Computational Journal of Mathematical and Statistical Sciences
3(2), 341–358
DOI:10.21608/CJMSS.2024.290167.1053
https://cjmss.journals.ekb.eg/

Predicting Sleep Disorders: Leveraging Sleep Health and Lifestyle Data with Dipper Throated Optimization Algorithm for Feature Selection and Logistic Regression for Classification

El-Sayed M. El-Kenawy[1]∗, Abdelhameed Ibrahim[2], Abdelaziz A. Abdelhamid[3,4], Nima Khodadadi[5]∗, Laith Abualigah[6], Marwa M. Eid[1]

[1] Delta Higher Institute of Engineering and Technology, Mansoura 35111, Egypt

[2] Computer Engineering and Control Systems Department, Faculty of Engineering, Mansoura University, Mansoura 35516, Egypt

[3] Department of Computer Science, Faculty of Computer and Information Sciences, Ain Shams University, Cairo 11566, Egypt

[4] Department of Computer Science, College of Computing and Information Technology, Shaqra University, 11961, Shaqra, Saudi Arabi

[5] Department of Civil and Architectural Engineering, University of Miami, Coral Gables, FL, USA

[6] Hourani Center for Applied Scientific Research, Al-Ahliyya Amman University, Amman 19328, Jordan. Al al-Bayt University, Mafraq, Jordan. Applied science research center, Applied science private university, Amman 11931, Jordan.

* **Correspondence:** skenawy@ieee.org & nima.khodadadi@miami.edu

Abstract: This paper is a thorough examination of the modeling of sleep disorders based on machine learning that is applied to the sleep-health-and-lifestyle data. The use of the Dipper Throated Optimization Algorithm for feature selection and Logistic Regression for classification is the basis of the study that explores the effectiveness of predictive models in identifying sleep disorders based on varied sleep metrics and lifestyle factors. The binary Dipper Throated Optimization Algorithm was the most successful with the lowest Average error of 0.71933 uses feature selection as the most effective method, which proves that it is successful the method of choosing the relevant features for predictive modeling. Moreover, Logistic Regression proved to be very efficient in classification; it got an Accuracy of 0.95. The results of these studies support the idea of the personalized treatment and earlier detection of sleep disorders; this, in turn, will be of great help to the progress in sleep health research and healthcare practice.

Keywords: Sleep health; Lifestyle factors; Feature selection; Dipper Throated Optimization Algorithm; Logistic Regression; Sleep disorder prediction.
Mathematics Subject Classification: 68N01; 68N30; 78M50
Received: 16 May 2024; Revised: 30 May 2024; Accepted: 8 June 2024; Published: 9 June 2024.

1. Introduction

In the diverse landscape of human health, sleep stands out as a basic element that is connected to physical health, mental ability and emotional strength. On the other hand, sleep is not only in terms of time but also has a diverse web of physiological processes, neurological rhythms, and environmental factors, which make our whole being and energy [1, 2, 3]. A large part of the existing studies proves the great influence of sleep health on overall health, which is, unfortunately, affected by not enough or disturbed sleep, which, in turn, is associated with many health issues that are both physical, cognitive and emotional.

The Sleep Health and Lifestyle Dataset, a robust compilation of sleep-related data and lifestyle factors, serves as a cornerstone in our sleep research. This comprehensive dataset encompasses a wide range of variables, from demographic characteristics to intricate sleep metrics and lifestyle behaviors. Its inclusion in our study provides a holistic perspective on sleep health and its implications for human development, bolstering the credibility of our findings [4, 5, 6].

In the modern healthcare environment, sleep health is now viewed as a subject of great importance, with sleep disorders being a major problem for most of the world. Insomnia, which is a typical sleep problem, is caused by many people who try to initiate or maintain their sleep for a long time and sleep apnea, which is another sleep disorder, is caused by many people who experience various episodes of their breathing being interrupted during sleep. Apart from their immediate effects on sleep quality, these disorders have other big consequences, like the diseases, mental health disorders, and overall morbidity that they cause [7, 8, 9].

Taking into account the rapid increase of sleep disorders and their main impact on human health, there is a pressing need to find ways of early detection and intervention. Predictive modeling, which is based on the principles of machine learning, provides a strong way for the complexities of sleep health to be dealt with and the risk factors for sleep disorders to be identified by people at high risk. Machine learning algorithms help researchers analyze the huge amount of data they have by finding hidden patterns and predictors that are not detectable by standard analytical methods [10, 11].

