with measurement of multiple physiological parameters like brain activity, heart rate, muscle tone and breathing. PSG offers full breakdown on the structure of sleep, from transitions into the various stages: wakefulness, NREM (N1, N2, and N3), and REM. Though PSG is an efficient tool in the diagnosis of sleep disorders, its use has various limitations that restrict access to it and reduce effectiveness. Such limitations include the fact that it is relatively expensive, has specialized equipment, and labor-intensively relies on the manual evaluation of sleep data. Moreover, human interpretation of PSG recordings is very subjective and vulnerable to variability and mistakes in diagnosis.

We have identified a significant gap in research regarding sleep diagnosis, especially in applying machine learning algorithms to automate sleep stage classification and diagnosis of sleep disorders. Traditional methods have relied on slow and error-prone manual analysis, but machine learning now offers an exciting alternative. It uses datasets to uncover complex patterns and relationships within sleep data. Our main research paper *[1]* tackles this issue by looking into how machine learning can improve the accuracy and efficiency of diagnosing sleep disorders. It’s the first reference in the list. Results of the study showed effectiveness of algorithms in classifying the stages of sleep and detecting potential sleep disorders.

In the main paper, 92.92% was achieved by Artificial Neural Network (ANN) optimized through Genetic Algorithm (GA). ANN, in its own right possessed impressive performance without optimization at 91.15%. Other algorithms for instance, include Random Forest at 88.5% and Decision Tree at 86.73%. The SVM, at a low accuracy of 64.6% when it had not been optimized, showed a dramatic efficiency rise to 92.04% when it had been optimized through Genetic Algorithm. This further demonstrates the essence of parameter tuning in increasing the efficiency of a model. The performance of algorithms such as KNN as well as unoptimized SVM were the lowest, so more research is required in them so that their efficiency might be increased. This highlights the strength of machine learning and the extent to which sleep medicine may eventually change but highlights also that algorithm improvement and performance analysis must continue in search.

It basically seeks to balance between traditional methodologies for diagnosing the sleep disorders and new automatic approach based on their comparison in view of evaluating their performances by many different machine learning algorithms used for the sleep disorder classifications. The main goal of this research is to determine the most suitable algorithms for the automation of sleep stage classification and the diagnosis of sleep disorders through a comparison of the accuracy, precision, and overall effectiveness of different models. The research also included the impact of the feature selection and dimensionality reduction techniques like Principal Component Analysis (PCA) on the model performance. These are in line with a bigger picture; better diagnosis accuracy and efficiency in sleep medicine which will in the end lead to better patient results and lighter burden on the healthcare systems.

This study covers a broad range of the analysis applied to machine learning techniques for sleep data. This evaluation process considers traditional models such as Logistic Regression, Decision Trees, and Random Forests; it considers even more sophisticated algorithms such as SVM, KNN, and XGBoost. The evaluation work is directed at these algorithms to determine how they might compare with metrics that include accuracy, precision, recall, F1 score, and computational efficiency. The study checks the stability and the generalizability of the models across different datasets for their real life applicability. It focuses on these factors to create a solid system for automatically identifying sleep stages and diagnosing sleep disorders.

This research has effects that are massive. Automation of the diagnostic process by machine learning will help in a significant reduction in time and resources spent in sleep disorder diagnosis. The use of the machine learning models would thus allow healthcare service providers to focus on running and managing plans along with the management of patients, thereby making the health system more efficient. Furthermore, machine learning models would reduce inconsistent evaluations and variation in judgments because of the subjective nature in the manual analysis currently used. The results would also be more precise and reproducible due to the machine learning model. This is important in light of the increased demand for sleep disorder diagnosis and management medical services brought about by the rising incidence of sleep disorders across the world.

II. Related Work

In [2], the authors used machine learning algorithms in predicting obstructive sleep apnea in Korean adults. They studied various algorithms including LR, SVM, RF, and XGB. The results obtained showed that SVM was the best model among them all with sensitivity of 80.33%, specificity of 86.96%, and a AUC of 0.87 that showed the best model for prediction of OSA. The opposite was observed where XGBoost performed the least among the algorithms under evaluation. The study concluded that SVM could be used for making predictions on OSA due to better results. The major limitation observed was the low dataset, which impacts the possible generalizability of findings with studies that used larger datasets.

