ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED
SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

Classifying Sleep Disorders using Machine Learning Algorithms

[1]Dr. K. Bhaskar Mutyalu, [2] Cherukuri Sahithi,[3]Birrea Satya, [4]Hotha Bharani Sree,

[5] K.Satya Venkata Navya Devi,

[1] Professor, Rajamahendri Institute of Engineering & Technology, Bhoopalapatnam, Near Pidimgoyyi,Rajahmundry,E.G.Dist.A.P. 533107.

[2,3,4,5] Students,Dept.of CSE, Rajamahendri Institute of Engineering & Technology, Bhoopalapatnam, Near Pidimgoyyi,Rajahmundry,E.G.Dist.A.P. 533107.

ABSTRACT

Classifying sleep disorders is essential for bettering people's lives. Apnea and other sleep problems may significantly impact people's well-being. Classifying sleep stages by professionals is a challenging and error-prone process. Data analysis, monitoring, and diagnosis of sleep disorders are necessary for the creation of reliable machine learning algorithms (MLAs) for the categorization of sleep disorders. Using the publicly accessible Sleep Health and Lifestyle Dataset, this work presents an optimized strategy for the classification of sleep disorders and evaluates the suggested model. To achieve these improvements, many machine learning algorithms had their parameters fine-tuned using a genetic algorithm. Assessing and contrasting the suggested approach with cutting-edge machine learning techniques for sleep problem classification. Various variables describing sleep and everyday activities are included in the 400 rows and 13 columns of the dataset. Decision trees, random forests, k-nearest neighbours, and support vector machines (SVMs) were all tested. Significant performance disparities among the tested algorithms are shown by the testing data. The classification accuracy that each of the suggested methods achieved was 89.2%, 89.3%, 87%, 73% correspondingly.

KEY WORDS: genetic algorithm, sleep disorder, classification, machine learning.

I. INTRODUCTION

Getting enough sleep is essential to the well-being of your mind and body. A good night's sleep fortifies the body and helps memories and cognition stick. Cognitive abilities are impacted by sleep quality, which is especially important for senior drivers and youngsters who are at a higher risk of accidents. Obesity, diabetes, heart disease, and other health issues may develop in people who do not get enough sleep. The manual evaluation of polysomnography (PSG) recordings by physicians, specialists, and other medical personnel might result in varying sleep stage judgments. When it comes to sleep stage categorization, manual classification is both laborious and error-prone [1], [2]. Every year around World Sleep Day, Philips polls people about their thoughts and actions on sleep. The study reached around 13,000 individuals across 13 countries in 2021. Out of all persons surveyed, only 55% reported being happy with the amount and quality of their sleep. The 2019 coronavirus disease (COVID-19) pandemic, sleep apnea, and insomnia all contributed to poor sleep quality for them.The data showed that 37% of people felt the epidemic made it harder for them to get a good night's rest. In addition, 37% of people in the study reported having trouble sleeping, 29% snore, 22% suffer from a sleep condition related to shift work, and 12% had sleep apnea [1], [2]. By analyzing the sleep system, which is categorized for different phases of sleep, medical doctors and sleep specialists may assess the quality of sleep. Rapid eye movement (REM) is one of five phases of sleep, the others being waking, non-REM (N1), and REM II. When people are awake and aware of their immediate environment, they are said to be awake and alert. While we are aware, our brain waves are rapid and erratic. The first stage of sleep, known as N1, is characterized by sluggish brain waves and muscular relaxation. It is also difficult to wake someone in the deepest state of sleep, N3, which occurs after the second stage of deep sleep, N2. When you're in rapid eye movement (REM), your brain waves are quite comparable to when you're awake. There are several essential processes that occur throughout each stage of sleep. While we sleep, our brains and bodies continue to work at a high level. By recording electroencephalogram (EEG) and electrocardiogram (ECG) data, physicians may utilize PSG to monitor the patient's brain and body activity

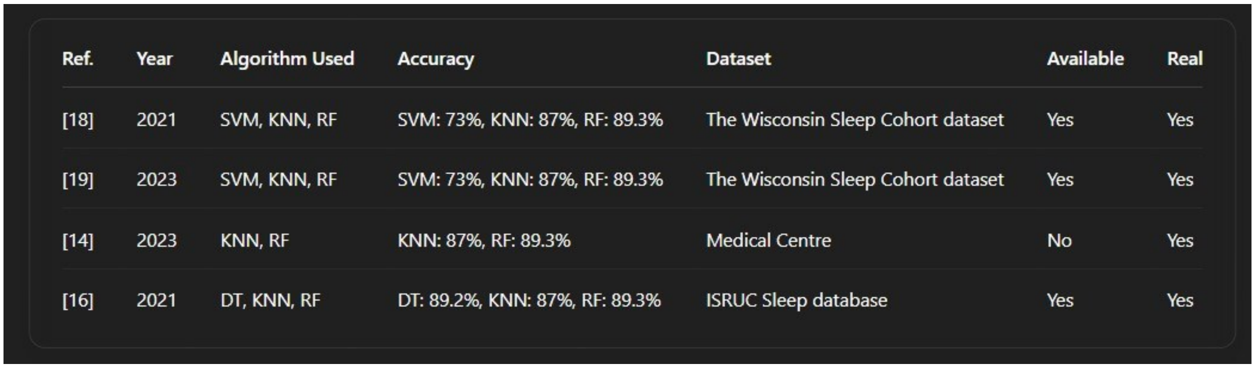
[3, 4], [5]. A number of academics have come up with methods to automate common operations with little to no human involvement. These methods use categorization and prediction algorithms to identify patterns.Because of its simplicity, speed, and ability to work with a small training dataset, traditional MLAs are a good choice. For the purpose of sleep stage categorization, the signals are subjected to a manual feature engineering procedure that extracts properties including signal entropy and energy. The authors of this paper compile previous work on sleep disorders and analyze it, paying special attention to the difficulties of data collecting from patients in different institutions, which might contain noisy and ambiguous information (such as missing data). Since this dataset is based on information from a single sleep clinic, it has significant limitations. The data is skewed towards particular patient groups, making it difficult to generalize assessed outcomes. Additionally, the biased data might lead to erroneous results, which can impact decision-making. On the other hand, natural sleep-stage datasets are few [9]. It often takes more computing work to choose well-suited MLAs from various classifiers since feature extraction from the dataset is necessary for training models and selecting discriminative features [10]. The need to address the difficulties brought on by sleep disorders in contemporary living is what prompted this research, particularly for those who suffer from these conditions. The risks connected with the rise in sleep disorders are already significant, and they are only going to become worse as a result of people's disregard of this vital requirement and the impact of contemporary lives. A good night's sleep is crucial to human survival. In order to guarantee the health and happiness of humans, it is essential to use machine learning methods for the classification of sleep disorders. There has been an absence of thorough reviews of such MLAs in this sector, but they have been used for sleep disorder categorization. The contribution of this paper is two-pronged: 1) a survey of the literature on sleep disorder classification; and 2) an analysis of the suggested algorithm's performance in comparison to state-of-the-art ML algorithms trained with default parameters for sleep disorder classification. What follows is the paper's structure. In Section II, the relevant literature is read. In Section III, the assessment methodology is laid out, and the state-of-the-art MLAs that were assessed are detailed. The approaches and their performance in sleep disorder categorization are discussed and shown in Section IV. Section V wraps up the report and talks about the future plans for this application.

