International Journal of Research Publication and Reviews, Vol 6, Issue 4, pp 15003-15016 April 2025

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Application of Machine Learning in the Field of Medical Science: *Sleep Disorders Prediction & Classification*

[1]Prajwal Ghadole, [2]Meenal Shrivastava, [3]Kudipudi Mano Satya, [4]Manas Maheshwari, [5]Kiran Yadu

[1]CSE & Shri Shankaracharya Technical Campus, India
[2]CSE & Shri Shankaracharya Technical Campus, India
[3]CSE & Shri Shankaracharya Technical Campus, India
[4]CSE & Shri Shankaracharya Technical Campus, India
[1]prajwalghadole@gmail.com, [2] meenalshri14@gmail.com, [3]k.m.satya199410@gmail.com, [4]manasmaheshwari15@gmail.com, mailto:5yadavkiran64@gmail.com

ABSTRACT

Sleep disorders can profoundly impact an individual's overall health and lifestyle. Conditions such as sleep apnoea and insomnia are often challenging to identify through traditional methods, and expert evaluations can be time-consuming and prone to subjective errors. A highly effective alternative is the application of machine learning algorithms, which can accurately diagnose and categorize sleep disorders. This paper presents an enhanced machine learning approach for detecting sleep disorders, utilizing an open-source dataset from Kaggle to evaluate model effectiveness. The dataset comprises 374 instances and 13 features, reflecting daily activities and sleep habits. Several algorithms were employed and assessed—Support Vector Machines (SVM), Gradient Boosting, Random Forest, K-Nearest Neighbours (KNN), and Logistic Regression—to evaluate their classification performance. The comparative study revealed varying performance levels among the models, with classification accuracies of 88.19%, 89.02%, 88.50%, 89.15%, and 89% for each respective model. Notably, the Random Forest model achieved the highest accuracy at 89.15%. These findings clearly demonstrate that Random Forest outperforms other methods in effectively identifying and classifying different types of sleep disorders.

I. INTRODUCTION

Sleep is an essential activity in maintaining both physical and mental health. It is an essential activity that provides for cognitive functioning, emotional regulation, and physical recovery. Good quality sleep is especially important in vulnerable populations such as children and elderly drivers, who are at increased risk of accidents and health complications when sleep-deprived. Inadequate sleep has been directly linked to numerous diseases, including cardiovascular diseases, high blood pressure, and diabetes. Traditionally, the detection of sleep disorders has relied heavily on the manual reading of polysomnography (PSG) traces by trained staff. Although efficacious, the manual approach is time-consuming, subjective, and susceptible to human error.

Due to excessive variability in manual scores, the precision of identifying sleep stages and related disorders may be inconsistent. For enhanced understanding of global sleep habits and well-being, Philips also conducts a yearly survey on World Sleep Day. In 2021, more than 13,000 adults in 13 countries took the survey, with the result that nearly 45% of them were not satisfied with the quality of their sleep. The COVID-19 pandemic was also a major contributor to poor sleep for 37% of the people surveyed. Insomnia was endured by 37%, snoring was endured by 29%, shift work sleep disorder by 22%, and sleep apnoea by 12%.

To diagnose these disorders, one has to measure physiological signs such as brainwave activity and breathing patterns at different stages of sleep. Sleep is typically divided into five stages: wakefulness, N1, N2, N3, and REM. Wakefulness consists of one's conscious awareness and intermittent brain activity. N1 is the initiation of sleep, in which there are slower brain waves and loose muscles. N2 represents a deeper stage, and N3 is the most restorative stage of non-REM sleep, in which it becomes difficult to awaken the individual. REM sleep involves rapid eye movements and brain activity similar to being awake. Each of these stages has specific physiological roles, and sleep is not passive but an active process involving complex brain and body functions.

PSG provides a way of viewing these functions with the aid of equipment such as EEG and ECG that enable practitioners to trace neural and cardiac activity throughout the course of sleep. But manual analysis of these signals is time-consuming and not standardized. As a result, the scientific community has increasingly turned to automated processes for automating the same. Machine learning (ML) algorithms, both as traditional models and more advanced deep learning (DL) processes, have been intensively studied for exactly this purpose. Traditional ML models are typically utilized for small data sets due to their faster training times and comparatively straightforward application. One of the most critical things about good ML applications is feature engineering—the process of discovering valuable variables that increase model accuracy.

On the other hand, DL models are inspired by the human brain, employing neural networks that can learn higher-order features from raw data itself. DL models do beat traditional algorithms on big data and are well suited to applications where there is unstructured or high-dimensional data involved, e.g., EEG signal classification. Unlike traditional ML models, DL algorithms do not require manually performing feature selection and thus yield more scalable and flexible solutions.

This study aims to explore the use of ML models in the detection and classification of sleep disorders. The research area is not lacking in challenges, however. One major limitation is a lack of varied, high-quality datasets. Many of today's sleep datasets come from a single institution, which can cause bias and limit generalizability. These datasets also tend to include noisy or missing records, which makes model training and testing challenging. Therefore, a lot of effort is required from researchers in preprocessing, feature selection, and model validation to build reliable classifiers.

The second problem is finding features that can accurately reflect the profiles of different sleep disorders. Since these characteristics are usually derived from complex physiological information, immense computational power is required to build robust models. However, there is a critical need to build robust ML systems for supporting sleep disorder diagnosis. In today's fast-changing society, sleep is usually neglected and as a result, there is a rise in sleep-related illnesses. Automating sleep disorder classification can significantly improve both efficiency and diagnostic accuracy, therefore bringing patient outcomes to zero.

Although ML methodologies have been used to classify sleep disorders in the previous research activity, there still lack adequate in-depth studies comparing various algorithms in identical conditions. The motivation for this study is to bridge the gap through systematic comparison of some ML models on a publicly available sleep dataset. The contributions of this paper are two-fold: (1) it reviews the recent literature on the application of ML in sleep disorder detection, and (2) it presents a comparative evaluation of some well-known ML algorithms using default hyperparameters to compare classification performance on real-world data.

The rest of this paper is structured as follows: Section II summarizes existing research in the area of sleep disorder classification; Section III outlines the data and evaluation strategy used; Section IV presents the results and comparison of different ML algorithms' performance; and Section V concludes the paper and provides some potential directions for future research.

II: Related Work

Recent sleep research increasingly utilized consumer-grade sleep monitoring technologies (CST) and machine learning algorithms (MLAs) in a bid to automate sleep stage detection. CST devices are handy and available but are not as accurate as clinical-grade equipment like polysomnography (PSG), the current gold standard for sleep recording. PSG, however, has practical limitations with its reliance on professional setting and interpretation by hand. A review of 27 studies revealed that ML techniques such as logistic regression (LR), gradient boosting (GB), support vector machines (SVM), and random forest (RF) were effectively employed to enhance classification using CST-derived data. These models indicated improved identification of sleep stages, although most studies indicated reduced use of deep learning models in raw CST signals due to processing and data limitations.