This study classifies the sleep stages by single-channel EEG signals analyzed through machine learning approaches. A wide variety of algorithms have been tested in this work such as SVM, Naive Bayes, KNN, Decision Trees, Bagging, AdaBoost, and ensemble methods. Of those algorithms, the ensemble SVM with bagging achieved a maximum accuracy of 96.85%. Because it gave better results than the other classifiers, the proposed authors brought forward automatic sleep staging using this ensemble SVM with bagging. Still, the study also revealed that scenario-independent tasks continued to be an obstacle to the effectiveness of the model, as in[3].

The authors discussed how simple and more interpretable methods can still be competitive for sleep scoring. They compared the traditional machine learning techniques with deep learning models and found that, although the methods were very simple, they surprisingly worked well when combined with detailed preprocessing and feature extraction. The study presented here obtained MF1 scores of 0.810 on the SC-20 dataset and 0.795 on the ST dataset. These results further suggest that sometimes simpler models can often perform well rather than the more complicated deep learning approaches. One of the great advantages of traditional models is interpretability-an extremely important factor in clinical situations, where understanding how decisions are being made is vital. Yet another point was made regarding the importance of how perhaps classic methods might miss important nonlinear patterns that deep learning models sense better. As mentioned in [4], automatic deep learning excels very much in automatically discovering complex data interaction but traditional machine learning should still play a major role as an alternative that is perfectly viable when interpretability becomes the first priority.

In this paper, halfwave decomposition technique has been applied to the EEG signal data so that the reduced-complexity set of data is achieved with most of the essential features. An innovative algorithm was proposed by the authors for diagnosing sleep disorders in [5]. The use of reduced data, extracted after the application of data reduction, for statistical feature extraction to obtain critical patterns from the simplified EEG signals was used. Those features were then classified by using the K-Nearest Neighbors algorithm, which has achieved high accuracy in different stage categorizations. The model was tested on the MIT-BIH Polysomnographic Database with 95.73% overall accuracy, 94.82% sensitivity, and 96.65% specificity. But the key limitation, according to which this is mentioned in the authors, is that this model had been validated only with a particular dataset. The limitation here presents a difficulty in terms of generalization since the model may not perform as well with other datasets, especially those having different characteristics or with clinical environments that varied. Testing should thus continue with more diverse datasets to establish how applicable the algorithm is across broader contexts in the diagnosis of sleep disorders. Classification methodology to classify the Obstructive Sleep Apnea Syndrome with various severities has been combined with clustering technique in [6].

Machine learning algorithms - Hierarchical Agglomerative, K-means, and Gaussian Mixture Models have been developed for clustering purposes. XGBoost, LightGBM, CatBoost, and Random Forest have been used as classification algorithms. Both models have achieved a very high success rate. Results have been achieved as 88.16% for mild, 87.84% for moderate and 91.06% for severe OSAS cases. Limitations of this study involve source data from single institution and difficulties in handling missing values, arguments in favor of the model suggest that it merits wider application for screening OSAS severity and therefore is useful in early diagnosis, and targeted intervention. More diversified and comprehensive validation will significantly add to the generalizability of these results. Another research proposal a system that used DT, KNN, and RF algorithms for the sleep stages classification from ECG in [7].

The authors of the work have exploited a widespread dataset ISRUC Sleep that was recorded from adult patients with two conditions: normal and sleep disorders. The signal record was randomly selected from the Hospital of Coimbra University Sleep Medicine Centre in PSG. Statistical features were utilized in the sleep attributes analysis. DT, KNN, and RF algorithms performed the best for automated sleep stages. The most significant observation was that the algorithm achieved a classification accuracy of over 90%, which is also better than the DT and KNN algorithms. Furthermore, investigators [8] proposed a model in which common MLAs, or DT, KNN, RF, and deep algorithms were utilized in detecting sleep apnea from one-lead ECG.