Central to our predictive modeling approach is the feature selection process, where we identify the most crucial attributes of the dataset to build our predictive models. In this study, we focus on the binary Dipper Throated Optimization Algorithm, a novel feature selection method renowned for its ability to handle complex datasets and pinpoint the most significant features. By employing this algorithm, our aim is to unravel the intricate sleep health patterns and uncover the factors that influence sleep patterns and susceptibility to disorders [12, 13, 14].

Moreover, we are starting a wide range of investigations of machine learning models to find out how effective they are in classifying people with sleep disorders. Logistic Regression is one of the most notable algorithms for this kind of task, as it is simple, interpretable, and effective in binary classification tasks. Through the application of Logistic Regression and other classification models to the evaluation and analysis of sleep disorder prediction, we want to study their strengths and weaknesses.

The combination of these many strands of research will enable us to reveal the complex interaction between sleep patterns, lifestyle, and sleep disorder risk. In order to understand the intricacies of sleep health, we aim to provide the necessary means for the creation of interventions and preventive strategies that will contribute to the well-being and healthy feelings of people from all races, social classes and backgrounds. We intend to facilitate a change in the way sleep health is approached, thus leading to a new era of personalized, data-driven interventions tailored to individual needs and requirements.

2. Literature Review

As shown in [15], The latest technology, multi-modal sensors and technology for the monitoring of physical activity, sleep, and circadian rhythm, has become very popular in recent years. This increase in the field has been the first time that large-scale, accurate sleep monitoring is possible. The amount of data that is produced is enormous, and the fields of the usage of this data are also many, for example, the research that investigates the relationship between sleep patterns and diseases or wellness applications like the sleep coaching of individuals with chronic conditions. On the contrary, to achieve the total advantages of these technologies for people and research, medicine, and research, many major barriers should be surpassed. There are vital problems with performance evaluation and data storage, which are extremely important fields. Also, a high-resolution camera has been used for the performance evaluation of the systems and for the purpose of sleep data collection, which is used for digital sleep from an interdisciplinary point of view. The newest sleep-monitoring technologies are introduced, and the future and current problems resulting from data collection and the final use of the insights in clinical and consumer settings are presented. In addition to that, the advantages and disadvantages of the present and future sensing methods, in particular the data-driven technologies that are currently being developed, such as Artificial Intelligence, are considered. The rising healthcare costs, together with the increasing demand for better health outcomes, have resulted in a new healthcare approach, which has moved from reactive to proactive. The knowledge of artificial intelligence (AI), which is about the investigative machines that are able to get a task done, and the expansion of mobile health technologies (e) is one of the widely emerging trends of the twenty-first century. The emergence of apps and wristbands are among the factors that contribute to the betterment of health care at a younger age. In [16], I found out that AI algorithms can be used to boost the evaluation of wellness, the creation of prevention strategies that will later be very useful in promoting a healthy lifestyle and the discovery of disease risk factors that were unknown before. The chapter is divided into three topics of wellness (physical, mental, and social), and it presents studies based on AI that use new data sources or analyze existing data sources to enhance preventive medicine.