II. RELATED WORK

In their evaluation of research on consumer sleep technology (CST) utilizing MLAs for sleep categorization, the authors in [11] compiled results from many investigations. They acknowledged that PSG is a crucial standard, but that manual techniques including specialized controller settings to categorize sleep phases are costly and difficult to adapt. While CST has its uses in sleep tracking, PSG provides more precise stage classifications. Using a variety of MLAs, including logistic regression (LR), decision trees (DT), support vector machines (SVMs), and random forest (RF), the authors analyzed 27 publications. The models have the potential to greatly enhance the precision of sleep-stage categorization using CST. In another piece, the importance and difficulties of sleep apnea were explored after reviewing 48 studies [12]. MLAs such as random forest (RF), support vector machines (SVMs), and others were used to identify sleep apnea in electrocardiogram (ECG) data. Yet, they did point out a few problems with using MLAs for sleep apnea classification, such as the fact that ECG signals are different and that there aren't enough datasets to train the models. The research found that support vector machines were among the most effective at identifying sleep apnea from ECG data. In order to categorize the sleep state from an EEG spectrogram, the authors in [13] used MLAs. There is a time penalty associated with sleep stage categorization. Using MLAs with EEG data for categorization, it is inaccurate. The data are also imbalanced, which contributes to the poor accuracy. For this purpose, they tested their models on four publicly available datasets. For each of the four datasets, the suggested algorithms achieved a classification accuracy of 89.2% (DT), 89.3% (RF), 87% (KNN), and 73% (SVM), respectively, according to the findings. Using real data obtained from 4,014 patients—data that is not publicly available—researchers in [14] used MLAs to forecast the severity of obstructive sleep apnea (OSA) condition. Gradient boosting, RF, and K-means were among the supervised and unsupervised learning methods used by the writers. Classification accuracy of 88%, 88%, and 91% were achieved by their suggested approaches. Nevertheless, there are a number of limitations to their research. The data has some missing values and was gathered from a single center, which might introduce bias. They came up with an easy-to-use and quick MLA model that can accurately predict OSA severity. For the purpose of identifying different phases of sleep using electrocardiogram data, another research [16] suggested a system that combined DT with KNN and RF algorithms. The adults' ISRUC-Sleep dataset, which is open to the public, was used. The dataset included two categories: healthy and sleep disorders. The Sleep Medicine Centre at the University Hospital of Coimbra randomly selected each recording from the PSG. The sleep variables were analyzed using statistical features. When it came to automatic sleep phases, the DT, KNN, and RF algorithms were on top. The RF algorithm outperformed the DT and KNN algorithms in terms of classification accuracy, with results over 90%.

For the early identification of people with high pretest OSA to recognize whether they have OSA or not, the authors [18] employed several MLAs, including extreme gradient boosting (XGB), light gradient boosting machine (LGBM), CB, RF, KNN, LR, and SVM. The suggested methods were tested using clinical data from the 1,479 entries in the Wisconsin Sleep Cohort database. Among these characteristics are physical measures, blood reports, and more. Model hyperparameters have been fine-tuned using Bayesian optimization and genetic algorithms, which indicated that routinely gathered clinical parameters might be used to overcome the constraints. With an F1-score of 75.96%, sensitivity of 88.76%, and specificity of 40.74%, the SVM method achieved a high accuracy of 73%.

To aid in the hunt for a novel cure for TB, the authors [21] devised an effective strategy that combined a genetic algorithm-based ensemble learning model with a heterogeneous feature representation to forecast antitubercular peptides. There are two separate anti-tubercular datasets including peptides (AtbPs) that were used to assess the suggested method. Outperforming competing algorithms, their suggested "iAtbP-Hyb-EnC" approach achieved 94.47% and 92.68% prediction accuracy, respectively.



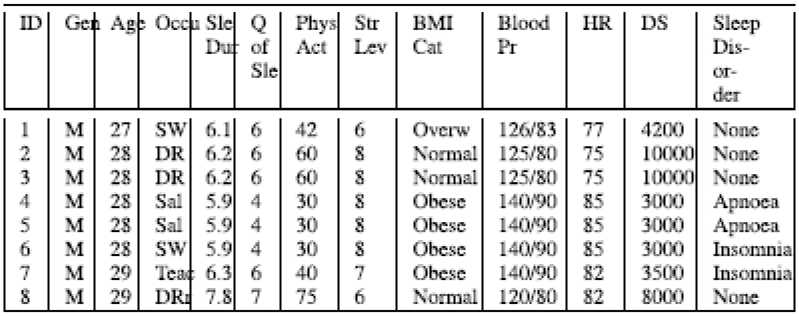
Ref.	Year	Algorithm Used	Accuracy	Dataset	Available	Real
[18]	2021	SVM, KNN, RF	SVM: 73%, KNN: 87%, RF: 89.3%	The Wisconsin Sleep Cohort dataset	Yes	Yes
[19]	2023	SVM, KNN, RF	SVM: 73%, KNN: 87%, RF: 89.3%	The Wisconsin Sleep Cohort dataset	Yes	Yes
[14]	2023	KNN, RF	KNN: 87%, RF: 89.3%	Medical Centre	No	Yes
[16]	2021	DT, KNN, RF	DT: 89.2%, KNN: 87%, RF: 89.3%	ISRUC Sleep database	Yes	Yes

III. METHODOLOGY

A. MATERIALS AND METHODS

The implementation of machine learning algorithms (MLAs) for the classification of sleep disorders is the main topic of this section. What follows is an explanation of the datasets, performance measures, and feature significance approach that will be used to evaluate the models and the suggested algorithms. Also included is a concise description of the classification method that was used for this study.Section B. The Actual Sleep Health and Lifestyle DatabaseDownloaded from the Kaggle website, the Sleep Health and Lifestyle Dataset was used in this work [22]. There are thirteen columns in the original dataset that include different kinds of data, and 400 observations overall. Every observation is a reflection of the real condition of sleep. The data may be grouped into thirteen categories that are pertinent to daily routines and sleep, including gender, age, profession, amount of sleep, and quality of sleep. The sleep problem for each participant is shown in Column 13. During the pre-processing stage, the labels "none," "sleep apnea," and "insomnia" were replaced with 1, 2, and 3, respectively, to better organize the data in this dataset into the three categories of sleep disorders. Here is an example of the dataset in Table 2.

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



ID	Gen	Age	Occu	Sle Dur	Q of Sle	Phys Act	Str Lev	BMI Cat	Blood Pr	HR	DS	Sleep Dis-or-der
1	M	27	SW	6.1	6	42	6	Overw	126/83	77	4200	None
2	M	28	DR	6.2	6	60	8	Normal	125/80	75	10000	None
3	M	28	DR	6.2	6	60	8	Normal	125/80	75	10000	None
4	M	28	Sal	5.9	4	30	8	Obese	140/90	85	3000	Apnoea
5	M	28	Sal	5.9	4	30	8	Obese	140/90	85	3000	Apnoea
6	M	28	SW	5.9	4	30	8	Obese	140/90	85	3000	Insomnia
7	M	29	Teac	6.3	6	40	7	Obese	140/90	82	3500	Insomnia
8	M	29	DR	7.8	7	75	6	Normal	120/80	82	8000	None

C. EXPERIMENT DESIGN

In order to categorize sleep disorders, this section suggests an evaluation methodology. Both techniques make up the methodology. The first method involves using 70% of the dataset as a training set and the remaining 30% as a test set to see how well the model performs. Without adjusting or optimizing the parameters, the model learns from the data and its performance is assessed using testing sets that are not visible to the model. Figure 1 exhibits the schematic of the ML model that was used for the purpose of sleep disorder classification.