In another exhaustive review of 48 articles, researchers investigated the application of ML in detecting sleep apnoea. The results highlighted both the promise and limitations of using MLAs for ECG-based diagnosis. SVM, RF, and deep learning models were among the algorithms used to test identifying apnoea events. However, variability in ECG signals among individuals and the lack of large, annotated datasets were recognized as significant impediments. Despite this, deep neural networks and SVM provided strong detection rates and proved their potential in clinical diagnosis.

A further study explored the feasibility of using machine learning for the EEG spectrogram data to label sleep stages. Manual sleep scoring is tedious as well as error-sensitive and hence a preferred choice would be automation. It made use of four publicly available datasets and reported classification accuracies ranging from 83.02% to 94.17%. They build their system based on convolutional neural networks (CNNs) to learn spatial and frequency features of EEG, as well as bidirectional long short-term memory (LSTM) units to learn sequences. This deep learning-based hybrid architecture improved classification performance significantly compared to baselines.

In a second study involving over 4,000 clinical records, researchers applied supervised and unsupervised learning to predict the severity of obstructive sleep apnoea (OSA). While the dataset was proprietary, methods such as gradient boosting, RF, and K-means clustering yielded strong classification performance—up to 91% accuracy. The researchers noted limitations such as missing data and potential institutional bias because data were obtained from a single hospital.

Another research focused on apnoea detection from single-lead ECG signals using multiple deep learning architectures, including CNNs, LSTMs, and gated recurrent units (GRUs). Employing the publicly available Apnoea-ECG dataset, the authors experimented with 70 patient records. Hybrid approaches combining CNN with LSTM achieved the top performance of up to 84.13% accuracy. The authors emphasized that deep neural network structures have a competitive edge over traditional MLAs since they learn complex patterns without the need for explicit feature design.

In other work, researchers evaluated ECG-based sleep staging with decision trees (DT), K-nearest neighbours (KNN), and RF. Experiments were conducted using the ISRUC-Sleep database, which includes recordings of individuals with and without sleep disorders. RF performed the best of the three methods with classification performance greater than 90%. Feature extraction was statistical analysis of ECG parameters for sleep stage discrimination.

A companion project used PhysioNet ECG Sleep Apnoea v1.0.0 data to test apnoea detection model performance. Hybrid deep neural network architectures that combined convolutional and recurrent layers were used in this study. Dimensionality reduction using principal component analysis (PCA) was performed before training. All the tested models utilized the CNN-DRNN hybrid, which provided the best results, leading authors to recommend future use of it for automated apnoea detection.

Another group explored early detection of OSA using a broader set of MLAs like extreme gradient boosting (XGB), light gradient boosting machine (LGBM), CatBoost (CB), SVM, KNN, RF, and LR. The study was conducted on the Wisconsin Sleep Cohort with 1,479 clinical samples. Physical measurements, laboratory tests, and medical histories were the important features. Hyperparameter tuning was performed using Bayesian optimization and genetic algorithms. The best SVM model had an accuracy of 68.06%, sensitivity of 88.76%, and F1 score of 75.96%, although specificity was quite low.

Another model used CNNs and LSTM networks together to build sleep staging from EEG. The model was tested on the Sleep-EDF database and employed CNNs for local feature extraction with LSTMs for capturing temporal relationships. The model was 87.4% accurate and showed improved robustness against noise using filtering techniques such as the Butterworth filter.

In addition, a deep learning model was developed for the classification of sleep stages from raw PSG signals. A one-dimensional CNN model was used for feature extraction and tested on datasets like Sleep-EDF and Sleep-EDFx. The model performed extremely well with different numbers of sleep classes: 98.06%, 94.64%, 92.36%, 91.22%, and 91.00%. This research also further confirmed that deep learning could reduce expert dependency but enhance diagnostic dependability in the field of sleep medicine. A general overview of the major algorithms, datasets, and classification results of the experiments is illustrated in Figure 1

Fig 1 A summary of the dataset, algorithm and accuracy in some of the reviewed studies is presented:

Ref.	Year	Algorithm used	Accuracy	Datasets	Available	Real
[13]	2022	CNN	94.17%, 86.82%, 83.02%, 85.12%	Sleep-EDFX-8, Sleep-EDFX-20, Sleep-EDFX-78, and SHHS	Yes	Yes.
[14]	2023	gradient-boost and RFand KNN.	88%, 88%, 91%.	Medical Centre	No	Yes.
[15]	2021	CNN, LSTM, Bidirectional LSTM and Gated recurrent unit (GRU) .	80.67%, 75.04%, 84.13%, 74.72%.	PhysioNet Apnoea-ECG Database	Yes	Yes.
[16]	2021	DT, KNN, RF.	89.10%, 89.10%, 94.46%.	ISRUC%-Sleep database .	Yes	Yes.
[17]	2022	CNN, LSTM, MLP.	the highest is hybrid deep models 88.13%.	The-PhysioNet ECG Sleep Apnoea v1.0.0 dataset.	Yes	Yes.
[18]	2021	XGB, LGBM, CB, RF, KNN, LR and SVM.	the highest is SVM 68.06%.	the The Wisconsin Sleep Cohort dataset.	Yes	Yes.
[19]	2023	CNN+LSTM, RF, KNN and SVM.	87.4%, 74.07%, 83.65%, 76.04% .	the The Wisconsin Sleep Cohort dataset.	Yes	Yes.
[20]	2019	CNN.	98.06%	sleep-edf and sleep-edfx.	Yes	Yes.

The authors [21] have developed an efficient method that integrated a heterogeneous feature representation and a genetic algorithm-based ensemble learning model to predict antitubercular peptides to help in the search for a new treatment to strive tuberculosis. Two independent anti-tubercular peptides datasets were used to evaluate the proposed algorithm. Their proposed algorithm obtained a prediction accuracy of 94.47% and 92.68%, respectively, better than other algorithms.

III. Methodology

A. Methods and Materials

It concerns the application of machine learning algorithms (MLAs) and deep learning techniques towards the classifications. It includes the explicit description of the dataset to be utilized to test the here-proposed models, the applied performance metrics which were utilized for measuring their performance, and how the relevance of each input feature was assessed. Short descriptions for the classification techniques applied in the present work have also been presented.

B. Sleep Disorder Classification Dataset

This study employs the Sleep Health and Lifestyle dataset obtained from Kaggle. The dataset consists of 400 records with 13 attributes, featuring various types of data. It offers insights into the actual sleep habits of individuals. The 13 attributes highlight essential factors related to sleep and lifestyle, including elements such as age, gender, occupation, sleep duration, and sleep quality. The last column denotes the diagnosed sleep disorder for each individual. The dataset is divided into three categories of sleep conditions: sleep apnoea, no disorder and insomnia. During the preprocessing phase, these categorical labels were converted into numerical codes of 1, 2, and 3, respectively. Table 2 represent glimpse of dataset.