Here, the PhysioNet ECG Sleep Apnea v1.0.0 dataset of 70 records applied hybrid convolution-recurrent CNNs for feature extraction. Deep DRRNs had captured the data's time-pattern, and these authors used the principal component analysis for dimensional reduction. Hybrid CNN-DRNN architecture made more accurate detections as compared with the other algorithms. The authors have recommended using hybrid deep neural networks for detection of sleep apnea from ECG. Researchers proposed a model by which conventional MLAs such as DT, KNN, RF, and deep algorithms are being used for single-lead ECG sleep apnea detection.
The authors in *[9]* have used the PhysioNet ECG Sleep Apnea v1.0.0 dataset that contains 70 records and applied hybrid convolutional-recurrent CNNs followed by DRRNs for feature extraction to establish the time pattern of the data. They have applied principal component analysis for dimensional reduction. The accuracy detection of the hybrid CNN-DRNN architecture was better than that of the other algorithms. They also proposed hybrid deep neural networks to detect sleep apnea from ECG. More than one MLA was adopted, such as extreme gradient boosting (XGB), light gradient boosting machine (LGBM), CatBoost (CB), random forest (RF), K-nearest neighbors (KNN), logistic regression (LR), and support vector machine (SVM), for a high pretest OSA patient diagnosis as OSA or non-OSA for early diagnosis.
They have utilized the Wisconsin Sleep Cohort database by using the algorithms they had designed to test the model. The number of clinical records stands at 1,479; the features found in these records are a set of blood reports and physical measurements amongst other types. Bayesian optimization and genetic algorithms addressed the problems arising when hyperparameter tuning of this model was done. The accuracy rate for the SVM algorithm was at 68.06%, sensitivity at 88.76%, specificity at 40.74%, and an F1-score of 75.96%. Other researchers have developed a sleep-staging model by combining the conventional machine learning with a deep learning approach in designing an automatic classifier of sleep stages on multimodal signals in *[10]*.
Their methodology utilized random forest (RF), K-nearest neighbors (KNN), support vector machine (SVM), and deep learning algorithms that combined CNN with LSTM. The CNN was to be used for feature extraction of distinct features from EEG signals, while the usage of LSTM was to model the temporal dynamics. This model was tested on a public database, Sleep-EDF, and reported a return accuracy of 87.4%, which is more than the other algorithms. The researchers also highlighted that the data recorded by the patient was noisy, but they use the Butterworth filter to filter the data. In another article under review, 48 papers were assessed to discuss the importance of sleep apnea and its challenges.
The authors highlighted that different MLAs, like SVM, RF, and deep learning algorithms, can be used effectively for sleep apnea detection from the ECG signals with proper preprocessing in *[11]*. At the same time, they pointed out challenges in using the MLAs for classifying sleep apnea issues, such as the variability of the signals in ECGs and limited availability of datasets for training the models. Their study showed that SVM and deep learning-based neural networks were the best for detecting sleep apnea from ECG signals.

III. METHODOLOGY

A. MATERIALS AND METHODS

This paper offers an integrated approach for classifying sleep disorders by combining powerful data analysis and machine learning algorithms. For this purpose, we used Python 3.12 for the analysis and uses Pandas for manipulating the data, Scikit-learn to implement models, XGBoost to apply ensemble learning, and Matplotlib and Seaborn to produce static plots, with Plotly enabled for interactive analysis. All these tools and resources provided a strong framework to perform pre-processing, feature engineering, and machine learning evaluation.

This encompasses defined steps like data collection and pre-processing, feature selection by genetic algorithms, model training, and evaluation. All of them have been so designed that it follows a systematic approach to reduce biases and gives reliable insights about sleep disorders.

A flowchart **Figure 1**, which explains the entire process ranging from acquiring data to the evaluation of the model, is given at the end for better clarity. This provides a visual approach towards methodology by demonstrating sequential steps and their interconnection.

B. REAL SLEEP HEALTH AND LIFESTYLE DATASET

The study #used data acquired from Kaggle under the sleep health and lifestyle. This is featured with a wide range of features regarding patterns of sleep, along with physiological measurements and lifestyle factors.

It included demographic characteristics like age, gender, and occupation; lifestyle factors like the level of physical activity, level of stress, and BMI category; and physiological parameters like heart rate, blood pressure, both systolic and diastolic, and metrics of sleep quality. The target variable was "Sleep Disorder," a means of classifying people into distinct conditions in terms of sleep. There were missing values, such as those in the column "Sleep Disorder", that were imputed. Then, to ensure reliability of the data, duplicate entries were removed.

Refer to **Table 1**, which gives a few observations from the dataset to provide an overview and better understanding of its structure and attributes.

