ISSN NO: 9726-001X

Volume 13 Issue 02 2025

Smart Diagnosis of Sleep Disorders using Machine Learning Techniques

[1] Reddy Abhisheku, [2] Vegesna Sai Sindhu, [3] Vendra Sri Suryanarayana, [4] Veligatla Hemanth Sai, [5] Ms. K. Siva Syamala,

[1,2,3,4,] Students, Dept. of CSE, DNR College of Engineering & Technology, Balusumudi, Bhimavaram, India.

[5] Assistant Professor, Dept. of CSE, DNR College of Engineering & Technology, Balusumudi, Bhimavaram, India.

ABSTRACT

Classifying sleep disorders is essential for bettering people's lives. The effects of sleep apnea and other similar conditions on human health can be substantial. Experts in the area have a difficult and error-prone duty when it comes to classifying sleep stages. Analyzing, monitoring, and diagnosing sleep disorders is necessary for developing effective machine learning algorithms (MLAs) for sleep disorder categorization. In order to categorize sleep disorders, this research contrasts traditional MLAs with deep learning methods. This work presents a strategy for the optimal classification of sleep disorders and tests it using the publically accessible Sleep Health and Lifestyle Dataset. The optimizations were carried out by adjusting the settings of several machine learning algorithms using a genetic algorithm. A review and comparison of the suggested approach with cutting-edge ML algorithms for sleep disease classification. Various variables describing sleep and everyday activities are included in the 400 rows and 13 columns of the dataset. We tested a variety of deep learning methods, including k-nearest neighbours, SVM, decision tree, random forest, and ANN. There are noticeable performance disparities among the algorithms that were tested, according to the experimental data. The classification accuracy that each of the suggested methods achieved was 83.19%, 92.04%, 88.50%, 91.15%, and 92.92%, correspondingly. With a recall of 91.93%, a precision of 92.01%, and an F1-score of 93.80% on the test data, the ANN attained the best classification accuracy of 92.92%. The artificial neural network (ANN) algorithm that outperformed the others in terms of accuracy.

The following terms are included in the index: genetic algorithm, sleep disorder, classification, deep learning, and machine learning algorithms.

I. INTRODUCTION

In order to maintain good mental and physical health, sleep is an essential physiological process. A good night's sleep fortifies the body and helps memories and cognition stick. Children and elderly drivers are more vulnerable to accidents due to poor sleep quality, which impacts cognitive processes. The human body is susceptible to the negative effects of sleep deprivation, which can lead to conditions such as diabetes, obesity, and heart disease. The manual evaluation of polysomnography (PSG) recordings by doctors, specialists, and other medical personnel might result in varying sleep stage judgments. Classifying sleep stages manually is laborious, error-prone, and time-consuming. (1) and (2). Philips surveys people's thoughts and actions around sleep every year on World Sleep Day. More than 13,000 individuals from 13 different countries were surveyed in 2021. Out of all persons surveyed, just 55% reported being happy with the amount and quality of their sleep. Issues with sleep apnea, insomnia, and the 2019 coronavirus illness pandemic all contributed to poor sleep quality. Among those who reported trouble sleeping, 37% stated it was because of the epidemic. Additionally, 37% of individuals reported having trouble sleeping, 29% snore, 22% suffer from a sleep condition related to shift employment, and 12% suffer from sleep apnea. (1) and (2). Sleep specialists and medical professionals assess the quality of sleep by analyzing the sleep system, which is categorized for different stages 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 conscious of their immediate environment, they are said to be awake and alert. During a state of awareness, brain waves are rapid and erratic. The first stage of sleep, N1, is characterized by sluggish brain waves and muscular relaxation. Also, while people are sleeping, they are in a deep stage of sleep called N2, and the deepest stage of sleep, N3, is even harder to rouse them from. Rapid eye movement (REM) sleep is characterized by brain waves that are comparable to those experienced while awake. Different processes rely on each stage of the sleep cycle. While we sleep, our brains and bodies continue to work at a high level. Consequently, PSG allows medical professionals to monitor the patient's physiological status by recording electroencephalogram (EEG) and electrocardiogram (ECG) data [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. There are two main categories into which these methods fall: deep learning algorithms and regular machine learning algorithms. The training dataset for traditional MLAs is quite modest, yet they are simple, rapid, and easy to deploy. 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. New machine learning algorithms (MLAs) with a biological basis aim to use neural networks to learn complex patterns from data in an effort to imitate the human brain. It is quite probable that deep learning algorithms will supersede more conventional forms of artificial intelligence. Any algorithm that uses layers to handle data is considered deep learning, and feature engineering is an automated process [6, 7]. Classification jobs with large amounts of data or complicated characteristics are ideal for deep learning models. Using an electroencephalogram (EEG) as input is the most used method for sleep-stage categorization [8]. 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. Data bias towards particular patient groups makes it hard to generalise evaluated outcomes, and erroneous results based on biased data might 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]. Aiming to alleviate the difficulties that persons with sleep problems face in today's fast-paced world is what prompted this research. The risks connected with the rise in sleep disorders are already significant, and they are only going to get worse as a result of people's disregard of this vital requirement and the impact of contemporary lives. One of the most fundamental need for human survival is sleep. Ensuring human well-being and quality of life requires the implementation of machine learning algorithms to categorize sleep disorders. The authors are unaware of any thorough reviews of MLAs used for sleep disorder categorization, despite their use in this area. There are two main points that this article brings up: 1) a survey of the literature on sleep disorder classification; 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; and 3) a thorough review of conventional MLAs trained with deep learning algorithms. Following is the outline of the paper. This paper's evaluation technique and the state-of-the-art MLAs are detailed in Section III, while Section II covers the relevant work. The approaches and their performance in sleep disorder categorization are discussed and shown in Section IV. The article is concluded in Section V, which also covers the anticipated future development for this application.