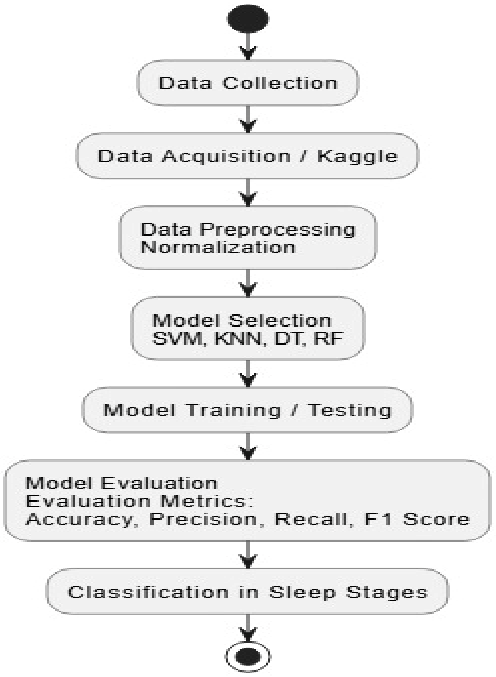
To fully understand the algorithms' strengths and areas for improvement, it is necessary to test them out without feature selection and optimization. This is the first stage in evaluating machine learning algorithms. In the alternative method, after getting the dataset ready, we fed each record into the models. Seventy percent of the dataset was used for training purposes. Using the optimization strategy, the suggested models were trained and evaluated. A fitness function was defined using the Genetic Algorithm (GA+MLAs) method, which combines GA and MLAs. Using GA, we applied feature selection to both the training and testing sets, and we found the best possible parameter settings.

In the training phase, the models were evaluated for their classification performance after learning from the data using the given features or parameters. Classification was then carried out using the trained GA with MLA. GA was used for feature selection in order to address the optimization issues with the classifiers. The classifiers may be fine-tuned by adjusting their various parameters. In order to get the most out of the suggested model, GA was used to fine-tune the parameters.

Figure 2 provides a high-level picture of how a genetic algorithm is put into action. This is how the suggested algorithm works:

- Step 1: Produce the starting population at random.
- Step 2: Find a fitness value that measures how well a collection of parameters (a potential solution) performs.
- Step 3: Choose parents who have more fit individuals to create healthy offspring.
- Step 4: Carry out the crossover process to produce new individuals (offspring).
- Step 5: Perform mutations to alter the DNA in an unpredictable way.
- Step 6: Repeat the process until the stop criteria are satisfied.

This method was used to perform feature selection and sleep disorder classification using MLAs.



D. PERFORMANCE METRICS

The suggested approach is tested and shown to work in this research for sleep problem categorization. Furthermore, the work-to-review ratio

because no two people are ever the same. As an example, sleep apnea might be responsible for a significant portion of the overall activity space. The majority class may achieve better accuracy using the classification accuracy measure, which is not suitable for this kind of dataset with imbalanced labels [23]. Take the accuracy measure as an example. It works well for balanced label classes but doesn't do much good for imbalanced ones. As a result, the F1-score, recall, precision, and classification accuracy were the four outcomes measured in this study [24]. The equations that follow define the mathematical formulations of these statistical indexes. The classification algorithms were tested using accuracy, which is defined as the proportion of correct predictions to the total number of predictions, as shown in (1), with TP standing for a true positive, TN for a true negative, FP for a false positive, and FN for a false negative:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (1)$$

The ratio of the number of anticipated false positives to the total number of expected TPs is called precision (2):

$$Precision = \frac{TP}{TP + FP} \qquad (2)$$

The percentage of true positives that were anticipated relative to the total number of true positives is known as recall (3):

$$Recall = \frac{TP}{TP + FN} \qquad (3)$$

A number's F1-score is the weighted average of its recall and accuracy. There are few FPs and FNs when the F1-score is ideal (4):

$$F1 = \frac{2 * TP}{2 * TP + FP + FN} \qquad (4)$$

E. CLASSIFICATION ALGORITHMS

1) SUPPORT VECTOR MACHINE

Classification and regression are two applications of support vector machines (SVMs), which are supervised learning algorithms [25]. The support vector machine (SVM) uses the class of hyperplanes, which is the line with the largest margin between two classes, to choose the optimum line to use as a decision boundary. As a measure of distance from the decision border to the nearest data points, the margin is measured vertically. In addition, SVM performs well when there are less samples than dataset dimensions. Furthermore, the model can train complex decision functions using SVMs and other kernel functions, including the RBF and linear functions [26]. #2: K-Nearest Neighbors If you want to know what a data point's nearest neighbor is worth, you may use the KNN, a nonparametric supervised learning method [25] for both classification and regression. Similarities between features form the basis of the KNN algorithms. The number of closest neighbor data points to include into the majority voting procedure is determined by the value of k, which is a process known as parameter tuning. Among the many distance metrics, you'll find the most common ones like the Minkowski, Manhattan, and Euclidean [27]. Thirdly, the Decision Tree When it comes to classification and regression issues, DTs are only one more nonparametric supervised learning technique to consider [25]. In terms of both comprehension and interpretation, the DT algorithms are on the simpler side. This means that the DT model uses the tagged data to learn basic state rules. You may use the DT with both numerical and categorical data. Even when presented with noisy data, the model manages to provide respectable results. Nevertheless, there are a few drawbacks to the DT. For instance, DT is vulnerable to overfitting, has issues with missing value handling, and may become unstable when exposed to small dataset modifications, which cause it to generate complicated trees that are ill-suited to handling new data [28]. 4) A Forest of Hope As an ensemble learning approach, RF classifiers generate several DTs at random and then combine them to enhance the accuracy of the model's predictions and control overfitting [25]. Two random processes are available to the model: bootstrapping and random feature selection. In order to make the model as resistant to conversions in the training data as possible, bootstrapping ensures that the model does not use the like data for each tree. By aggregating the trees and randomly selecting their features, we can lower their correlation [29].[31]. F. WHAT THIS FEATURE DOES One way to determine how important each input feature is before feeding it into the model is using the feature importance approach. The accuracy of the model is greatly affected by the maximum score of the features. Figure 3 shows that the model accuracy is greatly affected by the feature significance in this research, which includes the following features: body mass index (BMI), blood pressure, sleep duration, profession, and age.