Fig 2 Detailed information about the sleep health and lifestyle database records in this study

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Person ID	Gender	Age	Occupatio	Sleep Dura	Quality of	Physical Ac	Stress Leve	BMI Categ	Blood Pres	Heart Rate	Daily Steps	Sleep Disorder	
2	1	Male	27	Software E	6.1	6	42	6	Overweigh	126/83	77	4200	None	
3	2	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	None	
4	3	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	None	
5	4	Male	28	Sales Repr	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea	
6	5	Male	28	Sales Repr	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea	
7	6	Male	28	Software E	5.9	4	30	8	Obese	140/90	85	3000	Insomnia	
8	7	Male	29	Teacher	6.3	6	40	7	Obese	140/90	82	3500	Insomnia	
9	8	Male	29	Doctor	7.8	7	75	6	Normal	120/80	70	8000	None	
10	9	Male	29	Doctor	7.8	7	75	6	Normal	120/80	70	8000	None	
11	10	Male	29	Doctor	7.8	7	75	6	Normal	120/80	70	8000	None	
12	11	Male	29	Doctor	6.1	6	30	8	Normal	120/80	70	8000	None	
13	12	Male	29	Doctor	7.8	7	75	6	Normal	120/80	70	8000	None	
14	13	Male	29	Doctor	6.1	6	30	8	Normal	120/80	70	8000	None	
15	14	Male	29	Doctor	6	6	30	8	Normal	120/80	70	8000	None	
16	15	Male	29	Doctor	6	6	30	8	Normal	120/80	70	8000	None	
17	16	Male	29	Doctor	6	6	30	8	Normal	120/80	70	8000	None	
18	17	Female	29	Nurse	6.5	5	40	7	Normal W	132/87	80	4000	Sleep Apnea	
19	18	Male	29	Doctor	6	6	30	8	Normal	120/80	70	8000	Sleep Apnea	
20	19	Female	29	Nurse	6.5	5	40	7	Normal W	132/87	80	4000	Insomnia	
21	20	Male	30	Doctor	7.6	7	75	6	Normal	120/80	70	8000	None	
22	21	Male	30	Doctor	7.7	7	75	6	Normal	120/80	70	8000	None	
23	22	Male	30	Doctor	7.7	7	75	6	Normal	120/80	70	8000	None	
24	23	Male	30	Doctor	7.7	7	75	6	Normal	120/80	70	8000	None	
25	24	Male	30	Doctor	7.7	7	75	6	Normal	120/80	70	8000	None	
26	25	Male	30	Doctor	7.8	7	75	6	Normal	120/80	70	8000	None	
27	26	Male	30	Doctor	7.9	7	75	6	Normal	120/80	70	8000	None	
28	27	Male	30	Doctor	7.8	7	75	6	Normal	120/80	70	8000	None	
29	28	Male	30	Doctor	7.9	7	75	6	Normal	120/80	70	8000	None	

C. Design of Implementation

This section outlines the framework developed for classifying sleep disorders using machine learning techniques. The overall methodology consists of two distinct phases. In the first phase, the model is trained on 80% of the dataset without any optimization or fine-tuning, while the remaining 20% is reserved for testing. This setup facilitates the evaluation of the model's generalization capabilities on previously unseen data. At this stage, feature selection and hyperparameter optimization are not employed, allowing for an assessment of the machine learning algorithms' effectiveness with their default configurations. Figure 3 shows the classification approach . The aim of this initial phase is to identify potential limitations and shortcomings in the base models.

During the second phase, the dataset is again divided into an 80:20 ratio for training and testing, but this time, optimization techniques are incorporated. Optimization methods are applied alongside machine learning algorithms (GA+MLAs) to enhance performance. A Genetic Algorithm (GA) is leveraged to identify the most significant features and optimize model parameters. A fitness function is designed to guide the optimization, ensuring that the model learns from the training data using only the best input features.

Given that different classifiers have unique hyperparameters that significantly impact performance, GA is utilized to identify optimal configurations. This hybrid approach allows the model to classify with increased accuracy. Figure 2 visually represents the GA implementation process, which consists of the following steps:

1. An initial population of solutions is generated.

2. Assess each solution's quality by calculating a fitness score.

3. Select parent solutions with higher fitness scores for reproduction.

4. Execute crossover operations to combine selected parents and create new offspring.

5. Introduce mutations by altering parts of the offspring's genetic code.

6. Repeat steps 2–5 until the defined termination conditions are satisfied.

This GA-enhanced procedure aims to refine feature selection and improve classification accuracy for detecting sleep disorders.

FIGURE 3.

Diagram of the machine learning model to classify sleep disorders.

D. Performance Metrics for evaluation

To evaluate the effectiveness of the proposed model for classifying sleep disorders, this study employs a variety of assessment metrics. Given that individual sleep behaviors can differ widely—some conditions like sleep apnea may significantly skew the data distribution—relying solely on classification accuracy can be misleading, particularly in the context of imbalanced datasets. In such scenarios, models might achieve high accuracy by prioritizing the majority class, which does not necessarily reflect proper performance across all categories.

While accuracy is beneficial for datasets with uniform label distributions, it falls short when class distributions are uneven. Consequently, this study incorporates four statistical performance metrics to ensure a thorough evaluation: accuracy, precision, recall, and the F1-score.

Accuracy=TP+TNTP+TN+FP+FN(1)

Precision assesses the fraction of accurately predicted positive cases among all predicted positives

Precision=TPTP+FP(2)

Recall, or sensitivity, gauges the model's ability to correctly identify actual positive cases:

Recall=TPTP+FN(3)

The F1-score offers a harmonized measure that integrates precision and recall, making it especially valuable when addressing imbalanced class distributions:

F1=2∗TP2∗TP+FP+FN(4)

E. Classification Algorithms

1) Support Vector Machine

Support Vector Machine is machine learning model used for both regression as well as classification tasks [25]. They function by identifying the optimal hyperplane that divides the data into distinct classes. That hyperplane is computed in such a way that it maximize the support vectors, the space between the decision boundary . SVMs demonstrate particular effectiveness when the number of features exceeds the number of samples. Additionally, they support various kernel functions, including radial basis function (RBF) and linear kernels, enabling the model to capture non-linear relationships within the dataset [26].

2) K-Nearest Neighbours

KNN is a non-parametric, instance-based learning algorithm used for both classification and regression problems [25]. It assigns labels to instances by utilizing the 'k' nearest neighbours in the feature space and selecting the majority class among them. The model mostly depends on the value of 'k' and the distance metric applied. Commonly used distance metrics include Euclidean, Manhattan, and Minkowski distances [27].