II. RELATED WORK

In [11], the researchers looked examined a number of studies that classified sleep using consumer sleep technology (CST) and MLAs. 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. An assortment of multi-level analysis (MLA) techniques, including logistic regression (LR), decision trees (DT), support vector machines (SVM), and deep learning, were examined in 27 publications included in the article. The accuracy of sleep-stage categorization using CST might be greatly enhanced by the models. But using raw inputs with deep learning algorithms is still a bit of a challenge. The importance and difficulties of sleep apnea were highlighted in another work [12], which analyzed 48 articles. Furthermore, MALAs like support vector machines (SVMs), random forests (RFs), and deep learning algorithms can be employed to identify cases of sleep apnea from electrocardiogram (ECG) data. Nevertheless, they did point out a few problems with using MLAs to classify sleep apnea, such as the fact that ECG signals are different and that there aren't enough datasets to train the models. The study found that neural networks based on deep learning and support vector machines were the most effective in identifying sleep apnea from electrocardiogram (ECG) data. In order to categorize the sleep state from an EEG spectrogram, the authors in [13] employed MLAs. There is a time penalty associated

ISSN NO: 9726-001X

Volume 13 Issue 02 2025

with sleep stage categorization. In order to classify EEG data, it use MLAs, which are inherently inaccurate. The data are also imbalanced, which contributes to the poor accuracy. For this purpose, they tested their models on four publicly available datasets. Across all four datasets, the suggested algorithms achieved classification accuracies of 94.17%, 86.18%, 83.10%, and 85.12%, respectively. To create a deep learning model for sleep stage classification, they employed deep learning techniques. The EEG spectrogram was processed using convolutional neural networks (CNNs) to extract time and frequency information. A key technique for identifying different stages of sleep using electroencephalograms (EEG) spectrograms, the model incorporates several hidden layers of bidirectional long short-term memory (LSTM) to identify prediction sequences.

In order to forecast the severity of obstructive sleep apnea (OSA) disease, researchers in [14] employed MLAs using real data obtained from 4,014 patients—data that is not publicly accessible. Use of supervised and unsupervised learning methods, including RF, K-means, and gradient boosting, was carried out by the writers. A respectable 88%, 88%, and 91% classification accuracy was achieved using their suggested approaches. Nevertheless, there are a number of limitations to their study. There are some missing numbers and the data comes from a single center, which might make it biased. They came up with an MLA model that is easy to use and takes little time to accurately predict the severity of OSA.

In order to identify sleep apnea from a single-lead electrocardiogram (ECG), the developers in [15] employed convolutional neural networks (CNNs), long short-term memories (LSTMs), bidirectional LSTMs, and gated recurrent units. In order to test the suggested methods, which incorporate a total of 70 records, the apnoea-ECG dataset was utilized. Their proposed hybrid models acquired a classification accuracy of 80.67%, 75.04%, 84.13% and 74.72%. The CNN obtained a higher accuracy than the other algorithms, and the findings revealed that the best-performing algorithm was the hybrid CNN and LSTM network. After comparing the results of several deep learning algorithms, they concluded that, unlike traditional MLAs, deep learning algorithms can learn to diagnose sleep apnea automatically. Another study [16] suggested a system that combined DT, knearest neighbours (KNN) and RF algorithms to classify sleep phases from the ECG. They utilized the publicly available ISRUC?Sleep dataset obtained from people, which included two states: healthy and sleep disorders. The Sleep Medicine Centre at the University Hospital of Coimbra randomly selected each recording from the PSG. In their analysis of the sleep characteristics, they made use of statistical features. In terms of automatic sleep phases, the top performing algorithms were DT, KNN, and RF. Compared to DT and KNN, the RF algorithm achieved a classification accuracy higher than 90%. Additionally, a model was suggested by researchers [16] for sleep apnea diagnosis utilizing a single-lead ECG that utilized traditional MLAs such DT, KNN, RF, and deep algorithms. Using 70 recordings from the PhysioNet ECG Sleep Apnea v1.0.0 dataset, the authors employed deep recurrent neural networks (DRRNs) to record the data's temporal pattern and hybrid convolutional-recurrent convolutional neural networks (CNNs) to extract features. In order to lower the dimensions, they used principal component analysis. Compared to the other methods, the hybrid CNN-DRNN architecture had superior accuracy detection. Their recommendation for detecting sleep apnea from electrocardiograms was a hybrid deep neural network. Additionally, a model was suggested by researchers [17] for sleep apnea diagnosis utilizing a single-lead ECG that utilized traditional MLAs such DT, KNN, RF, and deep algorithms. Using 70 recordings from the PhysioNet ECG Sleep Apnea v1.0.0 dataset, the authors employed deep recurrent neural networks (DRRNs) to record the data's temporal pattern and hybrid convolutional-recurrent convolutional neural networks (CNNs) to extract features. In order to lower the dimensions, they used principal component analysis. Compared to the other methods, the hybrid CNN-DRNN architecture had superior accuracy detection. Their recommendation for detecting sleep apnea from electrocardiograms was a hybrid deep neural network. For the early identification of individuals 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. In order to test the suggested algorithms, they consulted the Wisconsin Sleep Cohort database, which contains 1,479 records of clinical data. Among these characteristics are physical measures, blood reports, and more. In order to optimize the model hyperparameters, evolutionary algorithms and Bayesian optimization have been applied. These algorithms have proposed that routinely gathered clinical parameters might be utilized to overcome these restrictions. A high accuracy of 68.06%, sensitivity of 88.76%, specificity of 40.74%, and an F1-score of 75.96% were achieved using the SVM method. In order to automate the process of sleep-stage categorization based on multimodal inputs, other researchers [19] suggested a model that combines traditional machine learning with a deep