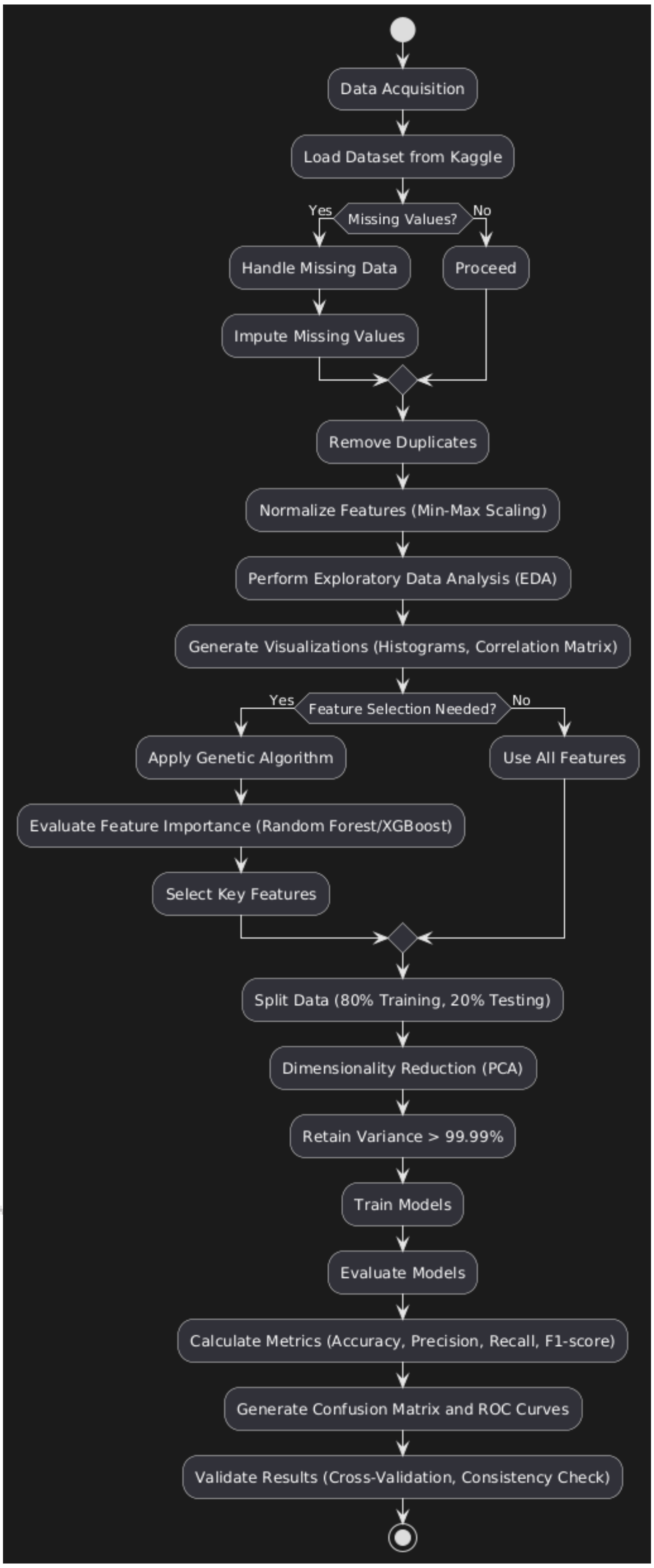


Figure 1. *Flowchart Illustrating the entire process*

Table 1. ***Overview of the sleep health and lifecycle dataset used in this study.***

Person ID	Gender	Age	Occupation	Sleep Duration (hrs)	Quality of Sleep (1-10)	Physical Activity Level	Stress Level (1-10)	BMI Category	Blood Pressure (mm Hg)	Heart Rate (bpm)	Daily Steps	Sleep Disorder
1	M	27	Software Engineer	6.1	6	42	6	Overweight	126/83	77	4200	None
2	M	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	None
3	F	32	Teacher	7.0	7	50	5	Normal	120/78	70	8500	None
4	M	45	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea
5	F	38	Nurse	6.0	5	40	7	Overweight	130/85	80	5000	Insomnia

C.EXPERIMENT DESIGN

Structured Approach: The overall experimental design led to a better evaluation of good classification algorithms, for there was an 80%:20% proportion of the whole dataset set out for training purposes and testing respectively. Missing data imputation occurred, duplication checking, and some outliers were filtered out using Interquartile Range (IQR). All features, which included "Blood Pressure," needed to be categorized into systolic and diastolic values and analyzed in minute detail. Other categorical variables like gender, occupation, and BMI category, were label-encoded to be compatible with machine learning models.