Sleep is the main part of a person's daily life that takes a lot of hours of time and has a huge impact on the health and well-being of a person. Smart Homes, famous for being able to recognize the human activities that take place in the houses, are the ones that create living environments that are better than the ones that we used to have [17]. The researchers used the ARAS dataset from Smart Homes, which has multiple residents, and they examined the residents' sleep behaviors in connection with other activities. They employed different machine-learning techniques to predict these behaviors. Logistic regression (RL), Linear Discriminant Analysis (LDA), K-Nearest Neighbor (KNN), Naive Bayes (NB), Classification and Regression Trees (CRT) and Support Vector Machine (SVM) were used for this investigation, and the purpose was to find out how these activities affect the residents' health. The experiments have shown that SVM was the best at the prediction of sleep activity for both House A and B, and thus, it proved to be an excellent algorithm for this task. In House A, Resident1 scored 90% accuracy, but Resident2 got 80%, while in House B, Resident1 got 100% accuracy, with Resident2 reaching 90%. These findings show that Resident 1 in both houses' sleep activities gets more time than the others because he/she has his/her own activities, and Resident 2 spends less time on sleep activities than the other activities. [18] aims to evaluate sleep quality and behavioral health through wearable devices and smartphones, and thus, an Adaptive Neuro-Fuzzy Inference System (ANFIS) is used for analysis. Smartwatches give users data about their physical activity and sleep, while a smartphone app gathers real-time usage data. The introduction of a new behavioral health indicator (B. Health) was conducted by the researchers in such a way that the study used real-time physical activity, sleep and screen time data to assess it. The features were sorted out, health being the top priority, and the top-ranked features were to be used for training the ANFIS model for the assessment of behavioral health. The Systematic Minority Oversampling Technique was used to augment the data, and as a result, the ANFIS model reached an accuracy of 91%. The research documented that the selection rate for sleep quality evaluation was 69%, and the selection rate for the class was 85%. 79% of patients get the exact same examination of their emotional and behavioral health.

Metabolic syndrome (MetS) is a condition that is very complicated and has big health consequences. MetS can be predicted in middle-aged people before it is known, which can, in turn, lead to better health outcomes, especially those related to cardiovascular diseases. [19] is trying to solve this acute problem by the use of the most modern machine-learning techniques to establish an optimum MetS prediction model for middle-aged Korean people. This research is very significant in the realm of public health and may even be the reason for saving lives. AI allows the processing of ”big data,” which is a product of the merger of clinical, environmental and laboratory-based objective data. Thus, it is possible to have a deeper knowledge of sleep and sleep disorders. The breakthrough in this area of science can change sleep medicine in the next years in the direction of better patient care and our ever-growing knowledge about human sleep. The [20] starts with a general introduction of the different parts and the analytic techniques of the AI field, and at the same time, it deals with the current state of the field. Sleeping disorders are among the many areas where AI is applied; genotyping, diagnosing, and treating are the ones that are talked about, and the context of precision/personalized sleep medicine is given. Various ways of AI that both improve and overpower patients’ clinical impact are discussed, along with the ethical points concerning the prognostic information from AI. In the clinical field, the early successes of AI in the field of radiology and pathology are discussed so that one can easily understand what sleep medicine is going to face when using this technology. In conclusion, the essential drawbacks that should be pointed out in order to make sure that the implementation of clinical AI will be carried out most efficiently and safely are discussed.

Precision medicine (PM) is an innovative technique that is being designed for the deterrence, detection, management, and therapy of diseases. It can be considered as the answer to the sleep healthcare problems of the future. The principle of PM is derived from evidence-based medicine, which means a treatment decision is made after considering the genes, environment, and lifestyle factors that influence individual variability. The reality that the chronic sleep loss problem is a big problem that affects millions of people every year thus, it is a significant public health issue is evident to everyone. The same methodology can also be applied in the sleep medicine field to treat sleep disorders, where physicians and scientists can decide on the most appropriate strategy for the prevention, diagnosis, treatment, and management of the issue. [21] is supposed to research the new functions of PM and its application in the field of sleep medicine through the combination of fresh technology such as artificial intelligence (AI), machine learning (ML) and the use of blockchain technology which will be helpful in the diagnosis and treatment. An extensive literature review was done in all the main databases, which are PubMed/MEDLINE, Web of Science, Scopus, and Google Scholar, with the keywords “Precision Medicine" AND "sleep disorders" OR "Sleep medicine," in order to collect and analyze the data. Although PM in Sleep Medicine is still in its early stages, it is very promising in the field of sleep medicine, and it is expected to be used in many fields of sleep healthcare in the near future. The application of AI and ML in the concept of PM has altered the testing, diagnosis and treatment method for sleep disorders, which is now anticipated to be helpful to the entire world through sustainable efforts and innovation. With the rise of the number of different sensor modalities [22], the advancement of wearable sensing and mobile computing technologies has been made possible. Hence, new plans to collect the health and well-being data longitudinally have been revealed. Wearable and mobile devices can be used for low-cost, objectively, clinically relevant data collection of physical activity, patient assessment and scalable behavior monitoring in large populations. The information from this study can be used in both interventional and observational studies in order to establish a link between behavior, health, and disease and to also improve the personalization and effectiveness of commercial wellness applications. Today, more than 400,000 people worldwide are being tracked prospectively through accelerometers for epidemiological studies. Thus far, epidemiologists and clinicians have been using self-reports of physical activity and sleep, which, although they are very useful in situations without any other choice, are biased and sometimes only provide a part of the information. Wearable devices gather physical behavior data, which helps in deriving the physically assessed and sensor-assessed physical behaviors. Thus, the limitations of self-reporting are removed. The main goal is to connect these data to clinical outcomes, and the future use of the results is to provide preventive and predictive medicine. Besides, the incorporation of artificial intelligence (AI), sensor fusion, and signal processing in the data of wearable sensors has led to the improvement of human activity recognition and behavioral phenotyping. This paper looks at the newest progress in wearable and mobile sensing technology in the fields of epidemiology and clinical medicine, and therefore, the role AI plays in this area is explained.