ISSN NO: 9726-001X

Volume 13 Issue 02 2025

JOURNAL OF CURRENT SCIENCE

learning strategy. Using a combination of convolutional neural network (CNN) and long short-term memory (LSTM) techniques, the researchers in this study were able to predict the temporal dynamics of EEG data and use CNN to extract unique characteristics. Public databases (sleep-edf) were used to assess the suggested system. Outperforming the other algorithms, the CNN/LSTM combination reached an accuracy of 87.4 percent. The patient data records also contained background noise. But to clean up the data, they utilized the Butterworth filter. In order to automatically categorize the various phases of sleep from raw PSG data, another study [20] suggested a method that makes use of a deep learning model. A basic convolutional neural network (CNN) is used to extract features by the model. They used two freely available online datasets, sleep-edf and sleep-edfx, to assess the suggested model. For two to six sleep classes, the suggested model achieved great accuracy with 98.06%, 94.64%, 92.36%, 91.22%, and 91.00%, respectively. To avoid the need for human specialists, the authors propose deep learning as an alternative to standard approaches for automated sleep-stage categorization. This method is vulnerable to mistakes made by humans. Table 1 provides an overview of the algorithm, dataset, and accuracy in a few of the papers that were examined. In their effective approach to predicting antitubercular peptides, the authors [21] combined a genetic algorithm-based ensemble learning model with a heterogeneous feature representation, which should aid in the hunt for a novel medication to combat TB. There are two separate anti-tubercular

TABLE 1. A summary of the algorithm, dataset 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.

ISSN NO: 9726-001X

Volume 13 Issue 02 2025

[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.

datasets including peptides (AtbPs) were utilized to assess the suggested method. Outperforming competing algorithms, their suggested "iAtbP-Hyb-EnC" approach achieved 94.47% and 92.68% prediction accuracy, respectively.

III. METHODOLOGY

A. MATERIALS AND METHODS

In this part, we will go over some methods for classifying sleep disorders using deep learning algorithms as well as traditional MLAs. 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. Furthermore, a concise explanation of the categorization algorithm utilized in this study is provided. B. The Lively Sleep Health and Lifestyle Research Pool Downloaded from the Kaggle website, the Sleep Health and Lifestyle Dataset was utilized in this work [22]. There are thirteen columns in the original dataset that include different kinds of data, and 400 observations overall. The real sleep state is represented by each observation. Gender, age, profession, number of hours slept, and quality of sleep are a few of the thirteen sleep-related characteristics that may be extracted from this data set. The sleep problem for each participant is displayed 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. The dataset is illustrated 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	DRr	7.8	7	75	6	Normal	120/80	82	8000	None

C. EXPERIMENT DESIGN

An evaluation procedure for the categorization of sleep disorders is presented in this section. Two methods make up the technique. In the first method, 70% of the dataset serves as a testing set, while the remaining 70% is used for model learning without performance scoring. As the model learns from the data without adjusting and optimizing the parameters, its performance is assessed on unknown testing sets. Figure 1 shows the schematic of the ML model that was employed for the classification of sleep disorders. The purpose of this stage is to assess the capabilities and shortcomings of machine learning algorithms in their whole, apart from feature selection and optimization. In the second 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. The implementation of the Genetic Algorithm (GA+MLAs) method defined a fitness function that integrates the 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 a 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. In order to fix the optimization issues with the classifiers, GA was employed for feature selection. 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 is a high-level depiction of how a genetic algorithm is put into action. This is how the suggested algorithm works: The first step is to produce the starting population at random. Step 2: Find a fitness value that measures how well a collection of parameters (a potential solution) performs. The next step is to

ISSN NO: 9726-001X

Volume 13 Issue 02 2025

choose parents that have healthy, fit individuals. In the fourth step, known as "crossover," two parents are combined to produce a new person. Fifth, alter the genetic code at random by performing mutation. Step 6: Keep going until the end requirements are satisfied. to perform feature selection and sleep disorder classification using MLAs.

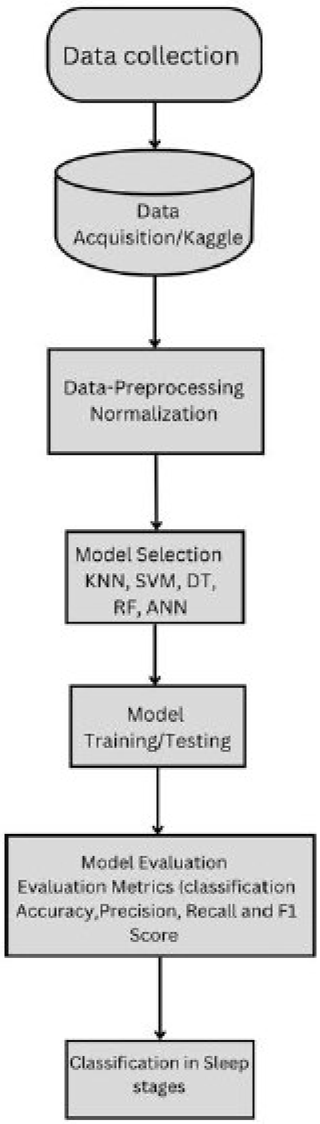


FIGURE 1. Diagram of the machine learning model to classify disorders.

