Making Sleep Disorder Classification Using Optimized Machine Learning Models

Chandana Yamani[1*], Dileep Kumar Modepalli[2], Abhishikth Dola[3], Yalavarthi Alimoon Shaik[4], Dr. S.Siva Nageswarara Rao[5], Dr. Sireesha Moturi[6]

[1]Asst.Professor,Department of CSE, Narasaraopeta Engineering College, Narasaraopet-522601, Palnadu, Andhra Pradesh, India.
[2,3,4]Department of CSE, Narasaraopeta Engineering College, Narasaraopet-522601, Palnadu, Andhra Pradesh, India.
[5]Professor,Department of CSE, Narasaraopeta Engineering College, Narasaraopet-522601, Palnadu, Andhra Pradesh, India.
[6]Assoc.Professor,Department of CSE, Narasaraopeta Engineering College, Narasaraopet-522601, Palnadu, Andhra Pradesh, India.
chandana.nrtnec@gmail.com

Abstract. Sleep disorders have a significant impact on human health and early diagnosis is crucial for the improvement of quality of life. In this work, machine learning models that classify sleep disorders have been developed from the data set of 374 participants that comprises features relating to sleep health and lifestyle, such as sleep duration, physical activity, stress levels, and BMI. It has used many machine learning algorithms, like GNB, KNN, SVM, RF, Logistic Regression, and ANN/MLP Classifier. Dataset was highly preprocessed to deal with missing values and normalize features, thereby offering the best performance to the models. Feature engineering and optimization through a Genetic Algorithm improved the predictive ability of these models. GA was highly effective especially in the extraction of the most relevant features, improving the classification accuracy, and dealing with the problem of small dataset size.The best accuracy by GNB was achieved in combination with GA, namely 94%, compared to all other models: SVM 93%, Logistic Regression 92%, and ANN 93%. Importantly, issues like dealing with missing data and feature extraction, where meaningful features such as systolic and diastolic blood pressure levels had to be extracted, were addressed in order to enhance the outcome. Our results show that the proposed model effectively detects patterns in the database, hence providing consequent results that can be used in real-time by health care professionals to diagnose sleep-related conditions. This consequently decreases manual effort towards diagnosis and improves its timing. It also allows for scalability, which may fit easily into wearable health devices and clinical decision systems that are intended to prevent long-term health risks due to sleep disorders left untreated. The proposed model facilitates automated classification of sleep disorders. This will be a valuable tool for advancing healthcare outcomes and fostering future innovations in sleep medicine.

Keywords: Feature Engineering · BMI (Body Mass Index) · Optimization · Real-Time Diagnostics · Automated Sleep Disorder Detection.

1 Introduction

Recently, sleeping disorders have emerged as a serious public health issue due to findings where sleep quality and duration directly impact the total well-being of an individual. Once a person's normal sleep pattern becomes disturbed, they begin to suffer from very critical health problems like heart diseases and metabolic syndrome or even cognitive impairment. The most common sleeping disorder is the most common form of sleep apnea. In general, it is described as the chronic disruption of sleep ventilation mainly due to upperairway obstruction. Untreated sleep apnea may lead to possibly life-threatening complications that come from health conditions such as heart disease, stroke, or hypertension. Overall treatment for sleep apnea remains Continuous Positive Airway Pressure, whereby one keeps the airway open throughout the night. More severe conditions may require surgical interference with the treatment of obstruction within the airway on a permanent basis. CPAP has proven to be one of the strongest treatments, and its symptoms have reduced the complications and problem of sleep apnea in many cases [1]. Insomnia is yet another very common sleep disorder. It makes falling asleep difficult, staying asleep, and waking up too early. There are two types: acute and chronic insomnia. Former usually does not last long. But the latter requires long time treatment; though pharmacological along with behavioural therapy is commonly used for treatment of chronic insomnia. Among the latter there is one type of cognitive behavioural therapy. This therapy, combined with the enhancement of sleep hygiene practice has been proven to result in considerable success in the treatment of the condition [2].In the past couple of years, technology has improved much-detection and management of sleep disorders. For instance, the design for portable diagnostic systems, usually based on CNNs, has transformed big classification techniques of sleep disorders. Such designs provide much more accurate and accessible tools for assessing sleep disorders outside the traditional clinical settings. This means that patients can now seek timely interventions to enhance their long-term outcomes [3]. Ensemble machine learning algorithms have also played a major role in the advancement of developing the diagnostic accuracy for sleep disorders. Algorithm that can classify a complex pattern of sleep can become highly useful for predictions and diagnosis related to sleep disorders. This has held special value in the differentiation of sleep disorders that are similar, but will have different treatment modalities [4]. These technological developments in practice would bring much improvement in diagnosis and treatments. Such developments do heighten the quality of lives of afflicted people suffering from sleep disorders, but they also open pathways for future discoveries in sleep medicine. The advancement of diagnostic tools and further research on treatment methodologies can help a healthcare provider deal better with the complexities of sleep-related disorders [5]. Hence, this approach can lead to better care and improved long-term health outcomes.