Feature engineering included the use of genetic algorithms for feature selection, which determined the most relevant features for classification. PCA was applied for dimensionality reduction that retained 99.999% of the variance with reducing the feature set. Every machine learning model was trained and tested using 5-fold cross-validation to avoid overfitting and get statistical strength. This ensured that meaningful patterns and relationships of the data were extracted, while fair comparison between classification methods could be achieved. See **Figure 1** for information.

D.PERFORMANCE METRICS

The metrics set for measuring the performance of the model included accuracy, precision, recall, F1-score, and stability. Accuracy was calculated on the basis of the correctness of the predictions overall, whereas precision calculated the reliability of the positive predictions. On the other hand, the recall measured the sensitivity of the models towards true positive cases, and the F1-score defined a harmonic mean of precision and recall, giving a balanced evaluation of performance. Stability, obtained from the standard deviation of cross-validation scores, measures the stability of the models on different data splits. The comparison of all the performance matrix excluding stability is given below in **Figure 2**.

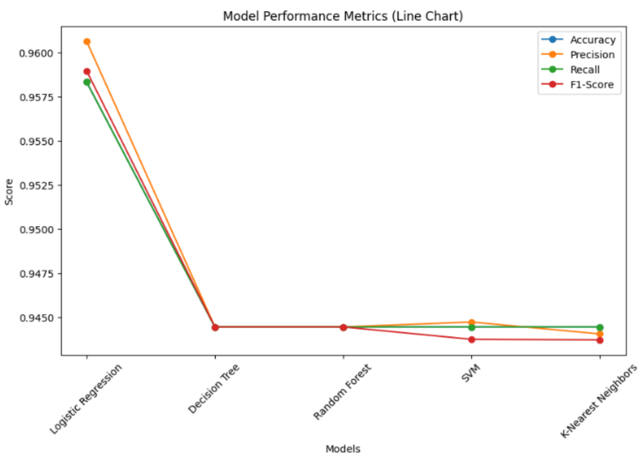


Figure 2. ***Line chart of performance matrix.***

E. CLASSIFICATION ALGORITHMS

Several machine learning models were used to predict sleep disorders and each has its advantages. The baseline model used is Logistic Regression which provides an interpretable linear classification model. Decision Trees ensured transparency in decision-making, therefore were useful to healthcare professionals. Random Forest and XGBoost, which are both ensemble models, provided strong performance through the use of many weak learners combined to reduce errors.

Support Vector Machines (SVM) were used due to their capability to manage non-linear relationships via kernel functions. K-Nearest Neighbors (KNN), a non-parametric algorithm, predicted data according to closeness in feature space, providing simple decision-making. All the algorithms were tested with the same experimental setup to allow for equal comparison of their performance regarding sleep disorder classification. You may see **Figure 2** for comparison of their performance.

F. FEATURE IMPORTANCE

An important part of the methodology is a feature importance analysis which determines the most important factors for the classification of sleep disorder. This is carried out by means of several approaches, we have use the Pearson correlation heatmap for identifying the important features which can be observed in **Figure 3**.

The analysis of feature importance would enable comprehension of the hidden patterns of sleep disorder, potentially helping health care professionals in useful ways. Such information could allow the targeting of these more relevant factors in diagnosis and treatment planning.

G. CORRELATION COEFFICIENT

In order to understand the relationships among various features in the dataset, correlation analysis is carried out. Pearson correlation coefficients are computed to measure the linear relationship between these parameters. It is used to find the redundant features and interdependencies among different sleep and lifestyle parameters.

It recognized these highly correlated features, thus efficiently eliminating redundancy in the data and reducing input to classification algorithms. Therefore, this analysis obtained meaningful information about relationships between lifestyle patterns and physiological measures, thus giving deeper insights into factors that affect sleep disorders.

The **Pearson heatmap** revealed several notable relationships:

- **Stress levels** were highly negatively correlated with **quality of sleep** (-0.91) and moderately with **sleep duration** (-0.80), indicating that increased stress significantly impacts sleep quality and duration.
- **Daily steps** showed a strong positive correlation with **physical activity levels** (0.82), emphasizing the relationship between activity and movement patterns.
- **Blood pressure values (systolic and diastolic)** were highly correlated with each other (0.98), which aligns with physiological expectations.