D. PERFORMANCE METRICS

Within the context of sleep disorder categorization, this study assesses and confirms the efficacy of the suggested model. Furthermore, the work-to-review ratio

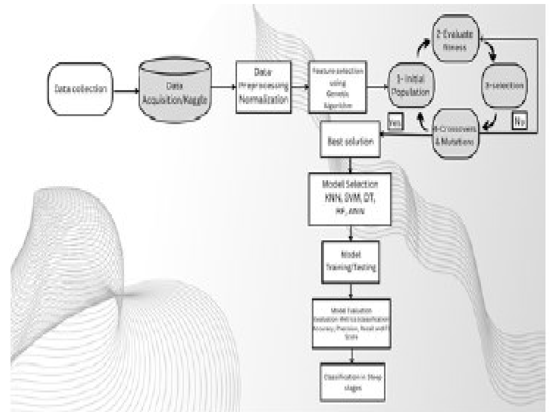


FIGURE 2. The proposed optimised model for sleep disorder classification.

in most cases, not everyone is the same. As an example, a large portion of the overall activity space might be attributed to sleep apnea. Due to the imbalanced nature of the labels in this dataset, the classification accuracy metric is not applicable, and the majority class can get better results [23]. Take the accuracy measure as an example. It works great for balanced label classes but doesn't do much good for imbalanced ones. Classification accuracy, recall, precision, and the F1-score were the four metrics employed for assessment in this research [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:

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

Precision is the ratio of the number of predicted TPs to the total number of predicted positives (2):

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

Recall is the ratio of the number of predicted TPs to the total number of actual TPs (3):

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

The F1-score provides a weighted average for the precision and recall of a number. A perfect F1-score provides low FPs and low FNs (4):

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

ISSN NO: 9726-001X

Volume 13 Issue 02 2025

E. CLASSIFICATION ALGORITHMS
1) SUPPORT VECTOR MACHINE

Classification and regression are two applications of support vector machines (SVMs), which are supervised learning algorithms [25]. Based on the class of hyperplanes, which is the line with the biggest margin between two classes, the SVM seeks to produce the optimal line, termed a decision boundary. The decision border and the nearest data points are vertically separated, and this distance is called the margin. In addition, SVM performs well when there are less samples than dataset dimensions. It is also possible to train the model to understand complex decision functions using SVMs in conjunction with other kernel functions, such as linear functions and the RBF [26]. Secondly, K-Nearest Neighbors Classification and regression [25] are two applications for the KNN, a nonparametric supervised learning technique that assigns a label to a data point based on its nearest neighbour. Feature similarity is the foundation of KNN algorithms. Parameter tuning determines the value of k, which is the number of neighboring data points to include in the majority voting procedure. Many other kinds of distance metrics exist, including Minkowski, Manhattan, and the one developed by Euclidean geometry [27]. thirdly, the decision tree For regression and classification issues, another nonparametric supervised learning technique is a DT [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. Both numerical and categorical data may be used with the DT. Even when presented with noisy data, the model manages to produce 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 Fuzzy Forest 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]. Bootstrapping and random feature selection are two random techniques that the model can employ. 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 randomly selecting features, we may aggregate the trees while reducing their correlation [29]. 5) A MACHINE THAT LABS WITH THE HINDS One kind of supervised learning algorithm that attempts to simulate brain function is the artificial neural network, or ANN. It is a network of synthetic neurons that communicate with one another. Between the input and output layers of an ANN are several hidden layers. A neural weight is included in every entry. Every neuron in the first layer receives an input, and each layer is fully connected to the next and given a weight. A threshold function passes a weighted sum on to the activation function. The activation function's output controls whether a neuron is active. In a process known as feed-forward propagation, the activated neuron is sent to the output neuron of the subsequent layer [30], [31]. Chapter F: The Significance of the Boundary One way to determine how important each input feature is before feeding it into the model is using the feature importance approach. When it comes to model correctness, the greatest score of features is king. Feature significance greatly affects model accuracy in this work, which includes data such as body mass index (BMI), blood pressure, sleep duration, profession, and age (Figure 3).

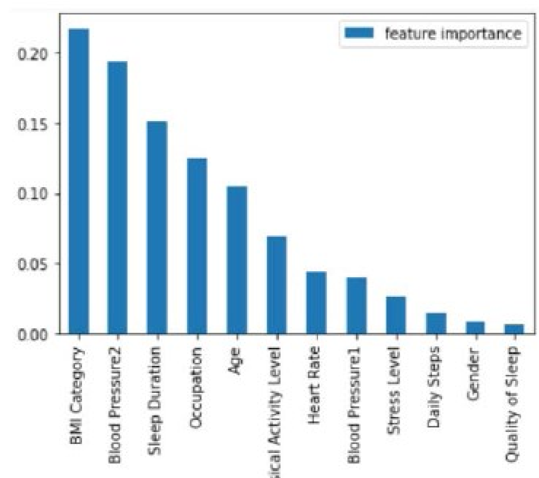


FIGURE 3. Feature importance.