The later few years have been marked by the fast development of AI-powered technology in the sleep medicine field, as shown in [23]. AI is the capacity of computer systems to do tasks that are generally considered to be the domains of human intelligence, like speech recognition, decision-making, etc. The task of recording and measuring physiological signals in sleep and the application of sleep tracking are commonly used. Thus, sleep monitoring in the lab as well as in the field outside the house gives us a huge amount of data, which, therefore, puts the sleep medicine field to benefit from AI. In [24], It is possible to classify sleep stages in real-time, which will be very useful, for instance, in noisy environments as it can, for example, help against noise and also provide clinical staff with the ability to analyze sleep patterns etc. Currently, the method of classifying sleep stages, polysomnography (PSG), is too complex, and too much time is needed to complete it. This paper takes a look at 27 studies that make use of Cloud Storage Technologies (CSTs) together with AI models to classify sleep stages. The AI models and their outputs are shown, and the comparison is made to make the present-day situation of the sleep stage classification and CSTs similar. Moreover, the existing methods, when they succeed and when they fail, are shown and how these AI models could be improved in the future. Furthermore, the problem of the development of simulations for people who are asleep is shown, which then led to the introduction of more interactive sleep interventions based on AI-infused CST solutions.

3. Proposed Methodology

3.1. Dataset

The Sleep Health and Lifestyle Dataset is the source of our research; it is the database that contains all the variables related to sleep health and lifestyle habits. The dataset includes 400 entries and 13 columns, which is complete data that will give information about people's sleep patterns and daily routines [25]. The dataset contains detailed information on the subjects, from demographic attributes like gender and age to physiological measures such as blood pressure and heart rate, and also on lifestyle factors like physical activity levels and stress ratings to subjective evaluations of sleep quality and duration, which gives a complete picture of sleep health.

Data Preprocessing Steps:

The preprocessing phase is crucial in guaranteeing that the dataset is of the highest quality and integrity that will be used for further analysis and modeling. This phase encompasses several key steps:

1. Handling Missing Values:
   The removal of data points is a norm in real-world datasets and has already been taken into consideration to prevent biases in the analysis. In our research, the missing values in the Sleep Health and Lifestyle Dataset are recognized and then replaced with the most efficient imputation methods. These methods may be the mean or the median imputation, where the missing values are replaced with the mean or median of the corresponding feature, or the more advanced methods like k-nearest neighbors' imputation, which uses the similarities between the data points to the degree of missing values quite accurately.
2. Encoding Categorical Variables:
   Categorical variables, for example, gender and occupation, are not numerical and, therefore, cannot be used for computational analysis directly. They must be converted into a format that is suitable for computational analysis. Through the use of encoding methods like one-hot encoding or label encoding, we are able to get this done. One-hot encoding of the categorical variables into binary columns is, on the one hand, where each category is represented by a binary indicator, and on the other hand, label encoding, which means that a unique numerical label is assigned to each category. Consequently, by encoding the categorical variables, we make sure that the algorithms can understand and use these variables properly during the modeling stage.
3. Feature Scaling:
   The characteristics of the dataset often are at different scales, thus affecting the performance of the machine learning algorithms. We fix this problem through feature scaling techniques and standardize or normalize the features to a comparable range. Standardization converts features to a mean of zero and a standard deviation of one, while normalization makes them have a predefined range between zero and one. In this way, the features are enhanced, and the convergence properties of the algorithms are improved; thus, model training becomes more efficient. By carefully cleaning and preprocessing data, we first make the Sleep Health and Lifestyle Dataset ready for further analysis and modeling. Through the process of ignoring missing values, coding of categorical variables, and normalization of features, we enable the next machine learning algorithms to extract the relevant insights from the data and make accurate predictions regarding sleep

disorders based on the complete sleep and lifestyle metrics included in the dataset.