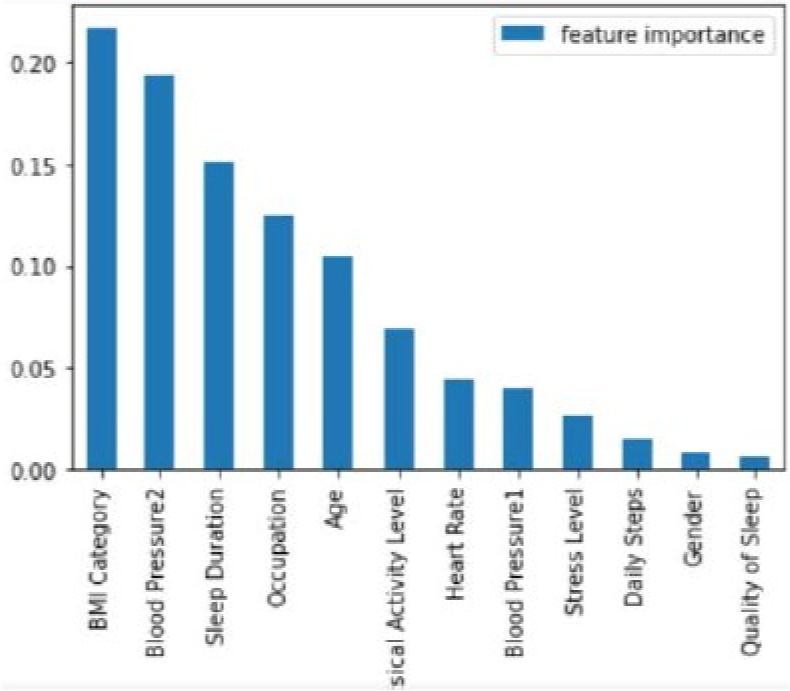


FIGURE 3. Feature importance.

G. CORRELATION COEFFICIENT

One statistical tool for examining the relationships between sleep-related and habit-related factors is the correlation coefficient. The value might be anything from -1 to 1. The coefficient of association between sleep length and sleep quality is greater than that of the other variables. Figure 4 summarizes the computed correlation between the characteristics. Section H: Python Graphics Export Figure 5 displays the DOT format, a text-based format that was achieved by exporting a DT using Python's graphviz. Section I: Genetic Algorithm An evolutionary algorithm, a genetic algorithm (GA) is an optimization method that draws inspiration from biology and the natural selection process. If an optimization issue has several possible solutions, GA may be used to modify the parameters and find the best one. As seen in Figure 6, the genetic algorithm adheres to a series of phases. The number 32.

IV. RESULTS AND DISCUSSION

Using MLAs for the proper classification of sleep disorders was shown in this research. The tests were carried out without using GA. The updated accuracy levels reached by the KNN, SVM, DT, and RF were **87%**, **73%**, **89.2%**, and **89.3%**, respectively. These results show strong performance across several algorithms. In order to find the best-performing kernel, this article examined training data from many SVMs and used the default setup for each classifier. While linear and polynomial kernels had the lowest accuracy, the RBF kernel achieved the bes performance using the SVM method.

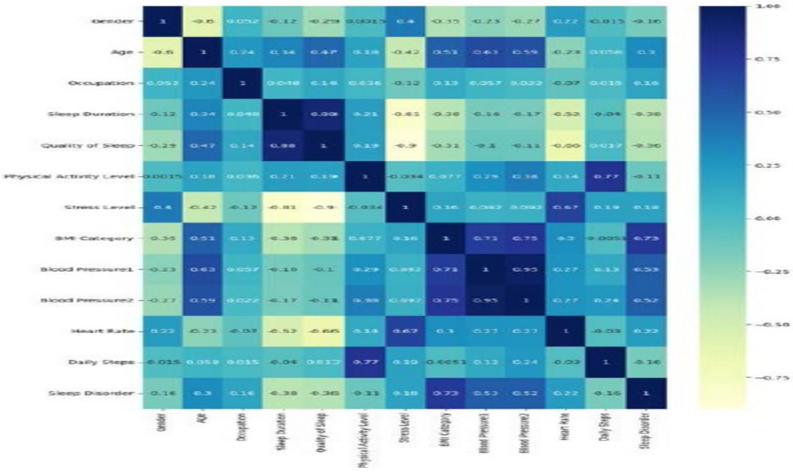


FIGURE 4. Correlation coefficient.

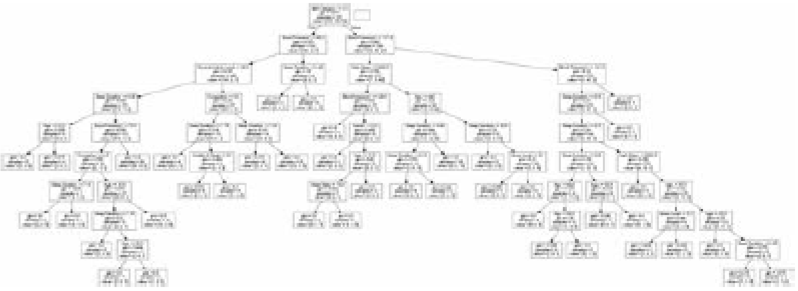


FIGURE 5. Export graphviz flow chart.

Nevertheless, determining the best parameter for each classifier is not an easy task, especially since there isn't yet an optimization technique that works well with MLAs on datasets with a lot of dimensions. A training and validation performance plot is shown in Figures 7–8. These plots were created to evaluate the value of the experiment. Although the points may show comparable loss curves, they may not be identical because of changes in the model's weights. This strategy provides a solid understanding of how learning performance evolves with the number of training iterations. It can also help determine whether adding additional training patterns improves validation scores or causes

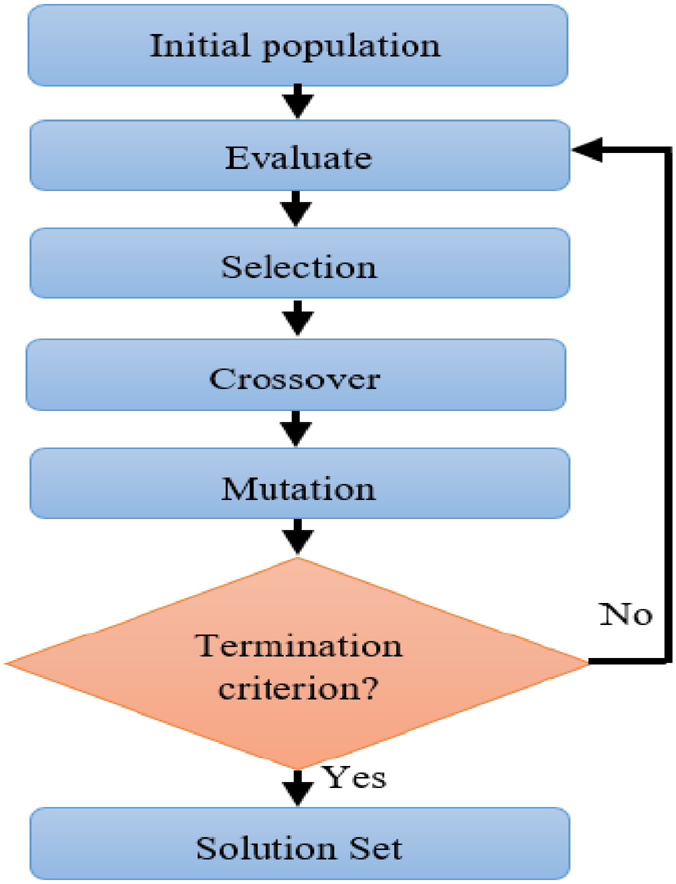
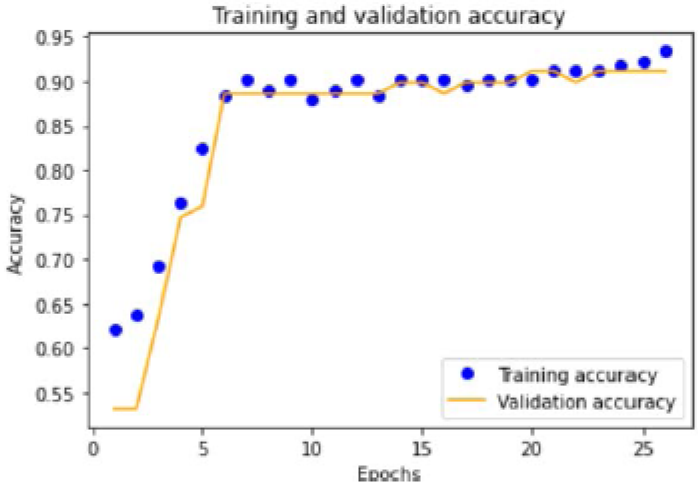