G. CORRELATION COEFFICIENT

A statistical tool that reveals the associations between factors that are related to sleep and daily behaviors is the correlation coefficient. It ranges from minus one to plus one. When compared to the other variables, the correlation coefficient between sleep length and sleep quality is the highest. Figure 4 summarizes the computed feature-to-feature correlation. Chapter H: Python Graphics Export Figure 5 shows the output of an export of a DT to the text-based DOT format created using the Python program export graphviz. Genomics Algorithm I An

ISSN NO: 9726-001X

Volume 13 Issue 02 2025

JOURNAL OF CURRENT SCIENCE

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

The results of this study show that MLAs are useful for the correct diagnosis of sleep problems. The trials were performed

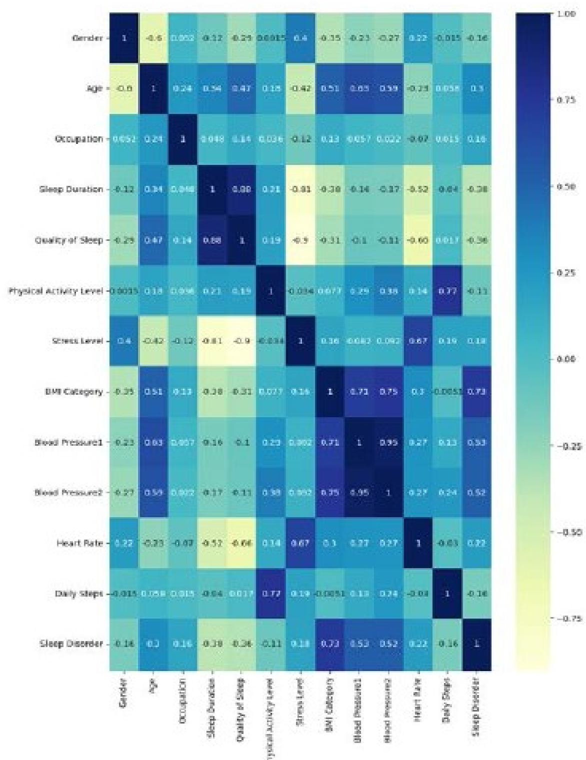


FIGURE 4. Correlation coefficient.

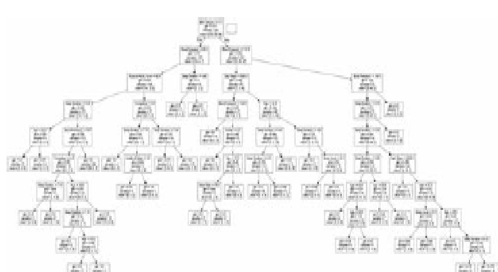


FIGURE 5. Export graphviz flow chart.

that do not employ GA. The accuracy levels achieved by the KNN, SVM, DT, RF, and ANN classifiers were 84.96%, 64.6%, 86.73%, 88.5%, and 91.15%, respectively. The best algorithms were able to reach a level of accuracy greater than 90%. 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 yielded the lowest accuracy, the RBF kernel achieved good performance with the SVM algorithm. Nevertheless, determining the best parameter for each classifier is not an easy task. Reason being, there isn't currently an optimisation method that works well with MLAs on datasets with a lot of dimensions. You can see a performance plot of the training and validation processes in Figures 7-8. This data was produced while trying to categorize

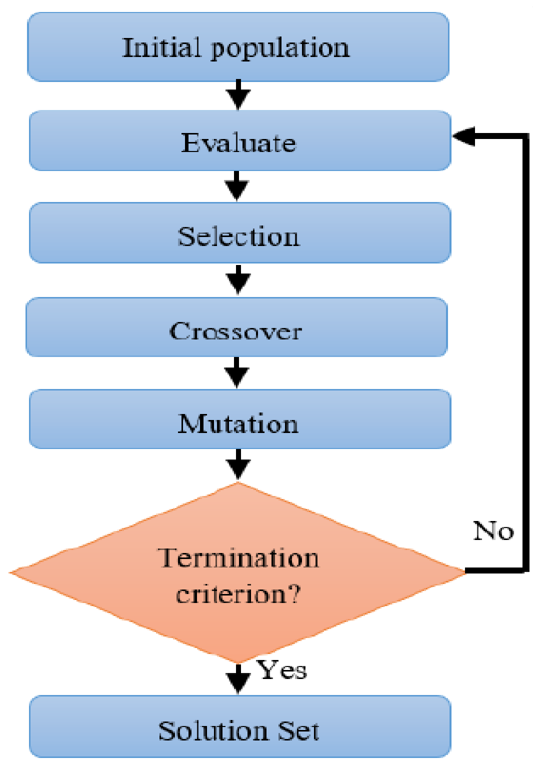


FIGURE 6. Basic architecture of the genetic algorithm [32].

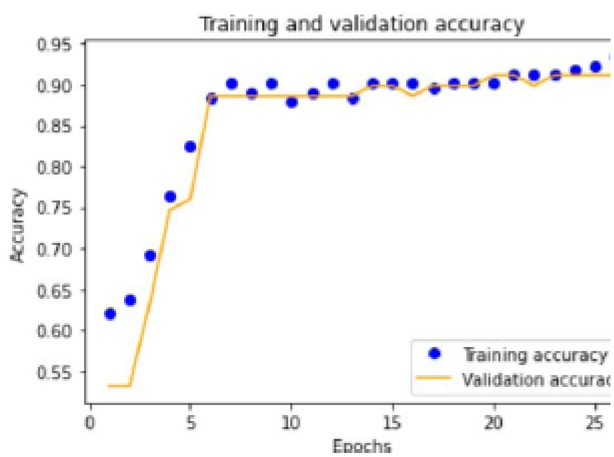
values found in the experiments. Points may not be comparable while showing similar loss curves because of changes in model weights. A solid understanding of how learning performance evolves with epoch count is provided by this training and validation loss, though. Using this strategy, you may find out if adding additional training patterns increases the validation score and if the model is overfitting during the learning phase. The results of the training phase for all the MLAs that were