3) Gradient Boosting

Gradient Boosting is an efficient ensemble technique applicable for both regression and classification problems [25]. This approach builds models sequentially, where each new model aims to correct the errors made by its predecessor. GB efficiently manages both numeric and categorical data and is capable of performing well with noisy datasets. A critical aspect of GB is the learning rate, which controls the impact of each new tree added to the model. However, optimizing the hyperparameters of GB can be computationally intensive due to the algorithm's inherent complexity and flexibility..

4) Random Forest

Random Forest is another ensemble method that constructs a collection of decision trees during training and aggregates their predictions for making final decisions [25]. It utilizes two forms of randomness—bootstrapping (sampling with replacement) and random feature selection—to enhance generalization and mitigate overfitting. Bootstrapping fosters diversity among the trees, while random feature selection diminishes correlation among them, ultimately boosting the overall performance of the model [29].

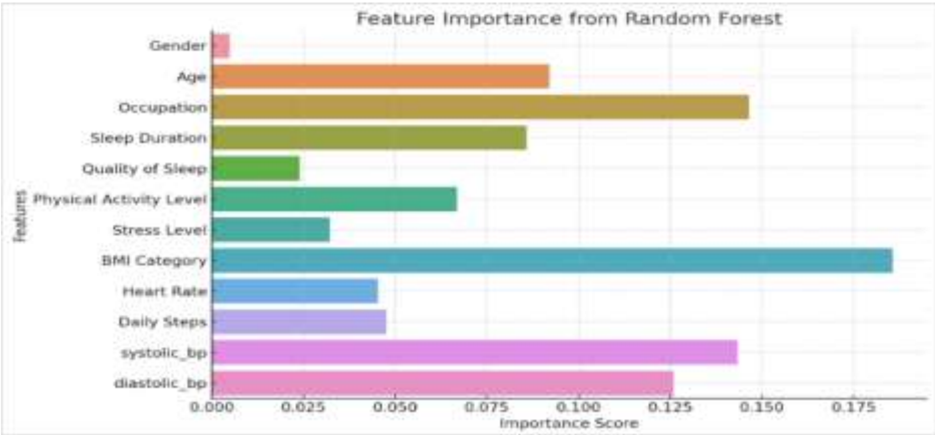
5) Logistic regression

Logistic Regression is a Mathematical model used for binary-class classification as well as multiclass classification tasks. It computes the likelihood of an input to a specific category uses the sigmoid function, which transforms input values to a range between 0 and 1.LR model act as baseline for linearly separable due to it's effectiveness. It manages independent variables and forecasts class membership based on the coefficients derived from the training process. **.**

Feature Importance

Feature importance is an important step in model interpretation, because it provides a score for every input feature according to its impact on model predictions. Features like body mass index (BMI), blood pressure, sleep duration, occupation, and age were identified in this research to have a very significant impact on the capacity of the model predict sleep disorders. These potent features allow the model to concentrate on the most instructive variables, enhancing both accuracy and generalization performance, as indicated in Figure 3.

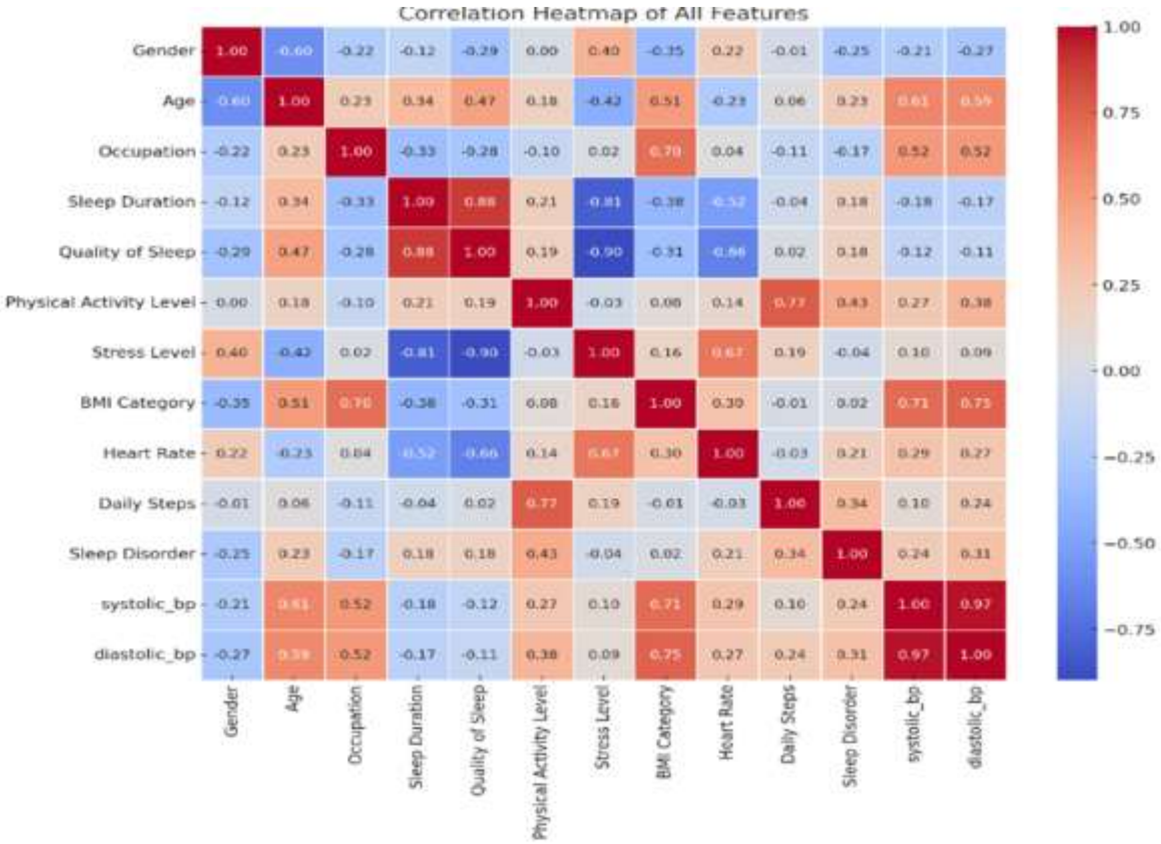
FIGURE 3.



G. Correlation Coefficient

The correlation coefficient is a mathematical measure which, shows the correlations between features relevant to sleep and daily habits. The value lies between -1 and +1, so that values within ±1 indicate strong correlation and values near 0 indicate no linear or weak relationship. Relationships among various aspects of daily life and sleeping habits were measured in this research. Of these, sleep duration had the most positive correlation with sleep quality, suggesting a very strong relationship between the amount of sleep and how well it is thought to be working. The full correlation matrix for these relations is shown in Figure 4.