FIGURE 6. Basic architecture of the genetic algorithm [32].

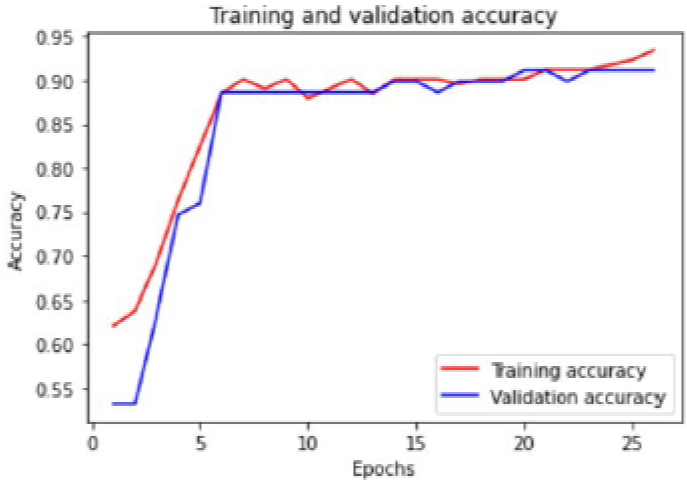
overfitting.

The results of the training phase for all the MLAs assessed are shown in Table 3, and the results obtained using 5-fold cross-validation are displayed in Table 4. Both Table 5 and Figure 9 summarize the testing phase outcomes for all MLAs using F1-score, recall, precision, and accuracy metrics. These results highlight how well the algorithms performed in comparison to each other.

Unfortunately, there isn't yet a universal optimization strategy that works effectively across all classifiers for high-dimensional datasets. To achieve the best results, each model's parameters must be carefully fine-tuned.



(a) Figure A



(b) Figure B

FIGURE 7. **Training and validation accuracy.**

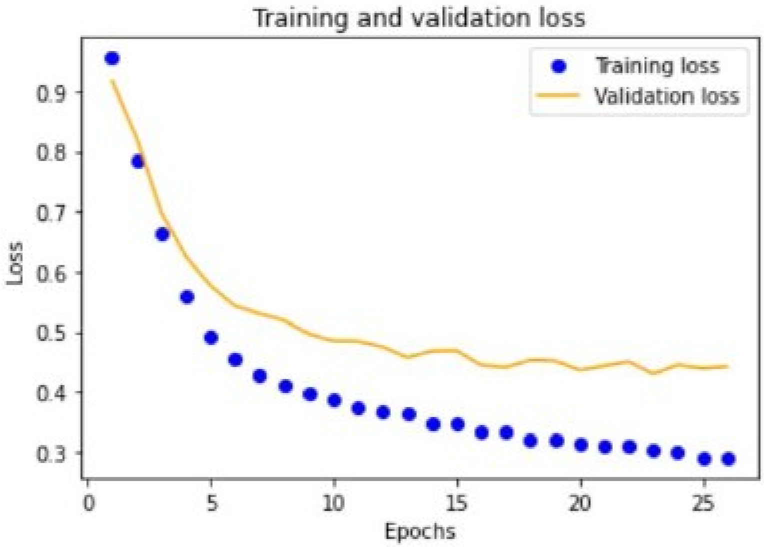


FIGURE 8. Training and validation loss.

Figure 10 and Table 6 show the outcomes of the experiment, where GA was used to find the best classifier parameters and search for an optimum one. The findings contrast the top-performing GA+MLA models with the top-performing MLA models. Another test is conducted using the t-test to demonstrate that there is a statistically significant difference in the best accuracy of the two models. All machine learning algorithms' test results for the dataset when measured against Precision, Recall,

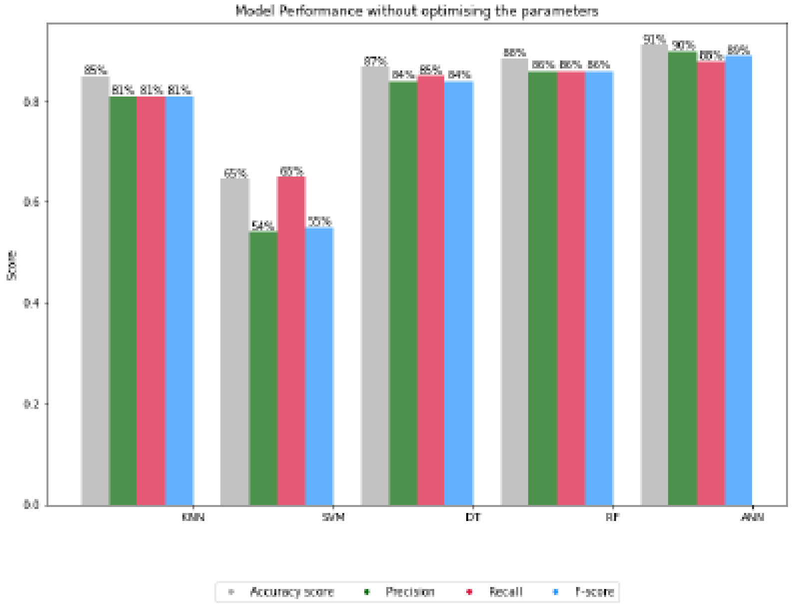


FIGURE 9. Results of the performance of all evaluated MLAs (As default parameters).

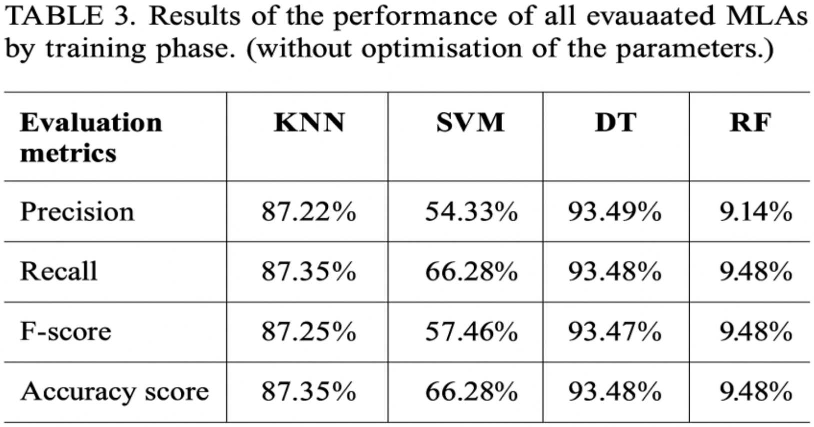


TABLE 3. Results of the performance of all evauaated MLAs by training phase. (without optimisation of the parameters.)