These all relationships can be verified from **Figure 4** which is on the next side.

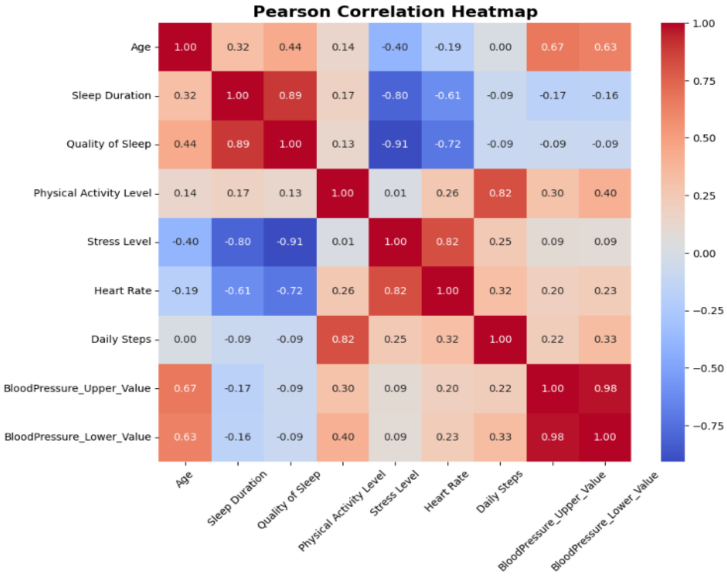


Figure 3: ***Pearson correlation heatmap between all features.***

IV. RESULTS AND DISCUSSION

A. MODEL PERFORMANCE ANALYSIS

The performance of several machine learning algorithms was analyzed on the Sleep Health and Lifestyle Dataset with features such as sleep habits, physical activity levels, stress, and other health factors. The mode was evaluated on an 80% training set and 20% testing set of dataset. We compared both cases with and without PCA to see how reducing dimensions impacts the performance of the algorithms. The performance metrics that were used for evaluation were Accuracy, Precision, Recall, F1-Score, and Stability. Stability was measured in terms of the standard deviation of the cross-validation accuracy.

B. RESULTS WITHOUT PCA

From **Table 2,** Logistic Regression, Random Forest, Support Vector Machine (SVM) and XGBoost proved to be excellent algorithms. Both the models from Random Forest and XGBoost separately obtained the accuracy of 94.44%. These models possessed outstanding precision and recall values; therefore, the models seem to work efficiently when classifying the sleep disorders. The Decision Tree and K-Nearest Neighbors (KNN) models, although registering fair accuracy, were somewhat unstable because of overfitting when the limited dataset was used.

The Random Forest and XGBoost models did very well without PCA being performed, with perfect precision and recall combination, pointing out how strongly these models were suited for requirements of the sleep disorder classification. Also their scores of stability lie very low because these models stand stronger and seem to be highly insensitive to variability in the set. Decision Tree and KNN models are still accurate but with greater variability, which makes them riskier sources of overfitting and thus less appropriate to apply in real scenarios.

Model	Accuracy	Precision	Recall	F1	Stability
Logistic Regression	94.44	94.50	94.44	94.44	0.0342
Decision Tree	93.06	93.37	93.06	93.18	0.0303
Random Forest	94.44	94.44	94.44	94.44	0.0205
SVM	94.44	94.50	94.44	94.44	0.0295
KNN	93.06	93.40	93.18	93.18	0.0321
XGBoost	94.44	94.44	94.44	94.44	0.0304

Table 2. ***Without PCA***

C. RESULTS WITH PCA

As presented in **Table 3**, the highest accuracy was obtained with Logistic Regression using PCA, at 95.83%. All other models performed worse than Logistic Regression, including precision (96.06%), recall (95.83%), and F1-score (95.90%) with the lowest stability score of 0.0176, meaning it is highly consistent. All other models (Decision Tree, Random Forest, SVM, KNN, and XGBoost) maintained an accuracy of 94.44%. PCA slightly reduced the stability of Decision Tree but had a negligible effect on accuracy for these models.