Figure 1 shows the heatmap visualization of the Sleep Health and Lifestyle Dataset, which, in turn, offers a detailed summary of the correlations between the different variables. The heatmaps are very useful for the visualization of the relations in the datasets since the colors stand for the force and the sign of the correlations. The heatmap shows that darker colors represent strong correlations, whereas lighter colors reveal weaker or negative correlations. Through the heatmap, we can get information about the connection between sleep and lifestyle metrics, which can be used in the feature selection and model development in future analyses. Thus, we can better understand the relationship between the various sleep and lifestyle metrics.

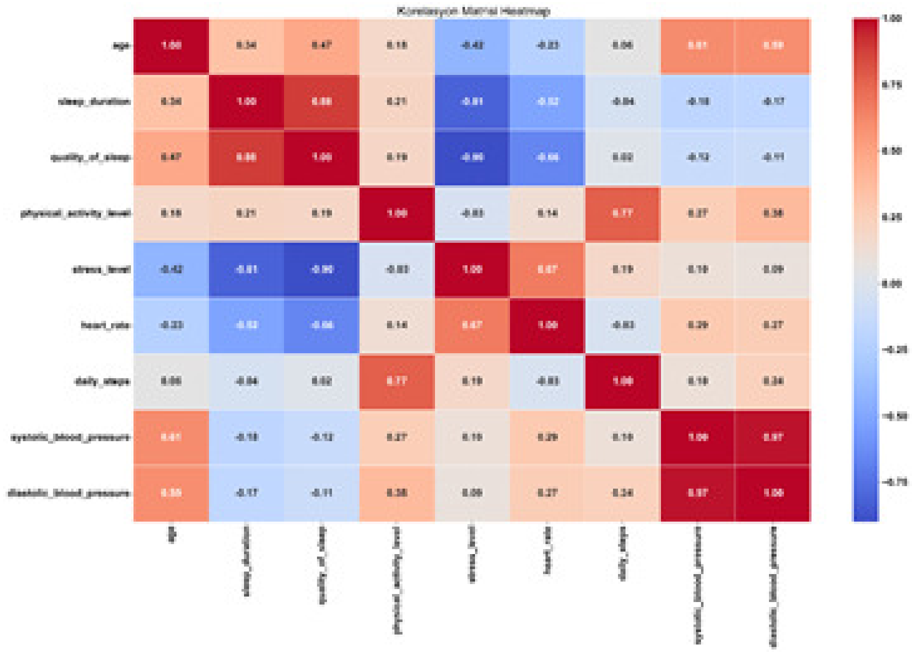

Figure 1. Heatmap of the Dataset

Figure 2 displays the charts of the important variables in the Sleep Health and Lifestyle Dataset, thus presenting a visual depiction of their distributions and relationships. The explicit analysis of these plots will help us to find the trends, the patterns and the potential outliers for different variables like sleep duration, physical activity levels, stress levels, and BMI categories. Through an image-based comparison of these variables, we can find out the possible associations and implications that can be used for the feature selection and model development techniques.

3.2. Feature Selection

Feature selection is the pivotal process in predictive modeling that consists of the extraction of the most important subset of features from the original dataset. Feature selection is aimed at enhancing the model performance by lowering the dimensionality, avoiding overfitting, boosting computational efficiency, and boosting interpretability. The modeling process is thus streamlined as a result of the selection of the most informative features that are done by the feature selection, which in turn leads to the production of more efficient algorithms and higher predictive accuracy [26, 27, 28].