2 Literature Review

Sleep disorders as insomnia and obstructive sleep apnea have become serious issues for an individual's well-being. Traditional approaches to diagnosis, such as PSG, are well resourced. Machine learning methods are increasingly applied to automate and improve the detection of sleep disorders, drawing on diverse data and advanced algorithms.

Recent works investigated different ML models to classify sleep disorders. For example, Alshammari attempted machine learning algorithms on the Sleep Health and Lifestyle dataset, where ANN obtained the highest accuracy of 92.92% [6]. This shows that lifestyle and health datasets can be used to predict with high accuracy, but the work does not provide advanced optimization approaches, such as feature selection algorithms, to improve the performance of the models presented.

Yadav et al. combined decision trees and Support Vector Machines (SVM) with physiological data, including EEG and ECG signals, to classify sleep disorders. The addition of physiological data greatly enhanced classification precision, with the PSDG dataset proving useful for training robust models [7]. This study does underscore the need for utilizing rich physiological datasets, though such datasets often require specialized equipment, making such applications not scalable.

Hidayat applied the Random Forest algorithm to the Sleep Health and Lifestyle dataset, which obtained 92% accuracy. This task showed that lifestyle-related factors, such as sleep patterns and health profiles, are highly indicative of the sleep disorders [8]. It lacks in-depth feature selection techniques which may be useful to further optimize the performance of the model. Similarly, Airlangga compared Logistic Regression (93% accuracy) and neural networks on the same dataset, elaborating on the strengths and weaknesses of traditional machine learning versus deep learning approaches [9]. These comparisons focus on how different methodologies are performing differently on similar datasets, although they did not discuss some techniques to increase feature relevance such as Genetic Algorithms (GA).

Ramesh et al. used machine learning models: Random Forest and neural networks, in EHRs data to classify OSA. This demonstrates the applicability of ML in healthcare with real-world clinical data [10]. However, the use of EHRs does not scale to wider applications for which simpler data sources like lifestyle information are more easily available.

Kim et al. developed prediction models for obstructive sleep apnea among Korean adults using demographic data combined with physiological signals, including ECG and respiratory inputs. The study underlined the requirement of models tailored specifically to population-specific data, and robust results were achieved [11]. Although useful for their specific population, the contribution of physiological data makes this approach less applicable to larger populations.

Tripathi et al. discussed the ensemble learning techniques, which include bagging and boosting, for the detection of insomnia using sleep ECG signals. The results showed significant improvements in the detection accuracy achieved by combining multiple machine learning techniques [12]. Limitations to real-world applicability arise due to the need for heavy, high-cost hardware-intensive physiological signals in comparison to more feasible simpler data sources, such as lifestyle or behavioral information, in certain settings.