Evaluation metrics	KNN	SVM	DT	RF
Precision	87.22%	54.33%	93.49%	9.14%
Recall	87.35%	66.28%	93.48%	9.48%
F-score	87.25%	57.46%	93.47%	9.48%
Accuracy score	87.35%	66.28%	93.48%	9.48%

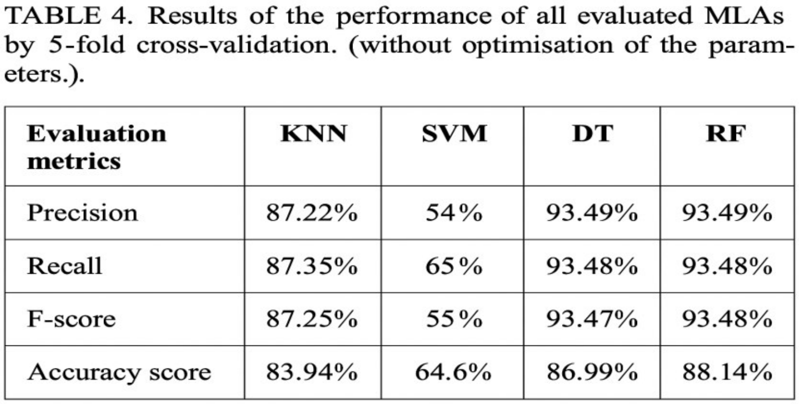


TABLE 4. Results of the performance of all evaluated MLAs by 5-fold cross-validation. (without optimisation of the parameters.).

Evaluation metrics	KNN	SVM	DT	RF
Precision	87.22%	54%	93.49%	93.49%
Recall	87.35%	65%	93.48%	93.48%
F-score	87.25%	55%	93.47%	93.48%
Accuracy score	83.94%	64.6%	86.99%	88.14%

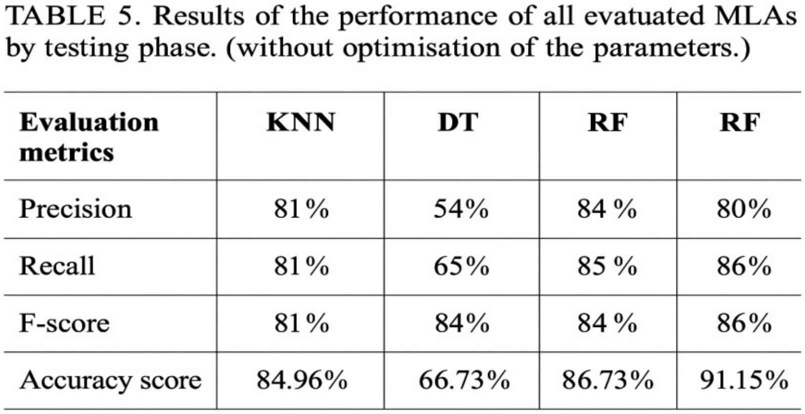


TABLE 5. Results of the performance of all evatuated MLAs by testing phase. (without optimisation of the parameters.)

Evaluation metrics	KNN	DT	RF	RF
Precision	81%	54%	84 %	80%
Recall	81%	65%	85 %	86%
F-score	81%	84%	84 %	86%
Accuracy score	84.96%	66.73%	86.73%	91.15%

Table 6 displays the t-test and F-Score. The results of the test reveal that while not all algorithms provide

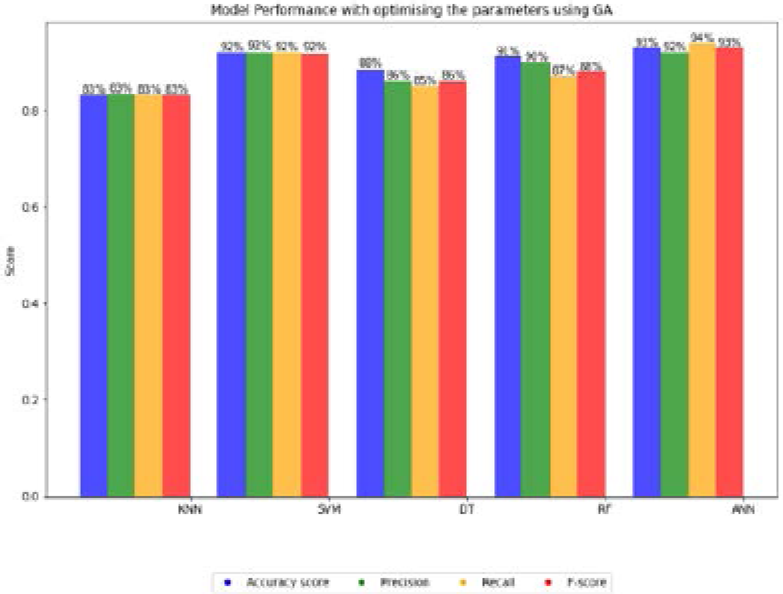


FIGURE 10. Results of the performance of all evaluated MLAs +GA (model performance with optimisation of the parameters using GA).

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

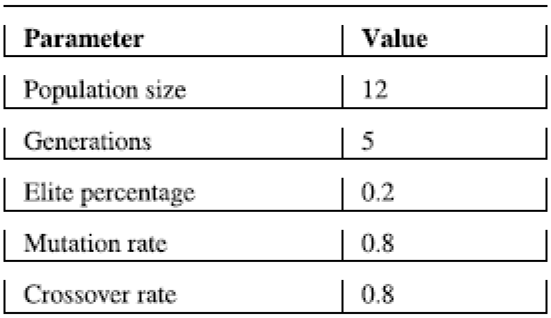
Evaluation metrics	**KNN**	**DT**	**RF**	**RF**
Precision	83.42%	92.1%	90.00%	92.00%
Recall	83.18%	92.0%	85.00%	87.00%
F-score	83.21%	91.8%	85.00%	87.00%
Accuracy score	83.19%	92.04%	88.50%	91.15%

Notable variations in results demonstrated that the suggested approach (GA+MLAs) outperformed MLAs with the default settings. The best values for the MLAs were determined using GA, which improved the classifiers' performance. Table 7 shows the results of running the GA with various parameter settings for five generations; Table 8 shows the results of using the fitness score to determine the optimal parameters and solutions. For instance, while optimizing the KNN model, the GA produced the best parameters, which were k = 2 and the metric for Euclidean distance. The KNN model was trained and tested on the complete dataset using these optimized parameters.

Classification accuracy after optimization was 87% for KNN, 73% for SVM, 89.2% for DT, and 89.3% for RF. Instead of using GA to optimize the SVM's hyperparameters, this paper also explored a grid search approach. With this method, training time can be reduced, results improved, and the best hyperparameter values for the SVM classifier identified by searching the hyperparameter space.