ISSN NO: 9726-001X

Volume 13 Issue 02 2025

JOURNAL OF CURRENT SCIENCE

assessed are shown in Table 3, and the results obtained utilizing a 5-fold cross-validation are shown in Table 4. Figure 9 and Table 5 summarize the overall performance of all the analyzed MLAs through the datasets in the testing phase using terms like accuracy, precision, recall, and the F1-score. The findings show how well the algorithms that were examined competed. Nevertheless, a classification accuracy of 91.15 percent was attained by the deep learning algorithms that relied on neural networks, which outperformed the other traditional machine learning approaches. However, not all classifiers in high-dimensional datasets can be optimized using the same approach, which is why certain models have demonstrated good accuracy. In order to get the best outcomes, the models' many parameters must be fine-tuned. With this



(a) Figure A

(b) Figure B

FIGURE 7. Training and validation accuracy.

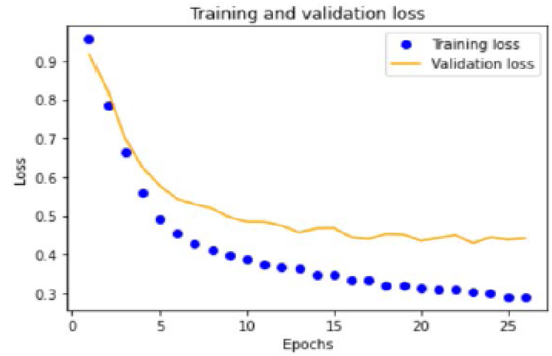


FIGURE 8. Training and validation loss.

Figure 10 and Table 6 display the outcomes of the experiment, where GA was used to find the best classifier parameters and search for an optimal 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. Precision, Recall, and Overall Score for each Machine Learning Algorithm Ranked on the Dataset

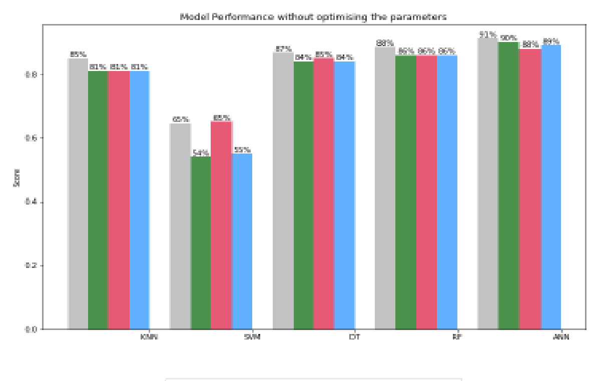


FIGURE 9. Results of the performance of all evaluated MLAs (As default parameters).

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

Evaluation metrics	KNN	SVM	DT	RF	ANN
Precision	87.22%	54.33%	93.49%	93.49%	91.33%
Recall	87.35%	66.28%	93.48%	93.48%	91.18%
F-score	87.25%	57.46%	93.47%	93.48%	91.23%
Accuracy score	87.35%	66.28%	93.48%	93.48%	91.18%

ISSN NO: 9726-001X

Volume 13 Issue 02 2025

JOURNAL OF CURRENT SCIENCE

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	ANN
Precision	87.22%	54%	93.49%	93.49%	92.25%
Recall	87.35%	65%	93.48%	93.48%	91.58%
F-score	87.25%	55%	93.47%	93.48%	91.55%
Accuracy score	83.94%	64.6%	86.99%	88.14%	91.58%

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

Evaluation metrics	KNN	SVM	DT	RF	ANN
Precision	81%	54%	84%	86%	90%
Recall	81%	65%	85%	86%	88%
F-score	81%	55%	84%	86%	89%
Accuracy score	84.96%	64.6%	86.73%	88.5%	91.1

Tabulated in Table 6 are the F-Score and the t-test. The results of the test reveal that although not all algorithms provide

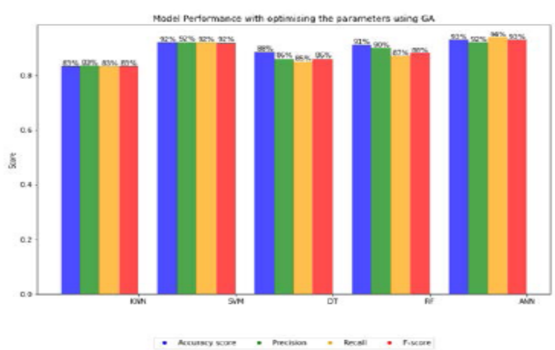


FIGURE 10. Results of the performance of all evaluated MLAs +GA (model performance with optimisation of the parameters using GA).

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

Evaluation metrics	KNN	SVM	DT	RF	ANN
Precision	83.42%	92.11%	86%	90.00%	92.01%
Recall	83.18%	92.03%	85%	87.00%	93.80%
F-score	83.21%	91.88%	86%	88.00%	91.93%
Accuracy score	83.19%	92.04%	88.50%	91.15%	92.92%

notable distinctions, the outcomes demonstrated that the suggested approach (GA+MLAs) outperformed MLAs using the default settings. Applying GA to find the best values for the MLAs 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 the KNN model, for instance, the GA recommended using the Euclidean distance metric and a value of k = 2. The KNN model was trained and tested on the complete dataset using these optimized parameters. The classification accuracy achieved by KNN, SVM, DT, RF, and ANN was 83.19%, 92.04%, 88.50%, 91.15%, and 92.92%, correspondingly. This paper used a grid search method to optimize the SVM's hyperparameters rather than the GA. With this method, you may speed up your training time, improve your results, and discover the best hyperparameter values for your SVM classifier by searching hyperparameter space. And as both studies made use of the same SleepHealth and Lifestyle Dataset, we were able to compare their findings [33]. Outperforming the most recent study, the suggested technique achieved greater outcomes by optimizing the parameters 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, 0.01, 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': 1, 'num-units-per-layer': 24, 'learning-rate': 0.004068331104981341)