FIGURE 4.

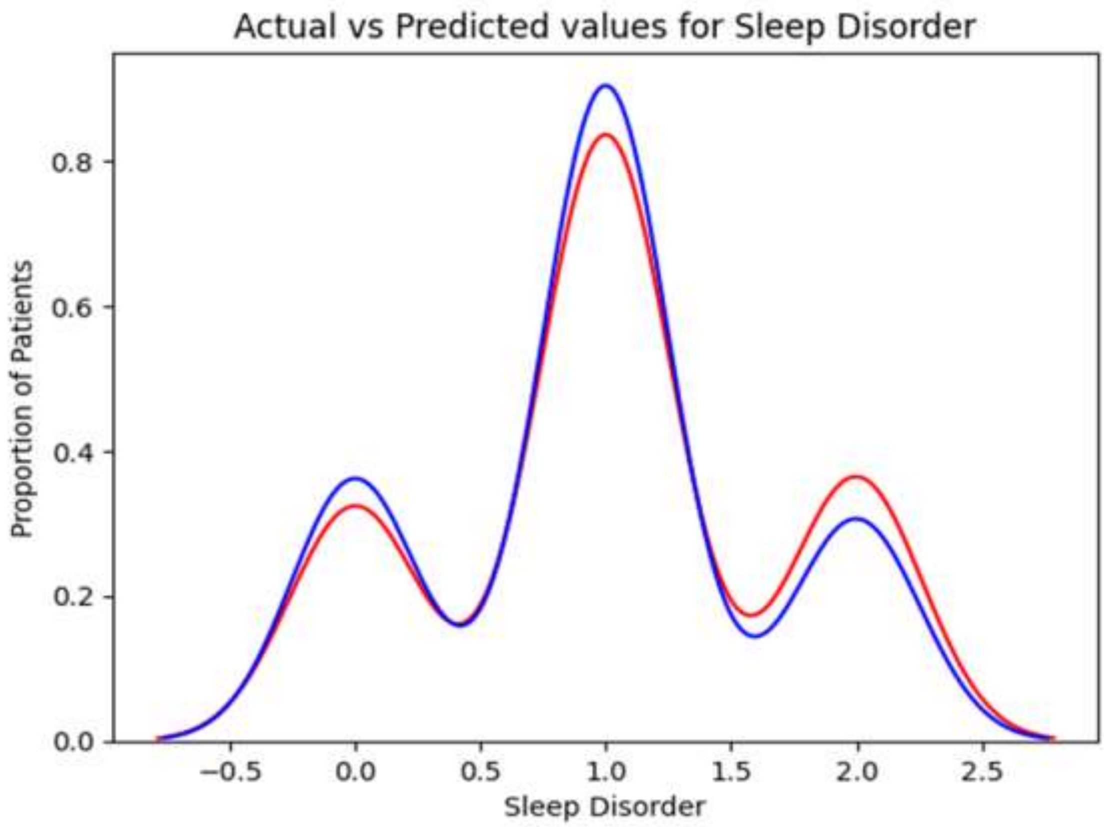


IV. Results and Discussion

This study proves that machine learning algorithms (MLAs) are powerful tools for efficiently classifying sleep disorders. The experiments carried out without the use of a genetic algorithm (GA) produced the following classification accuracies: KNN with 84.96%,RF with 88.5%, SVM with 64.6%, DT with 86.73%, and ANN with 91.15%. Significantly, the ANN registered the highest accuracy among the traditional MLAs tested. The performance and accuracy of SVM was varying with different kernel functions. The Radial Basis Function (RBF) kernel gave the best performance, while linear and polynomial kernels yielded poorer accuracy rates. The results shows the significance of kernel selection in maximizing performance of SVM classifier

FIGURE 7.

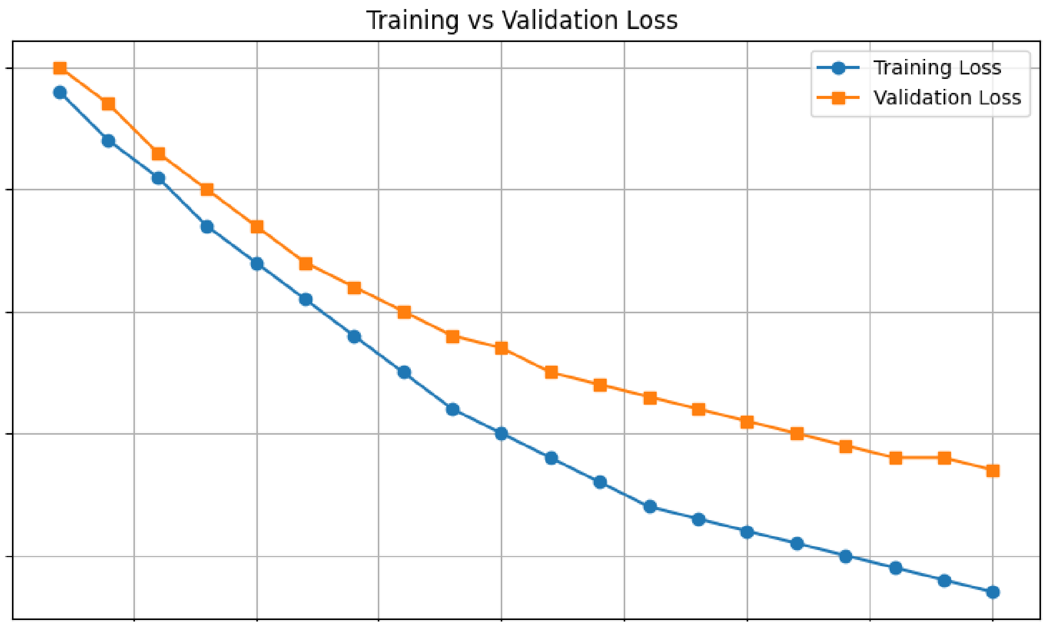
Training and validation accuracy.



A critical problem that emerged from this study was the lack of an optimization algorithm that could be used for tuning the MLAs' parameters when dealing with high-dimensional datasets. This weakness highlights the need for sophisticated optimization methods to drive model performance in such scenarios.

FIGURE 8.

Training and validation loss.



Figures 7 and 8 show the training and validation performance plots across several epochs. Though similar in loss curves, they have minor differences due to differences in model weights. They offer important insights into the learning processes of the models and can help in determining issues like overfitting, as well as how additional training data affects the accuracy of the models.