Limitations persist in previous works. Most models fail to scale or do not utilize optimization techniques in feature selection, leading to suboptimal performance. This study fills these gaps by integrating Genetic Algorithms (GA) for feature selection with ML models like GNB, SVM, and ANN. GA not only reduces feature redundancy but also increases model accuracy, particularly in smaller, diverse datasets like Sleep Health and Lifestyle. Unlike previous work, we extracted latent features, such as the levels of systolic and diastolic blood pressures, to further improve the model's predictiveness. This is a substantial contribution because it combines sophisticated feature engineering with scalable inputs, good for real-time applications.

Building from previous strengths and avoiding their weaknesses, our work offers a new way toward the classification of sleep disorders, based on both advanced optimization techniques and accessible datasets.

3 Methodology

The methodology in the fig.1 presents a typical workflow for machine learning. It begins with using the collected data and preprocessing followed by feature extraction/normalization to prepare it to feed into the system. Then,

the dataset needs to be divided to be split for training and testing purposes. For identifying which features were most appropriate for the training of the model, a genetic algorithm was employed with the objective of feature selection while optimizing the input variables of the model being trained.Lastly, the model was trained, and the performance of that model was estimated by evaluating accuracy results.

3.1 Dataset

The data is collected from the Kaggle named as "Sleep Health and Lifestyle" This dataset covers 374 participants with 12 features [13]. It dives into their habits, work life, physical activities, health markers like heart and blood pressure. Researchers generally use it to see how these factors might affect sleep quality or even lead to issues like sleep apnea and Insomnia. Although the dataset was useful, the number of participants in it which was just 374 necessarily limited it from the perspective of generalization. Thus, to minimize these problems, careful preprocessing and feature selection steps were followed to reduce overfitting. However, the study recognizes that including more diverse participant groups into the dataset in future research will strengthen the finding. In addition, linear patterns of the dataset and demographic representation also influenced the choice of models, though the results need validation through future studies on larger more varied datasets.

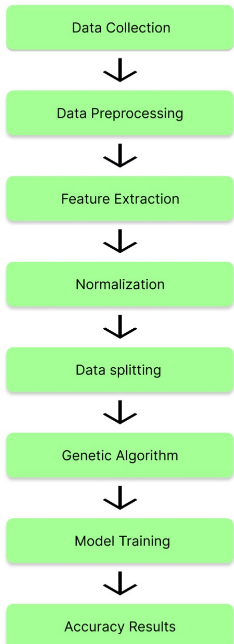

Fig. 1. Workflow of Models

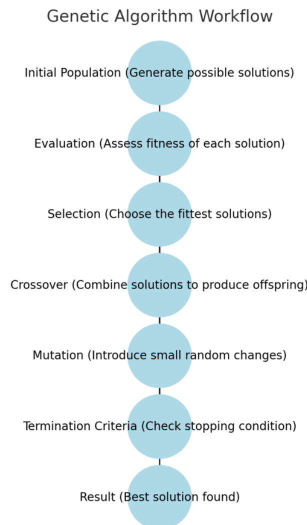

Fig. 2. steps of Genetic Algorithm

3.2 Data Preprocessing

i) Handling Null Values : Null values can cause the data inconsistency and may the model cannot be trained well at the same time there will be the downfall in achieving accuracy. In our dataset the target column contains the null values we replace it with the 'none' this might help to train the model without null values. For the target attribute we cannot use the statistic method such as mean and mode due to dataset contains the people with not having any sleep disorder.

ii) Feature Extraction : This method is mainly used to extract the new feature that is hidden and which we cannot see. we can make the meaningful variables by extracting them[14].so we can improve our model performance rapidly. As coming to our dataset there is a hidden feature that can be extracted from the blood pressure attribute can be called as systolic and diastolic which might be helpful for increasing the model performance.

iii) Label Encoding : label encoding technique is mainly used to transform the categorical columns into numerical columns. It helps in applying mathematical operations and improving performance. It assigns each category to a unique identification number. Our dataset contains the categorical attributes like gender, occupation, bmi and sleep disorder these will be transformed into a numerical columns.

iv) Standardization : Scaling of features to have similar ranges of values is very important in machine learning. Standardization - or normalization - is the process of rescaling one or more attributes to a common range, usually 0 to 1.This method will be of much help in the data that contains various of features with various units.