Enhanced Model Performance: Feature selection is based on the most important features that make

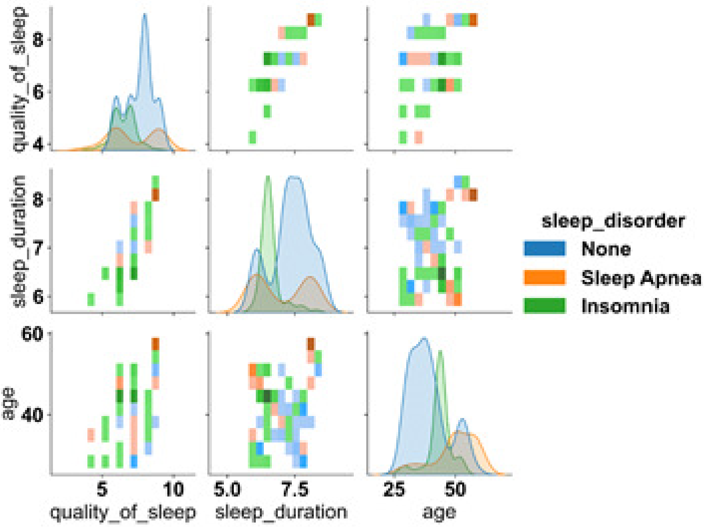

Figure 2. Comparison Plots of the Dataset

the model simpler and more interpretable, and at the same time, it decreases the noise and useless information from the dataset. Thus, the result is a greater generalization to the data that has yet to be seen and an improvement in the predictive accuracy of the model.

- Prevention of Overfitting: Eliminating redundant or superfluous features will greatly help prevent the model from overfitting, which is like a mode that learns noise in the training data rather than reading the underlying patterns. Feature selection is the act of choosing the factors that the model will use for predicting, which, in turn, makes the model generalize well to new data and does not memorize the training data.
- Reduction of Computational Complexity: The feature selection reduces the dimension of the dataset, so the model learning and inference time is faster. The important features that the algorithm identifies become the ones that the computer processes; thus, the computer uses its resources more efficiently, which allows the analysis of bigger datasets and more complicated models.
- Improved Interpretability: The models created by using a smaller number of features are easier to explain and comprehend, making it easier to interpret the main causes of the predictions. This, in turn, improves the model's transparency and trustworthiness, providing the stakeholders with the necessary data to make the right decisions based on the model's outputs.
  Binary optimization algorithms are specific methods aimed at tackling feature selection problems most efficiently. These algorithms repeatedly choose a part of the original feature space to add to the final output, with the purpose of maximizing a preset objective function, for example, model accuracy or predictive performance. Below, we provide a detailed description of the binary optimization algorithms commonly used for feature selection:
- bDTO (Binary Dipper Throated Optimization Algorithm): The bDTO emulates the Dipper Throated bird's behavior, and hence, it is able to select the features that are the most useful and that are the most informative from the dataset by means of a fitness function that assesses the feature subsets based on their predictive performance.
- bGWO (Binary Grey Wolf Optimizer): The bGWO is based on the Grey Wolf Optimizer, which reproduces grey wolf hunting activity to find the most appropriate features. The algorithm, on the

other hand, repeatedly modifies the wolves' locations in the search space to reduce the distance to the optimal feature subset.

- bGWO_PSO (Binary Grey Wolf Optimizer Particle Swarm Optimization): This fusion of the Grey Wolf Optimizer and Particle Swarm Optimization aims to improve feature selection performance by combining the best of both algorithms.
- bPSO (Binary Particle Swarm Optimization): The bPSO process is inspired by the collective behavior of bird flocks or fish schools. This means that the particles in the search space are updated based on the individual and collective experiences of the group, and therefore, the optimal subset of features is found.
- bBA (Binary Bat Algorithm): BBA involves the process of echolocation like that of bats, where the modeled random walk and frequency tuning are used to explore the search space and select features based on their fitness values.
- bWOA (Binary Whale Optimization Algorithm): Humpback whales, due to their predatory hunting activities, have inspired the bWOA algorithm, which uses a population-based approach to iteratively search for the best subset of features by imitating whale exploration and exploitation strategies.
- bBBO (Binary Biogeography-based Optimizer): According to the principles of biogeography, b-BBO treats features as islands and applies migration and mutation operations to iteratively select the features that are relevant to the given fitness values.
- bMVO (Binary Multiverse Optimization): BMVO is inspired by the idea of multiple universes. Thus, it creates several search spaces at the same time, which allows it to efficiently search for the best combination of features using the principles of quantum mechanics.
- FSBO (Binary Bowerbird Optimizer): BSBO uses the same strategy as bowerbirds for nesting and models feature subsets based on their attractiveness. Its main purpose is to find the most informative subset of features through iterative refinement.
- bGWO_GA (Binary Grey Wolf Optimizer Genetic Algorithm): This hybrid method combines the Grey Wolf Optimizer's social hierarchy and the Genetic Algorithm's genetic operators. Introducing crossover and mutation significantly enhances feature selection performance.
- bFA (Binary Firefly Algorithm): The bFA draws inspiration from the behavior of fireflies, using light intensity and attractiveness as guiding principles. It mimics the movement of fireflies in the search space, iteratively searching for the best subset of features.
- bGA (Binary Genetic Algorithm): The conventional optimization method that is based on natural selection is the bGA, which creates candidate solutions (feature subsets) from one generation to another through selection, crossover, and mutation operations to improve the feature selection performance over time systematically.