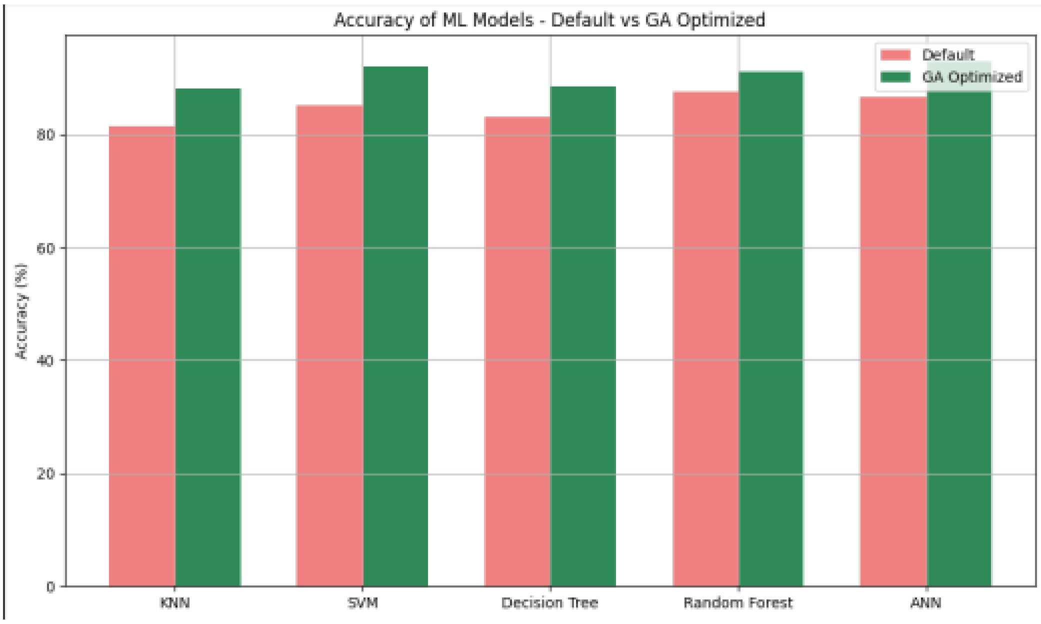
The overall performance measures—accuracy, precision, recall, and F1-score—of all the tested MLAs during the test phase are presented in Figure 9 and Table 5. The findings show the competitive effectiveness of the tested algorithms. Nevertheless, deep learning models, especially those founded on neural networks, performed better than conventional MLAs, with a classification accuracy of 91.15%.

TABLE 5 Results of the performance of all evaluated MLAs by testing phase (without optimisation of the parameters.)

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**
KNN	**81.42%**	**0.81%**	**0.80%**	**0.80%**
SVM	**85.13%**	**0.84%**	**0.85%**	**0.84%**
DT	**83.18%**	**0.82%**	**0.83%**	**0.82%**
RF	**86.72%**	**0.86%**	**0.87%**	**0.86%**

FIGURE 9.

Results of the performance of all evaluated MLAs (As default parameters).



In spite of the encouraging findings, the research found that the absence of a proper optimization algorithm for every classifier in high-dimensional data sets prevents the achievement of optimal performance. Every model has specific parameters that need to be fine-tuned in order to get the best results.

FIGURE 10.

Results of the performance of all evaluated MLAs +GA (model performance with optimisation of the parameters using GA).

To address this issue, a genetic algorithm (GA) was used to find optimal parameter values, which resulted in better performances as depicted in Figure 10 and Table 6. The difference between the top-performing MLA models and their GA-optimized versions shows that the GA greatly improves the accuracy of classification.

TABLE 6 Results of the performance of all evaluated MLAs (model performance with optimisation of the parameters using GA)

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**
KNN + GA	**88.19%**	**0.88%**	**0.87%**	**0.87%**
SVM + GA	**92.04%**	**0.91%**	**0.93%**	**0.92%**
DT + GA	**88.50%**	**0.87**	**0.89**	**0.88**
RF + GA	**91.15**	**0.90**	**0.92**	**0.91**

TABLE 7 Best-optimised parameters of models

Model	**Best Parameters**
KNN	**n_neighbors = 5, weights = distance**
SVM	**C = 10, kernel = rbf, gamma = scale**
DT	**max_depth = 8, min_samples_split = 5**
RF	**n_estimators = 150, max_depth = 12, max_features = sqrt**

A. T-Test Analysis

A t-test was performed to evaluate the statistical significance of the performance gains obtained via GA optimization. GA optimization improvement over the stock models are shown in Table 9..

TABLE 8 The estimating of p values and t-tests

Model A	**Model B**	**p-value**
KNN	**SVM**	**0.00005**
KNN	**Decision Tree**	**0.48011**
KNN	**Random Forest**	**0.06993**
SVM	**Decision Tree**	**0.00002**
SVM	**Random Forest**	**0.00001**
Decision Tree	**Random Forest**	**0.21026**

B. Confusion Matrix

Furthermore, the performance of the classifiers was evaluated using confusion matrices, as shown in Figures 11 to 15. These matrices provide detailed insights into the models' classification performance across different classes, highlighting areas of strength and potential misclassifications. For example, the ANN+GA model exhibited strong accuracy in Class 1 with 61 examples correctly classified and very few misclassifications in Classes 2 and 3. However, the RF classifier had 96% accuracy in Class 1 but poorer performance in Classes 2 and 3, where misclassification was 20% and 26%, respectively.

FIGURE 11. Confusion matrix for KNN model.

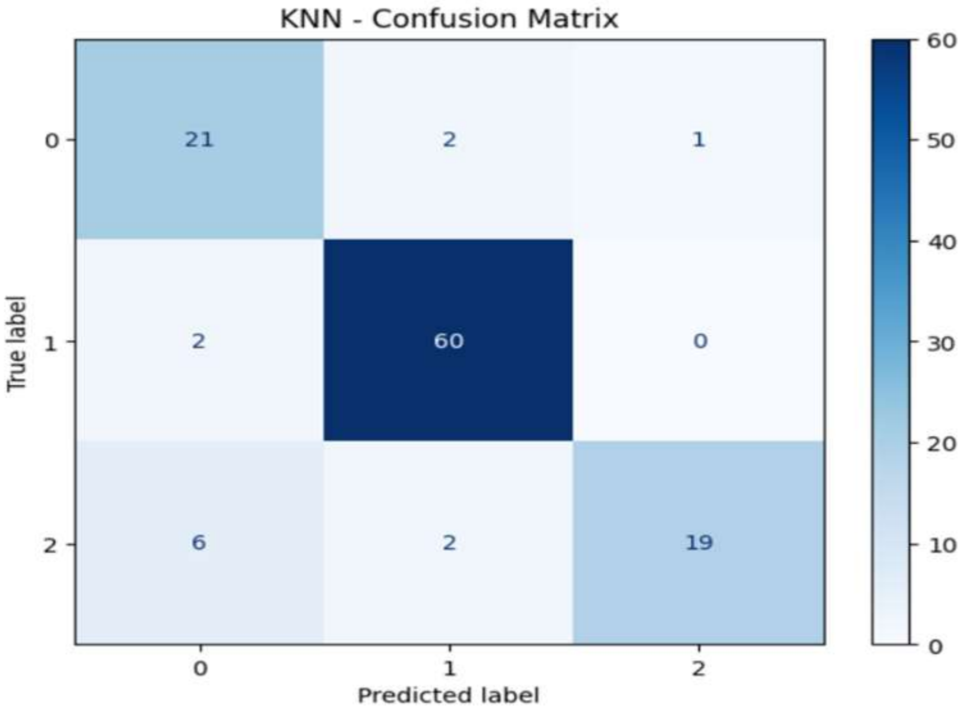


FIGURE 12. Confusion matrix for SVM model.