3.3 Analysis of Features

Relation with Target Feature : The scatter plots are suitable for visualizing and simplifying the relationship between the two numerical variables. They can show the patterns, trends, and interconnection between the variables. Outliers can be easily identified. In this case, a scatter plot can help us to identify on how the age might be related to sleep disorders. If you have a categorical target variable, you might consider other visualizations like bar plots.In the below we can see the scatter plot that are associated with Age and Sleep Disorder and bar plot graph of Gender related with Sleep Disorder.

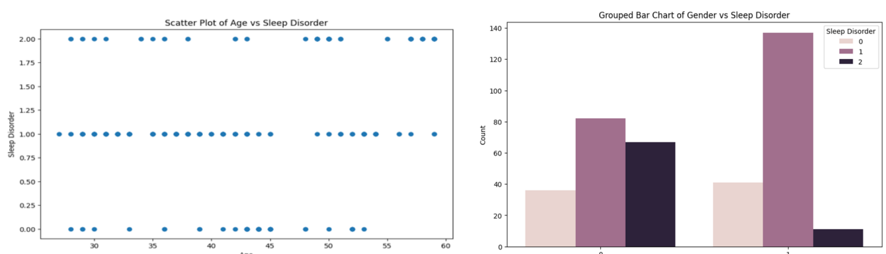

Fig. 3. Relationship between features

3.4 Models

1. Support Vector Machine (SVM) : Support Vector Machine is one among those used supervised learning algorithms that have been widely applied for the classification problems. It can adapt to the two types of data distributions, linear and non-linear, by using different kernel functions, such as a linear kernel, polynomial kernel, and Radial Basis Function kernel. In this experiment, the features of the dataset were aligned linearly, so a linear kernel(kernel='linear') was used [15].

2. Random Forest (RF) : For dealing with complex data distributions and possible overfitting, the ensemble learning technique of Random Forest has been applied. Hyperparameters have been tuned to optimize the model. These include values such as n-estimators = 600 that creates many decision trees, max-depth = None; that is, letting them get as deep as possible so that splits on a node are optimal. As compared to criterion which uses entropy to maximize information gain on a split, the parameters min-samples-split and min-samples-leaf prevent growth trees from getting unbalanced [16].

3. K-Nearest Neighbour (KNN) : KNN is a non-parametric algorithm that is quite simple and yet powerful, since it bases its predictions about the class of a given data point on its nearest neighbors. For this small dataset, n-neighbors=2 has been used as hyperparameter to optimize performance, combining the use of Euclidean for the distance metric in order to preserve very high accuracy in classification.

4. Gaussian Naive Bayes (GNB) : Though the assumptions are rather simplistic, GNB is computationally efficient and proves to be good when dealing with small-scale datasets. The hyperparameter var-smoothing=1e-9 had been used in the order to achieve numerical stability. Classifiers proved efficient and trained.

5. ANN (MLP Classifier) : Artificial Neural Network (ANN) or Multi-Layer Perceptron (MLP) was applied because it mimics the structure of the human brain consisting of interconnected neurons. That network was set using hidden-layer-sizes=(24,), properly optimized for the size and complexity of this particular dataset. ANN finds best to capture anything but linear patterns as well as relationships between features.

6. Logistic Regression (LR) : Logistic regression was a very efficient classification algorithm, which was interpretable and suitable for a smaller dataset. max-iter=1000 was used as a hyperparameter during the whole training process so that the model must converge to stable and reliable results.