These alternatives to binary optimization algorithms are used in different ways to select the most informative subset of features from a given dataset, each based on a different principle of nature or optimization. Through the study of the search space and the evaluation of the fitness of feature subsets, these algorithms are to be of the standard for creating models that are both sturdy and efficient throughout different domains and applications.

3.3. Machine Learning

The research conducted in our study was a thorough investigation of the different machine learning models that are used to provide accurate predictions of sleep disorders using the large Sleep Health and Lifestyle Dataset. Each model uses a different method, hence the combination of various algorithms and methodologies that are used to examine the relationship among sleep patterns, lifestyle factors and the presence of sleep disorders [29, 30, 31]. Here is an in-depth look at the machine-learning models we scrutinized:

1. Logistic Regression: Logistic regression, a linear model of the past, is often used for binary classification tasks. It is a statistical tool that estimates the probability of a binary outcome based on one or more predictor variables. Logistic regression, even if it is quite simple, can be used to model linear relationships and is especially helpful when the relationship between the predictors and the outcome is clear.
2. Decision Tree: Decision trees are online models that make decisions by recursively partitioning the feature space into regions, with each region denoting the class label. They produce decision-making processes that are based on common sense and can grasp the complexity of the relations among the features. Nevertheless, decision trees are susceptible to overfitting, even if the data is noisy or the data set is high-dimensional.
3. Random Forest: A Random Forest, an ensemble learning method, creates several decision trees and unites their forecasts to improve the accuracy and robustness of the prediction. By averaging the predictions of many trees, Random Forests reduce overfitting and often generalize well to data that was not in the training set. They are world-famous for their ability to process high-dimensional data and complex relationships between features.
4. Support Vector Machine (SVM): SVM, an impressive model of supervised learning, creates a hyperplane or a set of hyperplanes in a high-dimensional space where the classes are separated. It pursues the goal of expanding the gap between classes, thus being a good choice in cases where the decision boundaries are not linear and the data dimensionality is high. SVMs are used in many different fields and are the choice for classifying tasks.
5. K-Nearest Neighbors (KNN): KNN, a non-parametric classification method, assigns the new data points to the most common class of their k nearest neighbors in the feature space. KNN is easy to use but efficient, especially in instances where the decision boundary is nonlinear and nonlinear. It does not need to be trained and is also easily adapted to changes in the dataset.
6. Naive Bayes: Naive Bayes is a classifier that is based on the subject probability, and it assumes that the features are independent of the class label. Although Naive Bayes makes a lot of simple assumptions, it is still an efficient and admirably performed algorithm with high-dimensional data. It is most of the time beneficial for text classification and tasks with categorical features.
7. Neural Network (Multilayer Perceptron - MLP): MLP, a deep learning model with multiple layers of interconnected neurons, outperforms in learning the complex patterns in the data. MLPs are very flexible and skilled at picking up complicated relations between features. Thus, they are the ones that are fit for nonlinear decision boundaries and huge amounts of data. It is the special feature of these models that allows us to learn the deep hierarchical structure from raw data instantly.
8. Gradient Boosting: Gradient Boosting, an ensemble learning technique, develops a series of weak

learners successively, each of which corrects the errors of its predecessor. Through the process of continuing to refine the new models to the residues of the previous models, Gradient Boosting boosts the predictive accuracy and, at the same time, is good at handling complex datasets. It is commonly used in contests and real-life projects because it is the best and most reliable.