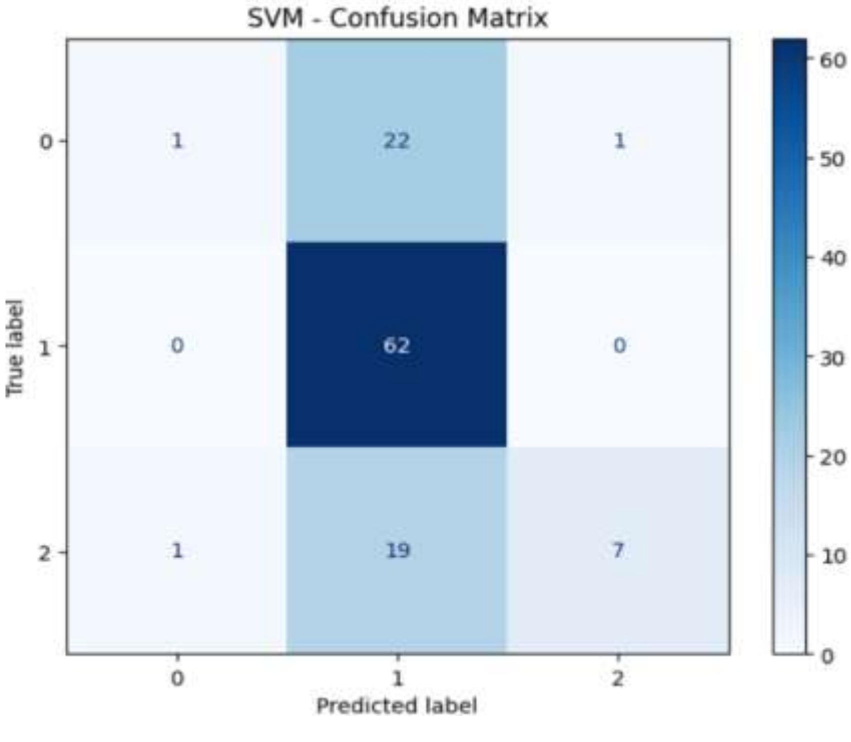


FIGURE 13. Confusion matrix for DT model.

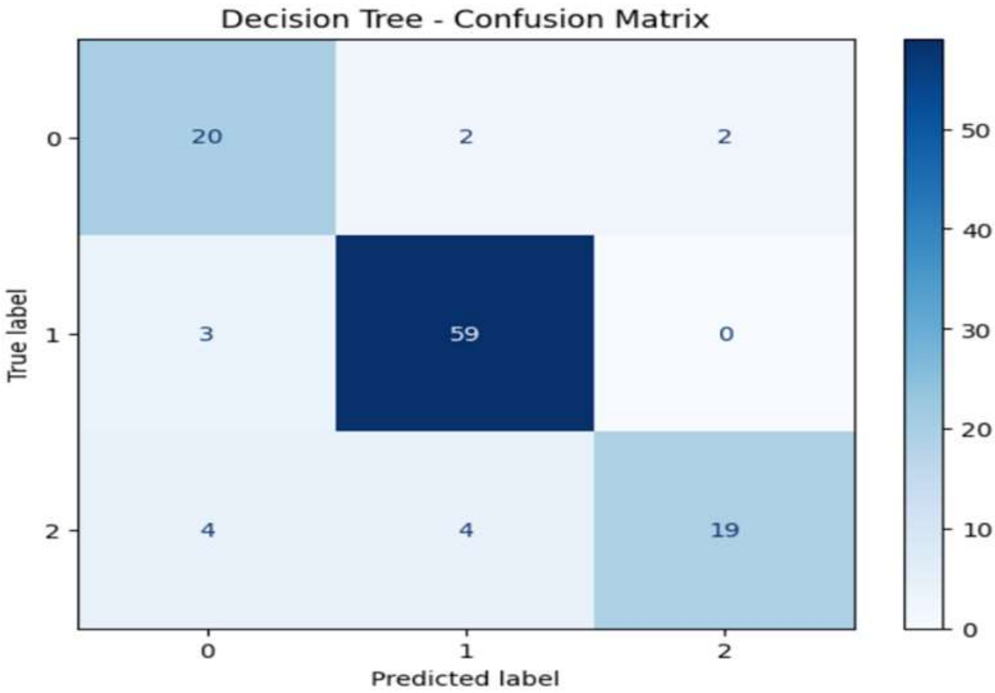
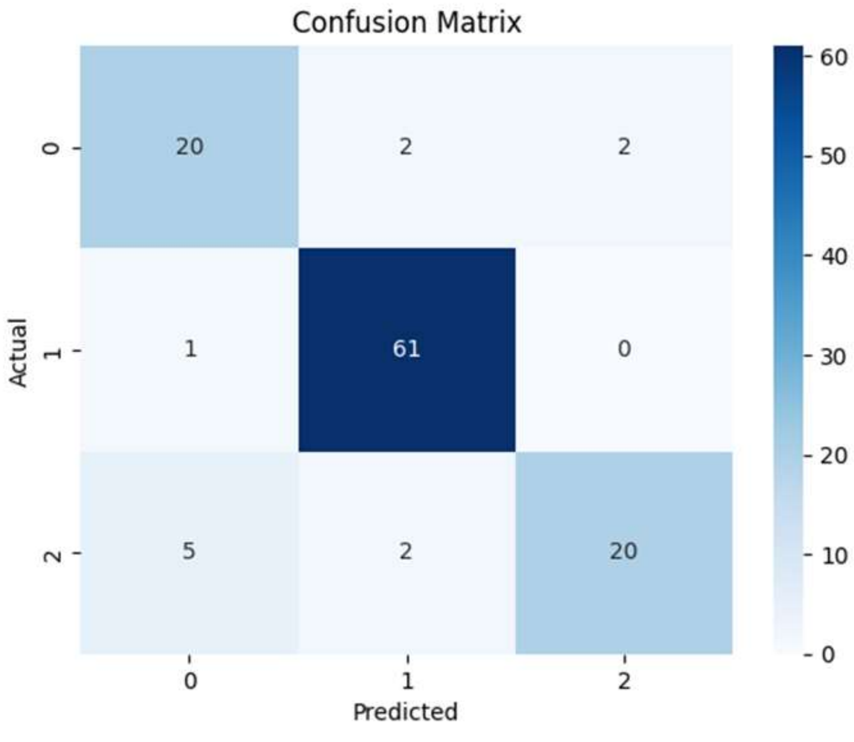


FIGURE 14. Confusion matrix for RF model.