ISSN NO: 9726-001X

Volume 13 Issue 02 2025

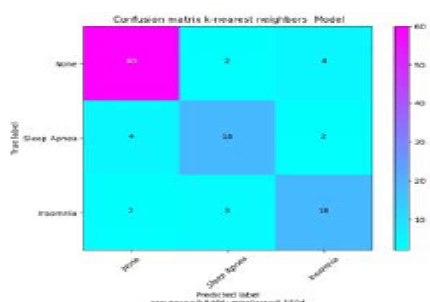
JOURNAL OF CURRENT SCIENCE

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

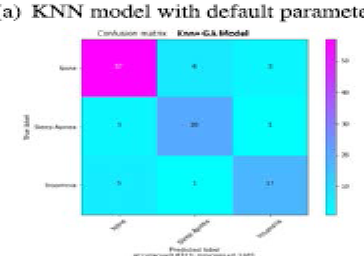
Model		t-test result	Conclusion
KNN	t-stat.	0.3375263702777991	No significant
—	p-value	0.7396243325450431	improvement
SVM	t-stat.	-1.5212776585113266	significant
—	p-value	0.14556309281759117	improvement
DT	t-stat.	0.3629888130588261	No significant
—	p-value	0.7208408332545125	improvement
RF	t-stat.	-1.3416407864997872	significant
—	p-value	0.19639447228341037	improvement
ANN	t-stat.	-1.186884852112364	significant
—	p-value	0.2507018182951046	improvement

A. T-TEST ANALYSIS

The statistical significance of the improvement achieved by the GA-optimized MLAs was assessed using a t-test for samples. According to the null hypothesis, there is a statistically significant difference between the baseline accuracy and the average accuracy of many GA-optimized MLAs models. Table 9 shows the results of the t-test, which showed that the GA-optimized MLAS classifiers significantly improved accuracy.

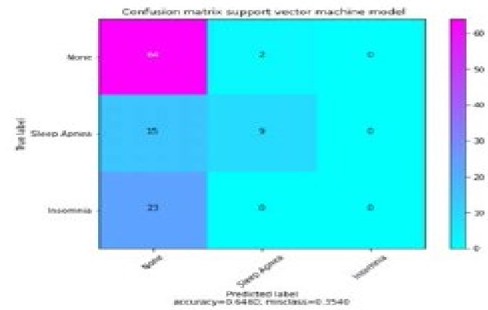


(a) KNN model with default parameters

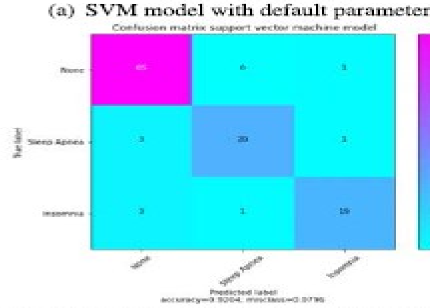


(b) KNN model with optimisation of the parameters

FIGURE 11. Confusion matrix for KNN model.



(a) SVM model with default parameters

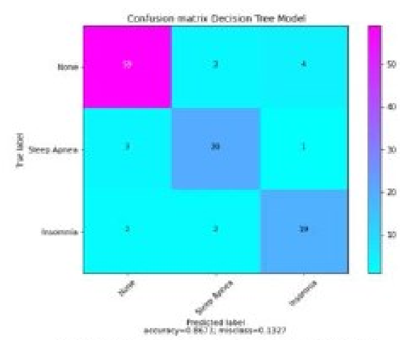


(b) SVM model with optimisation of the parameters

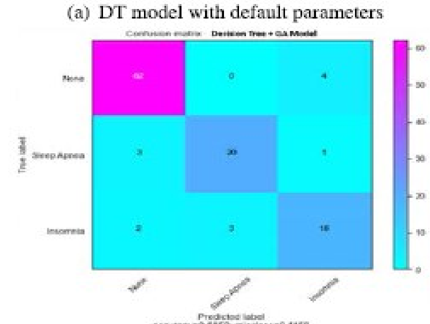
FIGURE 12. Confusion matrix for SVM model.

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

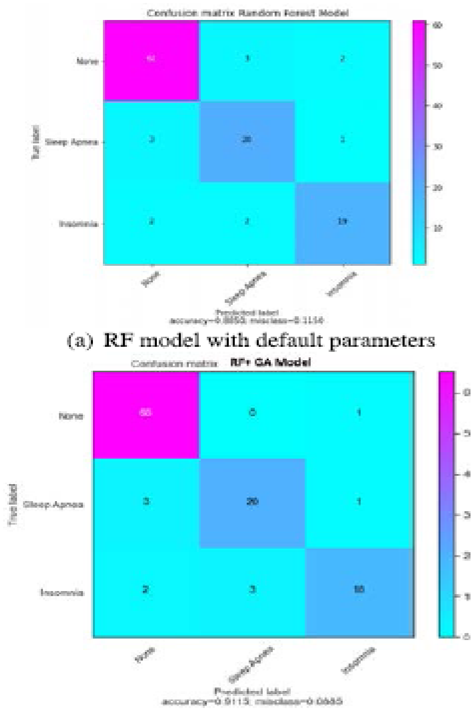


(b) DT model with optimisation of the parameters

FIGURE 13. Confusion matrix for DT model.