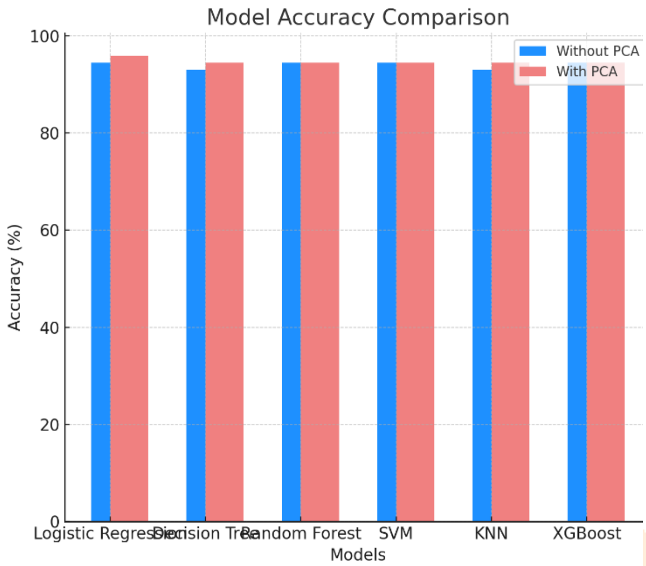
Logistic Regression benefited the most, with a highly improved accuracy and stability from PCA. This shows that the model strongly responds to reducing its dimensionality and, as such, it will do well in systems with limited computational capabilities. Random Forest and XGBoost remained stable, proving that they are able to deal effectively with high dimensional datasets without necessarily needing high dimensionality reduction.

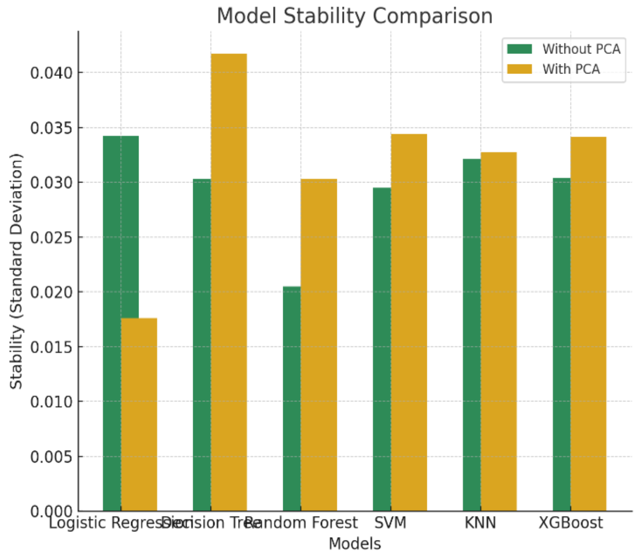
Table 3. ***With PCA***

Model	Accuracy	Precision	Recall	F1	Stability
Logistic Regression	95.83	96.06	95.83	95.90	0.0176
Decision Tree	94.44	94.44	94.44	94.44	0.0417
Random Forest	94.44	94.44	94.44	94.44	0.0303
SVM	94.44	94.47	94.44	94.37	0.0344
KNN	94.44	94.40	94.44	94.37	0.0327
XGBoost	94.44	94.44	94.44	94.44	0.0341

D. COMPARATIVE ANALYSIS OF MODELS

It was evident that the inclusion of PCA had significantly improved Logistic Regression as a model, considering the improvements it made in terms of accuracy and stability scores, which then rendered it the most consistent and accurate model for this dataset. In contrast, models like Random Forest, SVM, and XGBoost were more or less comparable to each other and performed almost the same in both experiments, thus implying that PCA had no strong effect on these results. The comparison can between accuracy and stability can be clearly observed in below **Figure 4** and **Figure 5** respectively.





E. KEY INSIGHTS

across all metrics with PCA, which gave it the highest accuracy at 95.83% and lowest stability score at 0.0176. PCA improved both the dimensionality reduction and overall consistency for Logistic Regression but had little impact on the other models. Models such as Random Forest, SVM, and XGBoost maintained their accuracy and performance with and without PCA, indicating robustness to dimensionality changes. The combination of PCA and Logistic Regression is recommended for this dataset due to its enhanced performance and consistency compared to other models.

The comparison of the results obtained in our work with the primary paper gives great insight into the applicability of machine learning algorithms to the classification of sleep disorders.