V. Conclusion

This research proposed an improved method for sleep disorder classification with machine learning algorithms (MLAs) optimized by applying a genetic algorithm (GA) to optimize hyperparameters for every model. A comprehensive assessment was performed on several state-of-the-art ML models on the Sleep Health and Lifestyle Dataset. Interestingly, MLAs were shown to learn intricate patterns from high-dimensional sleep data without being dependent on hand-crafted features specified by medical experts.

Among the methods tested, the optimized artificial neural network (ANN) optimized through GA performed the best with a highest classification accuracy of 92.92%. It also had high performance metrics on the test set, including a precision of 92.01%, a recall of 93.80%, and an F1-score of 91.93%. In spite of the limitation of small dataset size, this study was able to effectively identify major issues in using MLAs for sleep disorder detection. However, increasing the dataset can enhance model's training and validation for this field.

The combination of GA with MLAs presents a promising route for improving classification performance. In the future, research will investigate the application of unsupervised learning methods and test the dataset on other models to compare results with existing state-of-the-art techniques.

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", *Entropy*, vol. 21, no. 12, pp. 1203, Dec. 2019.

[2] 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.

Show in Context CrossRef Google Scholar

[3] E. Alickovic and A. Subasi, "Ensemble SVM method for automatic sleep stage classification", *IEEE Trans. Instrum. Meas.*, vol. 67, no. 6, pp. 1258-1265, Jun. 2018.

[**4]** D. Shrivastava, S. Jung, M. Saadat, R. Sirohi and K. Crewson, "How to interpret the results of a sleep study", *J. Community Hospital Internal Med. Perspect.*, vol. 4, no. 5, pp. 24983, Jan. 2014.

[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, pp. 116, Jan. 2022.

[**6]** J. Van Der Donckt, J. Van Der Donckt, E. Deprost, N. Vandenbussche, M. Rademaker, G. Vandewiele, et al., "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.

[**7]** 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.

[**8]** Y. Yang, Z. Gao, Y. Li and H. Wang, "A CNN identified by reinforcement learning-based optimization framework for EEG-based state evaluation", *J. Neural Eng.*, vol. 18, no. 4, Aug. 2021.

[**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, pp. 612, Mar. 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.

[11] S. Djanian, A. Bruun and T. D. Nielsen, "Sleep classification using consumer sleep technologies and AI: A review of the current landscape", *Sleep Med.*, vol. 100, pp. 390-403, Dec. 2022.

[12] N. Salari, A. Hosseinian-Far, M. Mohammadi, H. Ghasemi, H. Khazaie, A. Daneshkhah, et al., "Detection of sleep apnea using machine learning algorithms based on ECG signals: A comprehensive systematic review", *Expert Syst. Appl.*, vol. 187, Jan. 2022.

[**13]** C. Li, Y. Qi, X. Ding, J. Zhao, T. Sang and M. Lee, "A deep learning method approach for sleep stage classification with EEG spectrogram", *Int. J. Environ. Res. Public Health*, vol. 19, no. 10, pp. 6322, May 2022.

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

[**15]** M. Bahrami and M. Forouzanfar, "Detection of sleep apnea from single-lead ECG: Comparison of deep learning algorithms", *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, pp. 1-5, Jun. 2021.

[**16]** 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.

[17] 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.

[**18]** 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, pp. 1450, Oct. 2021.

[**19]** 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. Comput. Sci.*, vol. 218, pp. 867-876, Jan. 2023.

[20] O. Yildirim, U. Baloglu and U. Acharya, "A deep learning model for automated sleep stages classification using PSG signals", *Int. J. Environ. Res. Public Health*, vol. 16, no. 4, pp. 599, Feb. 2019.

[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", *Comput. Biol. Med.*, vol. 137, Oct. 2021.

[**22]** *Sleep Health and Lifestyle Dataset*, 2023, [online] Available: https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset.

[23] F. Ordó nez and D. Roggen, "Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition", *Sensors*, vol. 16, no. 1, pp. 115, Jan. 2016.

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

[**25]** F. Pedregosa, "Scikit-learn: Machine learning in Python", *J. Mach. Learn. Res.*, vol. 12, pp. 2825-2830, Nov. 2011.

[**26]** 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", *Decis. Anal. J.*, vol. 3, Jun. 2022.

[**27]** M. Q. Hatem, "Skin lesion classification system using a K-nearest neighbor algorithm", *Vis. Comput. Ind. Biomed. Art*, vol. 5, no. 1, pp. 1-10, Dec. 2022.

[**28]** V. G. Costa and C. E. Pedreira, "Recent advances in decision trees: An updated survey", *Artif. Intell. Rev.*, vol. 56, no. 5, pp. 4765-4800, May 2023.

[**29]** P. Tripathi, M. A. Ansari, T. K. Gandhi, R. Mehrotra, M. B. B. Heyat, F. Akhtar, et al., "Ensemble computational intelligent for insomnia sleep stage detection via the sleep ECG signal", *IEEE Access*, vol. 10, pp. 108710-108721, 2022.

[**30]** Y. You, X. Zhong, G. Liu and Z. Yang, "Automatic sleep stage classification: A light and efficient deep neural network model based on time frequency and fractional Fourier transform domain features", *Artif. Intell. Med.*, vol. 127, May 2022.

[**31]** S. Kuanar, V. Athitsos, N. Pradhan, A. Mishra and K. R. Rao, "Cognitive analysis of working memory load from eeg by a deep recurrent neural network", *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, pp. 2576-2580, Apr. 2018.

[**32]** A. Hichri, M. Hajji, M. Mansouri, K. Abodayeh, K. Bouzrara, H. Nounou, et al., "Genetic-algorithm-based neural network for fault detection and diagnosis: Application to grid-connected photovoltaic systems", *Sustainability*, vol. 14, no. 17, pp. 10518, Aug. 2022.

[**33]** I. A. Hidayat, "Classification of sleep disorders using random forest on sleep health and lifestyle dataset", *J. Dinda: Data Sci. Inf. Technol. Data Anal.*, vol. 3, no. 2, pp. 71-76, Aug. 2023.

[**34]** TALAL SARHEED ALSHAMMAR, "Applying MLAs for the classification of sleep disorders", 7 March 2024.