7. Genetic Algorithm (GA) : GA Feature selection and model improvement used an optimization technique. Critical hyperparameters of the GA were used, including population-size=12 evaluated 12 feature subsets per generation and n-generations=9, assuring sufficient evolution to guarantee optimal solutions; a relatively high mutation-rate=0.8 was applied in order to maintain diversity and avoid local optima. GA assisted considerably in the optimization procedure related to feature selection, especially for a small dataset, to reduce dimensionality and improve model performance.

4 Results and Discussion

4.1 Steps in Genetic Algorithm

In the fig. 2 we can see the steps of Genetic Algorithm

- Phase 1: First Generation Generation
- Randomly creates an initial population of individuals (chromosomes), as binary strings, who carry a potential solution.
- Phase 2: Evaluation
- Evaluate individual fitness by using a problem-specific function according to which the effectiveness of an attribute is judged for those individuals.
- Phase 3: Selection
- Chooses parents for the subsequent generation based on performance ranking, where the better ones have better fitness.
- Phase 4: Crossover
- Combines two chromosomes from parent pairs using single-point, multi-point, or uniform crossover technique to produce offspring.
- Phase 5: Termination
- Its primary check is whether convergence criteria are met or not. They check for maximum fitness or minimal improvement after a certain number of generations. If not met, then reevaluate.

4.2 Confusion Matrix and ROC Curve

In fig.6 a confusion matrix for a GNB classifier that gives a visual summary of the model performance

The confusion matrix for the GNB classifier shows that this model correctly classified three classes of sleep disorder. The classification was correct in 11 instances for class 0, 44 for class 1, and 16 for class 2. There were misclassifications for 1 instance of class 0 miscoded as class 1, 2 instances of class 1 miscoded, (1 as class 0 and 1 as class 2), and 1 instance of class 2 miscoded as class 1. There is overlap in such classes that were found from the results. Further exploration of feature distributions and model tuning could lead to this reduction in errors and an improvement in the distinction among similar sleep disorders, thus giving better real-world performance in clinical applications.

In Fig.5 we can see the ROC curve analysis showed that the model performs best for Class 0 (AUC = 0.96), followed by Class 1 (AUC = 0.94) and Class 2 (AUC = 0.93), showing a slight drop in performance for Class 2. Curves are illustrated for misclassifications; false positives show the tendency to be higher in classes with lower AUCs, indicating greater challenges in the discrimination of particular classes. These misclassifications can have real-world consequences, particularly when Class 1 and Class 2 represent similar sleep disorders, leading to the very real consequences of mistaken diagnoses. Misclassifications could well improve the model's general effectiveness when put to use in practice.

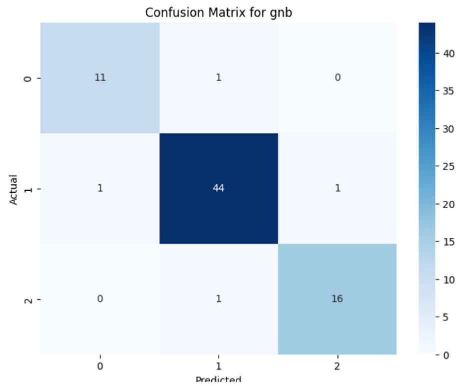

Fig. 4. Confusion Matrix of GNB

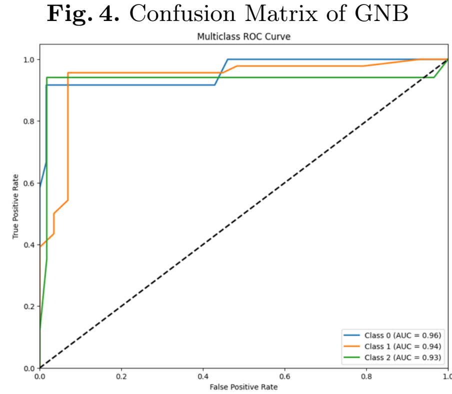

Fig. 5. ROC Curve of GNB

4.3 Performance Metrics

In the Table.1, We can see the different performance metrics like Accuracy,F1-Score,Precision and Recall related to the different classifiers with various results are mentioned.