In the main paper, the best accuracy achieved was 92.92% through an Artificial Neural Network optimized using a Genetic Algorithm. Without tuning, ANN remained the leader with an accuracy of 91.15%, while Random Forest (RF) and Decision Tree (DT) followed with accuracies of 88.5% and 86.73%, respectively. The performance of SVM was outstandingly improved from 64.6% without tuning to 92.04% with GA, as this again points out that proper parameter tuning is essential. Their results for algorithms such as KNN and SVM without tuning were relatively lower than others.

Our results showed better performance for most algorithms, especially with PCA. The best accuracy achieved with Logistic Regression at 95.83% outperformed the best results found in the main paper. Other algorithms such as Decision Tree, Random Forest, SVM, KNN, and XGBoost obtained a consistent 94.44% accuracy using PCA, thus having minimal variance in their performances. Without PCA, performance was a bit worse: logistic regression and Random Forest held up to 94.44% accuracy while the decision tree and KNN plummeted to 93.06%. Our result stability measures also establish robustness: variability in the accuracy of models is lower in our experiment as well, when compared with the main paper.

This comparison further reveals the strength of feature dimensionality reduction techniques, such as PCA, in terms of improving the model's performance and stability. Our results not only surpass the accuracy achieved by the main paper but also show consistent stability across multiple algorithms, thereby establishing the robustness of our approach.

IV. Conclusion

In this paper, an optimized machine learning model for sleep disorder classification is presented by using multiple algorithms such as Logistic Regression, Random Forest, XGBoost, SVM, KNN, and Decision Trees. The best accuracy of 95.83% was achieved by the model of Logistic Regression with PCA, and its precision, recall, and F1 scores were reported to be 96.06%, 95.83%, and 95.90%, respectively, along with unparalleled stability with a variance of 0.0176.

The dataset is quite limited, which would probably affect the generality of the models. Nonetheless, the research solves the problem of classifying sleep disorders by showing the capacity of MLAs in handling high-dimensional data and credible predictions.

Future work would focus on enhancing the accuracy even more with the application of deep learning techniques, much larger and diversified datasets, and optimized models toward real-time applications in the clinic. This research proves the MLAs, particularly Logistic Regression, to be quite promising in classification of sleep disorders and further asserts their relevance to both clinical practice and research.

References

[1] Talal Alshammari, "Applying machine learning algorithms for the classification of sleep disorders", *IEEE Access*, 2024.

[2] 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, Mar. 2021.

[3] H. O. Ilhan, ''Sleep stage classification via ensemble and conventional machine learning methods using single channel EEG signals,'' Int. J. Intell. Syst. Appl. Eng., vol. 4, no. 5, pp. 174–184, Dec. 2017.

[4] 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.

[5] Yash Paul, Rajesh Singh, Surbhi Sharma, Saurabh Singh, In-Ho Ra, "Efficient Sleep Stage Identification Using Piecewise Linear EEG Signal Reduction: A Novel Algorithm for Sleep Disorder Diagnosis", Sensors, vol.24, no.16, pp.5265, 2024.

[6] H. Han and J. Oh, ''Application of various machine learning techniques to predict obstructive sleep apnea syndrome severity,'' Sci. Rep., vol. 13, no. 1, p. 6379, Apr. 2023.

[7] S. Satapathy, D. Loganathan, H. K.Kondaveeti, and R. Rath, ''Performance analysis of machine learning algorithms on automated sleep staging feature sets,'' *CAAI Trans. Intell. Technol.*, vol. 6, no. 2, pp. 155–174, Jun. 2021.

[8] M. Bahrami and M. Forouzanfar, ''Sleep apnea detection from single-lead ECG: A comprehensive analysis of machine learning and deep learning algorithms,'' *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[9] J. Ramesh, N. Keeran, A. Sagahyroon, and F. Aloul, ''Towards validating the effectiveness of obstructive sleep apnea classification from electronic health records using machine learning,'' *Healthcare*, vol. 9, no. 11, p. 1450, Oct. 2021.

[10] S. K. Satapathy, H. K. Kondaveeti, S. R. Sreeja, H. Madhani, N. Rajput, and D. Swain, ''A deep learning approach to automated sleep stages classification using multi-modal signals,'' *Proc.*

[11] Ghasemi, H. Khazaie, A. Daneshkhah, and A. Ahmadi, ''Detection of sleep apnea using machine learning algorithms based on ECG signals: A comprehensive systematic review,'' *Expert Syst. Appl.*, vol. 187, Jan. 2022, Art