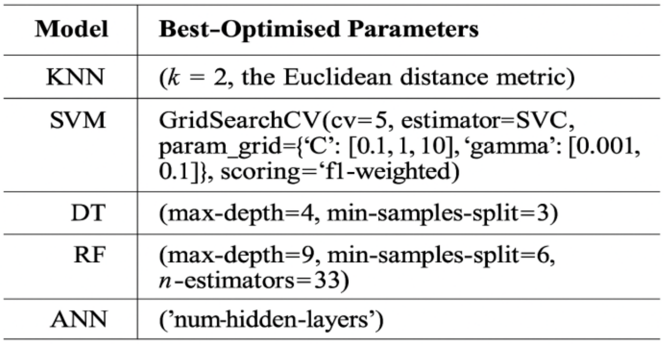
In addition, the Sleep Health and Lifestyle Dataset was utilized to compare this research to previous studies [33]. Outperforming the most recent research, the suggested technique achieved improved outcomes through parameter optimization using GA.

TABLE 7. parameter of the GA settings used.



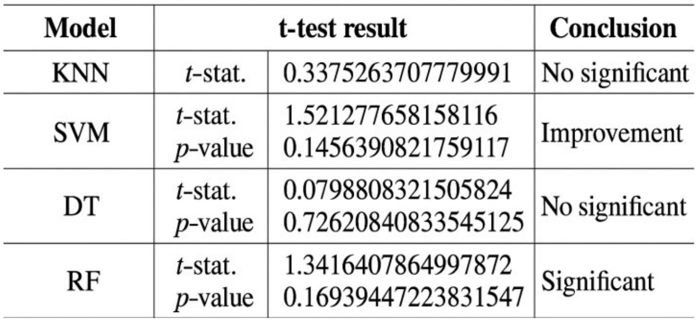
Parameter	Value
Population size	12
Generations	5
Elite percentage	0.2
Mutation rate	0.8
Crossover rate	0.8

TABLE 8. Best-optimised parameters of models.



Model	Best-Optimised Parameters
KNN	(k = 2, the Euclidean distance metric)
SVM	GridSearchCV(cv=5, estimator=SVC, param_grid={‘C’: [0.1, 1, 10], ‘gamma’: [0.001, 0.1]}, scoring=‘f1-weighted)
DT	(max-depth=4, min-samples-split=3)
RF	(max-depth=9, min-samples-split=6, n-estimators=33)
ANN	(’num-hidden-layers’)

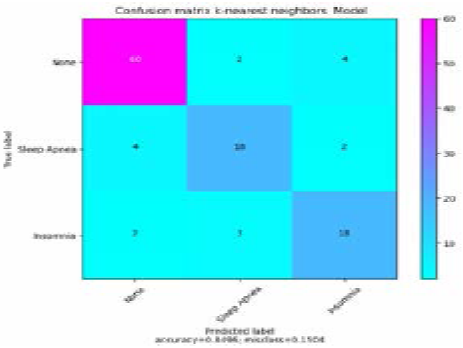
TABLE 9. The estimating of p values and t-tests.



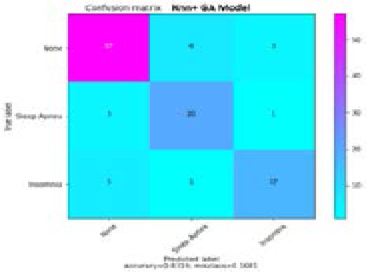
Model	t-test result		Conclusion
KNN	t-stat.	0.3375263707779991	No significant
SVM	t-stat. p-value	1.521277658158116 0.1456390821759117	Improvement
DT	t-stat. p-value	0.0798808321505824 0.72620840833545125	No significant
RF	t-stat. p-value	1.3416407864997872 0.16939447223831547	Significant

A. T-TEST ANALYSIS

Using a t-test for samples, we looked at how much the GA-optimized MLAs improved things. The average accuracy of some GA-optimized MLAs models deviates considerably from the baseline accuracy, according to the null hypothesis. Table 9 displays the results of the t-test, which showed that the GA-optimized MLAS classifiers dramatically improved accuracy.

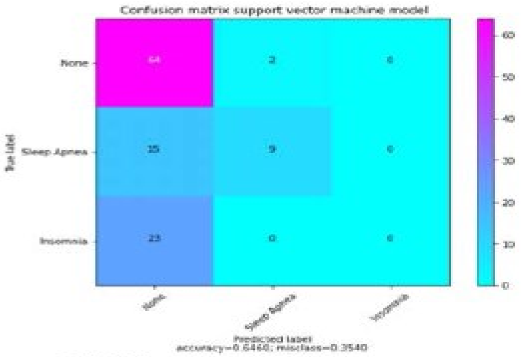


(a) KNN model with default parameters

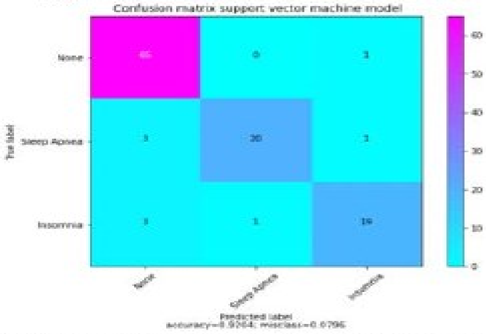


(b) KNN model with optimisation of the parameters

FIGURE 11. Confusion matrix for KNN model.



(a) SVM model with default parameters

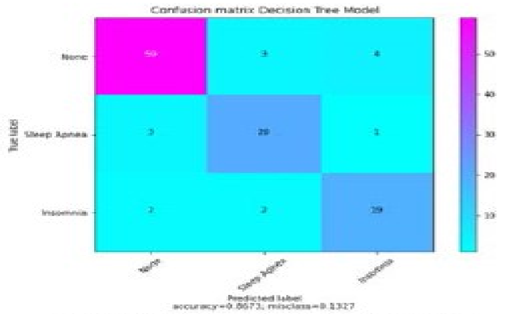


(b) SVM model with optimisation of the parameters

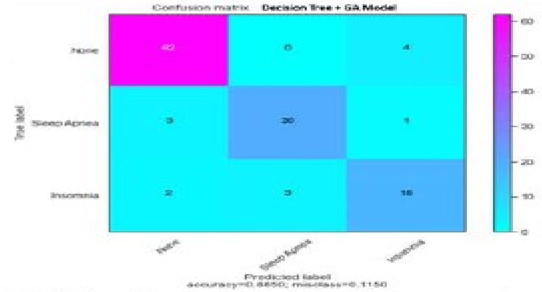
FIGURE 12. Confusion matrix for SVM model.

B. CONFUSION MATRIX

A confusion matrix, which summarizes the classification findings, was used to assess the MLAs classifiers' performance. You can see the confusion matrix for the in Figure 11-15.