Table 1. Performance Metrics Of Classifiers Using Genetic Algorithm.

Models	**Accuracy**	**F1-Score**	**Precision**	**Recall**
GNB	0.94	0.94	0.94	0.94
MLP	0.93	0.93	0.93	0.93
SVM	0.93	0.93	0.93	0.93
Random Forest	0.92	0.91	0.92	0.92
Logistic Regression	0.92	0.91	0.92	0.92
KNN	0.85	0.86	0.87	0.85

4.4 Anova Test

To assess the statistical significance of the performance differences between the classifiers, we conducted an ANOVA test. The results showed a significant difference in performance, with an F-statistic of 14.16 and a p-value of 3.49×10^{-12}

. This indicates that at least one classifier outperforms the others in terms of accuracy or fitness score.

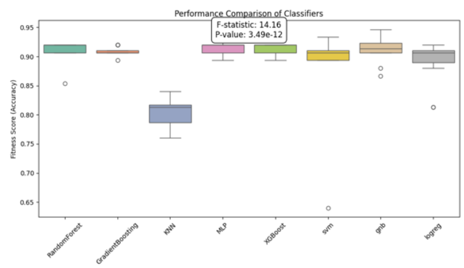

Fig. 6. Anova test for Classifiers

4.5 Computational Efficiency and Trade-offs:

GNB has greater computational efficiency than MLP. This model would train in just 0.0020 seconds and predict in 0.0012 seconds, making the same perfect for working in real-time or low-resource embedded systems. MLP takes 0.1741 seconds for training and 0.0015 seconds for predictions, thus is slow.

This means that GNB has a much smaller model size, compared to MLP which is around 1.48 KB, and MLP at 13.48 KB, which is ideal in case the devices have constraints regarding storage. GNB performance competes with MLP while giving it a higher fitness score: 0.9467 compared to 0.9333 while being significantly cheaper in terms of calculation. Therefore, in applications where speed, efficiency, or resource limitations become critical, GNB seems to be the better choice, balancing accuracy and efficiency.

4.6 Comparitive Analysis

In this study, we implemented these multi-class algorithms: GNB, MLP, SVM, Random Forest, Logistic Regression, and KNN using a genetic algorithm that also performs the feature selection process and enhances it. It is so important to establish that the genetic algorithm powered the enhancement process in feature subsets for enhancing model performance. GNB achieved an accuracy of 94%, with F1-score, precision, and recall all in strong agreement. MLP and SVM followed closely with nearly balanced results at 93%, while the Random Forest and

Logistic Regression showed robust accuracy at 92%, though the F1-score was a little lesser, at 91%. KNN had very poor performance as its accuracy rested at 85%, which seems less effective even after the optimization process. These results have been found, which are underlining the effectiveness of using genetic algorithms in a direct feature selection process and more so in enhancing GNB, MLP, and SVM but may require further adjustments on KNN in order to boost accuracy.

5 Conclusion

The best feature selection was carried out by applying a Genetic Algorithm which showed the most marked performance on various classifiers. In this, the GNB classifier reached the highest accuracy that was 94%, which was better in all other models. This reflects the feasibility of GAs in filtering out unnecessary attributes to yield efficiency on lightweight models, such as GNB, so as to attain better predictive accuracy. This coupling of GA and GNB provides an efficient and practically viable approach to enhancing performance with predictive modeling tasks sans the entrance into the realms of complex algorithms.

The implications are very high, especially in applications where wearable health devices and real-time diagnostics are concerned. GNB is light and, along with the feature reduction capability of GA.However, there are still several limitations to the study namely a relatively small size of the dataset and possible biases that may decrease the generalizability of the results; in fact, subsequent studies will be able to cross the problem with the help of larger and diversified datasets and deep learning models that can be deployed for the evaluation of performance in similar situations. Further investigation of the effect of GA on other lightweight and complex classifiers would offer broader insights into the applicability of these across domains.