ISSN NO: 9726-001X

Volume 13 Issue 02 2025

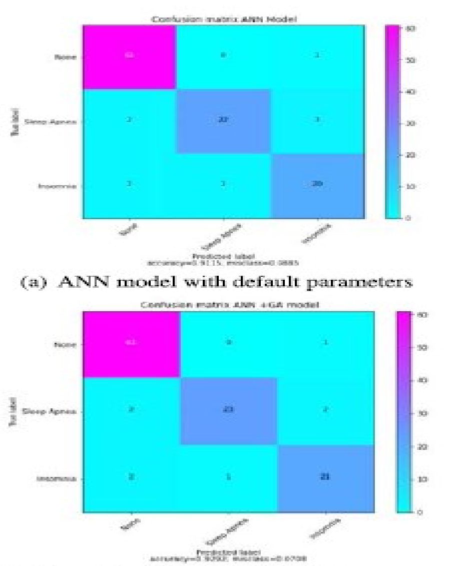


(a) RF model with default parameters

(b) RF model with optimisation of the paramete

FIGURE 14. Confusion matrix for RF model.

undertaking a classification assignment 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.



(a) ANN model with default parameters

(b) ANN model with optimisation of the parameter

FIGURE 15. Confusion matrix for ANN model.

the errors in categorization. Using the ANN+ GA model as an example, 61 out of 100 occurrences were accurately categorized as Class 1, 0 as Class 2, and 1 as Class 3. The same holds true for Class 3: 21 cases were correctly categorized, but 2 were mistakenly placed in Class 1 and 1 in Class 2. In Class 2, 23 examples were correctly identified, but 1 was mistakenly placed in Class 2. If you want to know how well the RF classifier did on the multi-class classification job, go no further than the confusion matrix. The matrix is utilized. With a 96% success rate, the model demonstrated excellent accuracy in Class 1. Class 2 and Class 3 also had lesser accuracy, with 20% and 26% of misclassifications, respectively.

V. CONCLUSION

This research proposes an improved model for sleep disorder categorization using MLAs and a genetic algorithm to get the best values for each model's hyperparameters. This study examined the effectiveness of MLAs in classifying sleep disorders and assessed many cutting-edge MLAs using the Sleep Health and Lifestyle Dataset, a real-world dataset. And MLAs don't need characteristics established by experts to learn from high-dimensional sleep data and try to categorize sleep disorders. With an accuracy of 92.92%, the suggested optimized ANN with GA outperformed the competing MLAs. On the testing data, the F1-score was 91.93%, the recall was 93.81%, and the precision was 92.01%. Regardless of the data constraint. The difficulties of using MLAs for sleep disorder categorization were the focus of this research. To train and evaluate models in this domain, however, huge datasets are still required. When it comes to sleep disorder categorization, MLAs with GA can make a world of difference. We will evaluate the dataset on a novel model and compare its performance to current state-of-the-art models; moreover, we will create MLAs using unsupervised learning.

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, p. 1203, Dec. 2019

[2] Y. Li, C. Peng, Y. Zhang, Y. Zhang, and B. Lo, ‘‘Adversarial learning for semi-supervised pediatric sleep staging with single-EEG channel,’’ *Methods*, vol. 204, pp. 84–91, Aug. 2022.

ISSN NO: 9726-001X

Volume 13 Issue 02 2025

[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, p. 24983, Jan. 2014.
[5] V. Singh,V. K. Asari, and R. Rajasekaran,''Adeep neural network for early detection 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 on traditional machine learning: Simple and interpretable techniques are competitive to deep learning for sleep scoring,'' *Biomed. Signal Process. Control*, vol. 81, Mar. 2023, Art. no. 104429.
[7] H. O. Ilhan, ''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, Art. no. 046059.
[9] Y. J. Kim, J. S. Jeon, S.-E. Cho, K. G. Kim, and S.-G. Kang, ''Prediction models for obstructive sleep apnea inKorean adults using machine learning techniques,'' *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 from single-channel EEG signals,'' *J. Neurosci. Methods*, vol. 324, Aug. 2019, Art. no. 108312.
[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, 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. no. 115950.
[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, p. 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, p. 6379, Apr. 2023.
[15] M. Bahrami and M. Forouzanfar, ''Detection of sleep apnea from singlelead ECG: 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, ''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, p. 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, p. 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, Art. no. 104778.
[22] (2023). *Sleep Health and Lifestyle Dataset*. [Online]. Available: ttp://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyledataset
[23] F. Ordóñez and D. Roggen, ''Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition,'' *Sensors*, vol. 16, no. 1, p. 115, Jan. 2016.
[24] D. M. W. Powers, ''Evaluation: From precision, recall and Fmeasure to ROC, informedness, markedness and correlation,'' 2020, *arXiv:2010.16061*.
[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 Knearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning,'' *Decis. Anal. J.*, vol. 3,

[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, C. C. Ukwuoma, A. Y. Muaad, Y. M. Kadah, M. A. Al-Antari, and J. P. Li, ‘‘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, Art. no. 102279.
[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,’’ in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 2576–2580.
[32] A. Hichri, M. Hajji, M. Mansouri, K. Abodayeh, K. Bouzrara, H. Nounou, and M. Nounou, ‘‘Genetic-algorithm-based neural network for fault detection and diagnosis: Application to grid-connected photovoltaic systems,’’ *Sustainability*, vol. 14, no. 17, p. 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.