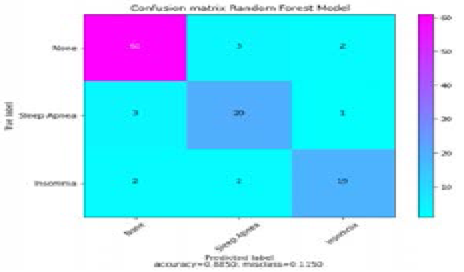


(a) DT model with default parameters

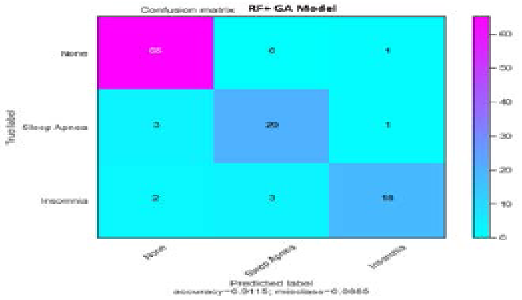


(b) DT model with optimisation of the parameters

FIGURE 13. Confusion matrix for DT model.



(a) RF model with default parameters



(b) RF model with optimisation of the parameters

FIGURE 14. Confusion matrix for RF model.

categorization job with several classes. By looking at the confusion matrix, we can see how many cases were categorized into each class and how well the model performed overall.

the misunderstandings. Important information about the RF classifier's performance on the multi-class classification job is revealed by the confusion matrix. By means of the chart. Class 1 was where the model really shone, with an impressive 96% accuracy rate. Class 3 had a misclassification rate of 26% while Class 2 had a rate of 20%, indicating inferior accuracy.

V. CONCLUSION

Using MLAs in conjunction with a genetic algorithm, this research proposes an improved model for sleep disorder categorization that investigates the best possible values for each model's hyperparameters. Using the real-world Sleep Health and Lifestyle Dataset, this article analyzed the efficacy of several state-of-the-art MLAs for sleep disorder classification. Also, MLAs may try to diagnose sleep disorders from high-dimensional sleep data without relying on criteria established by experts. Regardless of the data constraint. Implementing MLAs for sleep disorder categorization is not without its difficulties, which were explored in this work. To train and evaluate models in this domain, however, huge datasets are still required. When it comes to sleep disorder categorization, MLAs with GA may make a world of difference. In addition to training MLAs using unsupervised learning, future work will include testing a new model on the dataset and comparing its results to those of current top-tier models.

REFERENCES

[1] F. Mendonça, S. S. Mostafa, F. Morgado-Dias, and A. G. Ravelo-García,''A portable wireless device for cyclic alternating pattern estimationfrom an EEG monopolar derivation,'' *Entropy*, vol. 21, no. 12, p. 1203,Dec. 2019.
[2] Y. Li, C. Peng, Y. Zhang, Y. Zhang, and B. Lo, ''Adversarial learningfor semi-supervised pediatric sleep staging with single-EEG channel,''*Methods*, vol. 204, pp. 84–91, Aug. 2022.
[3] E. Alickovic and A. Subasi, ''Ensemble SVM method for automaticsleep 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 tointerpret the results of a sleep study,'' *J. Community Hospital Internal Med.Perspect.*, vol. 4, no. 5, p. 24983, Jan. 2014.
[5] V. Singh,V. K. Asari, and R. Rajasekaran, ''Adeep neural network for earlydetection and prediction of chronic kidney disease,'' *Diagnostics*, vol. 12,no. 1, p. 116, Jan. 2022.
[6] J. Van Der Donckt, J. Van Der Donckt, E. Deprost, N. Vandenbussche,M. Rademaker, G. Vandewiele, and S. Van Hoecke, ''Do not sleep ontraditional machine learning: Simple and interpretable techniques arecompetitive to deep learning for sleep scoring,'' *Biomed. Signal Process.Control*, vol. 81, Mar. 2023, Art. no. 104429.
[7] H. O. Ilhan, ''Sleep stage classification via ensemble and conventionalmachine 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 reinforcementlearning-based optimization framework for EEG-based state evaluation,'*J. Neural Eng.*, vol. 18, no. 4, Aug. 2021, Art. no. 046059.
[9] Y. J. Kim, J. S. Jeon, S.-E. Cho, K. G. Kim, and S.-G. Kang, ''Predictionmodels for obstructive sleep apnea inKorean adults using machine learningtechniques,'' *Diagnostics*, vol. 11, no. 4, p. 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 fromsingle-channel EEG signals,'' *J. Neurosci. Methods*, vol. 324, Aug. 2019,Art. no. 108312.
[11] S. Djanian, A. Bruun, and T. D. Nielsen, ''Sleep classification usingconsumer 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, and A. Ahmadi, ''Detection of sleep apnea using machinelearning algorithms based on ECG signals: A comprehensive systematicreview,'' *Expert Syst. Appl.*, vol. 187, Jan. 2022, Art. no. 115950.
[13] C. Li, Y. Qi, X. Ding, J. Zhao, T. Sang, and M. Lee, ''A deeplearning method approach for sleep stage classification with EEGspectrogram,'' *Int. J. Environ. Res. Public Health*, vol. 19, no. 10, p. 6322,May 2022.
[14] H. Han and J. Oh, ''Application of various machine learning techniquesto predict obstructive sleep apnea syndrome severity,'' *Sci. Rep.*, vol. 13,no. 1, p. 6379, Apr. 2023.
[15] M. Bahrami and M. Forouzanfar, ''Detection of sleep apnea from singleleadECG: Comparison of deep learning algorithms,'' in *Proc. IEEE Int.Symp. Med. Meas. Appl. (MeMeA)*, Jun. 2021, pp. 1–5.
[16] S. Satapathy, D. Loganathan, H. K.Kondaveeti, and R. Rath, ''Performanceanalysis of machine learning algorithms on automated sleep stagingfeature 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-leadECG: A comprehensive analysis of machine learning and deep learningalgorithms,'' *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.
[18] J. Ramesh, N. Keeran, A. Sagahyroon, and F. Aloul, ''Towards validatingthe effectiveness of obstructive sleep apnea classification from electronichealth records using machine learning,'' *Healthcare*, vol. 9, no. 11, p. 1450,Oct. 2021.

[19] Agrawal, K. K. ., P. . Sharma, G. . Kaur, S. . Keswani, R. . Rambabu, S. K. . Behra, K. . Tolani, and N. S. . Bhati. "Deep Learning-Enabled Image Segmentation for Precise Retinopathy Diagnosis". *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 12s, Jan. 2024, pp. 567-74, https://ijisae.org/index.php/IJISAE/article/view/4541.

[20] Samota, H. ., Sharma, S. ., Khan, H. ., Malathy, M. ., Singh, G. ., Surjeet, S. and Rambabu, R. . (2024) "A Novel Approach to Predicting Personality Behaviour from Social Media Data Using Deep Learning", *International Journal of Intelligent Systems and Applications in Engineering*, 12(15s), pp. 539–547. Available at: https://ijisae.org/index.php/IJISAE/article/view/4788

[21] S. K. Satapathy, H. K. Kondaveeti, S. R. Sreeja, H. Madhani, N. Rajput,and D. Swain, ''A deep learning approach to automated sleep stagesclassification using multi-modal signals,'' *Proc. Comput. Sci.*, vol. 218,pp. 867–876, Jan. 2023.

[22] O. Yildirim, U. Baloglu, and U. Acharya, ''A deep learning model forautomated sleep stages classification using PSG signals,'' *Int. J. Environ.Res. Public Health*, vol. 16, no. 4, p. 599, Feb. 2019.