The research presented here focuses on the computational effectiveness as well as real-world usability furthering the efficient integration of feature selection based on GA with lightweight classifiers in real-world applications, potentially benefitting from the same higher accuracy and speed in even more resource-constrained settings.

References

1. P. Levy, V. Viot-Blanc, and J.-L. Pépin, "Sleep disorders and their classification–an overview," *Sleep Apnea*, vol. 35, pp. 1-12, 2006.
2. C. R. Soldatos, J. D. Kales, T. L. Tan, and A. Kales, "Classification of sleep disorders," *Psychiatric Annals*, vol. 17, no. 7, pp. 454-458, 1987.
3. A. Atianashie Miracle, E. Armah, and N. Mohammed, "A portable GUI-based sleep disorder system classification based on convolution neural networks (CNN) in Raspberry Pi," *J Eng Appl Sci Human*, vol. 6, pp. 13-23, 2021.

4. M. Sharma, J. Tiwari, V. Patel, and U. R. Acharya, "Automated identification of sleep disorder types using triplet half-band filter and ensemble machine learning techniques with EEG signals," *Electronics*, vol. 10, no. 13, p. 1531, 2021.
5. T. Wongsirichot and A. Hanskunatai, "A classification of sleep disorders with optimal features using machine learning techniques," *J Health Res*, vol. 31, no. 3, 2017.
6. T. Alshammari, "Applying machine learning algorithms for the classification of sleep disorders," *IEEE Access*, 2024.
7. P. K. Yadav, U. K. Singh, J. J. A. Kovilpiaali, and R. Tamilarasi, "Sleep disorder detection using machine learning method," in *2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS)*, IEEE, pp. 1530-1532, 2023.
8. I. A. Hidayat, "Classification of sleep disorders using random forest on sleep health and lifestyle dataset," *Journal of Dinda: Data Science, Information Technology, and Data Analytics*, vol. 3, no. 2, pp. 71-76, 2023.
9. G. Airlangga, "Evaluating machine learning models for predicting sleep disorders in a lifestyle and health data context," *JIKO (Jurnal Informatika dan Komputer)*, vol. 7, no. 1, pp. 51-57, 2024.
10. 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, 2021.
11. Y. J. Kim, J. S. Jeon, S.-E. Cho, K. G. Kim, and S.-G. Kang, "Prediction models for obstructive sleep apnea in Korean adults using machine learning techniques," *Diagnostics*, vol. 11, no. 4, p. 612, 2021.
12. P. Tripathi, M. A. Ansari, T. K. Gandhi, R. Mehrotra, M. B. B. Heyat, F. Akhtar, C. C. Ukwuoma, et al., "Ensemble computational intelligence for insomnia sleep stage detection via the sleep ECG signal," *IEEE Access*, vol. 10, pp. 108710-108721, 2022.
13. Sleep Health and Lifestyle Dataset. Available at: http://www.kaggle.com/datasets/uom190346a/sleep-health-andlifestyledataset, 2023.
14. M.Sireesha, S. N. TirumalaRao, Srikanth Vemuru, Optimized Feature Extraction and Hybrid Classification Model for Heart Disease and Breast Cancer Prediction International Journal of Recent Technology and Engineering Vol - 7, No 6, Mar - 2019 ISSN - 2277-3878, Pages – 1754 – 1772
15. Sunayna, S. Siva, SN Thirumala Rao, and M. Sireesha. "Performance evaluation of machine learning algorithms to predict breast cancer." In Computational Intelligence in Data Mining: Proceedings of ICCIDM 2021, pp. 323-335. Singapore: Springer Nature Singapore, 2022.
16. Moturi, Sireesha, Jhansi Vazram Bolla, M. Anusha, M. Mounika Naga Bhavani, Srikanth Vemuru, SN Tirumala Rao, and Sneha Ananya Mallipeddi. "Prediction of Liver Disease Using Machine Learning Algorithms." In International Conference on Data Science and Applications, pp. 243-254. Singapore: Springer Nature Singapore, 2023.