9. AdaBoost: AdaBoost, a part of the ensemble learning methods, combines a number of weak classifiers to create a strong classifier. It is focused on the cases that are dismissed by its predecessor; hence, it makes the model more efficient by adjusting the weights of the training instances. AdaBoost is well known for its simplicity and effectiveness, especially in situations where other algorithms fail to generalize.

Through a careful analysis of these various machine learning models, our aim is to find the best way to correctly estimate sleep disorders based on the provided sleep health and lifestyle information.

Evaluation Metrics Used for Model Performance Assessment:

We thoroughly assessed the performance of each machine learning model in predicting sleep disorders by using the whole range of evaluation metrics. The statistical measurements give information on the specific aspects of the model performance that, in turn, help the researchers to make a thorough evaluation of the predictive powers of the model. The evaluation metrics employed in our study encompass:

- Accuracy: The ratio of the correctly classified instances to the total number of instances, thus conveying the all-around measure of the model performance.
- Sensitivity (True Positive Rate): The ratio of true positive instances (sleep disorder cases correctly identified) to all actual positive instances, which means the model is able to detect sleep disorders.
- Specificity (True Negative Rate): The ratio of true negative instances (properly recognized non-sleep disorder cases) to all actual negative instances, which is evidence of the model's capability of correctly determining the non-disordered sleep.
- Positive Predictive Value (PPV): Also called precision, PPV is the term for the ratio of true positive instances to all instances that were predicted as positive. This ratio gives us an idea of how reliable positive predictions are.
- Negative Predictive Value (NPV): The ratio of the true negative instances to the total negative instances predicted is the measure of the accuracy of negative predictions.
- F1-Score: The F1-score is the harmonic mean of precision and recall and it is the optimal balance between the two, which is particularly useful for imbalanced datasets. It measures the model's accuracy in terms of the cases of both false positives and false negatives.

Through the examination of these evaluation metrics, we can acquire a complete view of the performance of each model and thus, determine the best way of predicting sleep disorders from the provided sleep health and lifestyle metrics.

4. Results

The feature selection process using a single-variable restriction least squares algorithm was one of the methods that yielded interesting results, as shown in Table Tab1. This process is of paramount importance because it allows the selection of the most appropriate features out of the dataset, thus reducing the dimensionality and improving the model performance. The comparison gives information about the average error (AE), the selected size, Average Select size (ASS), the best fitness (BF), the worst fitness (WF), the average fitness (AF), and the standard deviation fitness (SDF) across different algorithms.

Looking at the table, it's clear that the binary Dipper Throated Optimization Algorithm (bDTO) was the one that achieved the lowest average error of 0.71933, which signifies its ability to choose the right features that are beneficial while, at the same time, it also reduces the prediction errors. This algorithm, too, has a relatively low standard deviation of fitness, hence, showing the stability of its performance across the iterations. Moreover, bDTO had competitive average fitness, best fitness, and worst fitness values, which were better than those of the other algorithms, thus proving its ability to select features.

The performance evaluation of machine learning models used for the prediction of sleep disorders is shown in Table Tab2. The assessment is important for finding the best model that can correctly classify people based on sleep health and lifestyle statistics. The table encompasses five parameters: precision, sensitivity (True Positive Rate), specificity (True Negative Rate), Positive Predictive Value (PPV), Negative Predictive Value (NPV), and F1-score for each model.

The results clearly show that logistic regression is the most accurate method, with a result of 0.95 out of the assessed models, which got the best rating. This means it is a better predictor of sleep disorders based on the sleep health and lifestyle metrics provided. Logistic regression also had high sensitivity and specificity. Thus, it showed its effectiveness in the correct identification of both the positive and negative cases of sleep disorders.

Figure 3 illustrates graphical data of the accuracy scores made by different machine learning models in predicting sleep disorders. Accuracy, the ratio of correctly classified instances to the total number of instances, is a major criterion for the assessment of classification models' performance. Through the plot of the accuracy scores of the models, we can compare their effectiveness to accurately classify people based on sleep health and lifestyle metrics. This number is an important source of insights into the relative performance of the different models, and hence, it guides the selection of the most suitable one for the practical application.

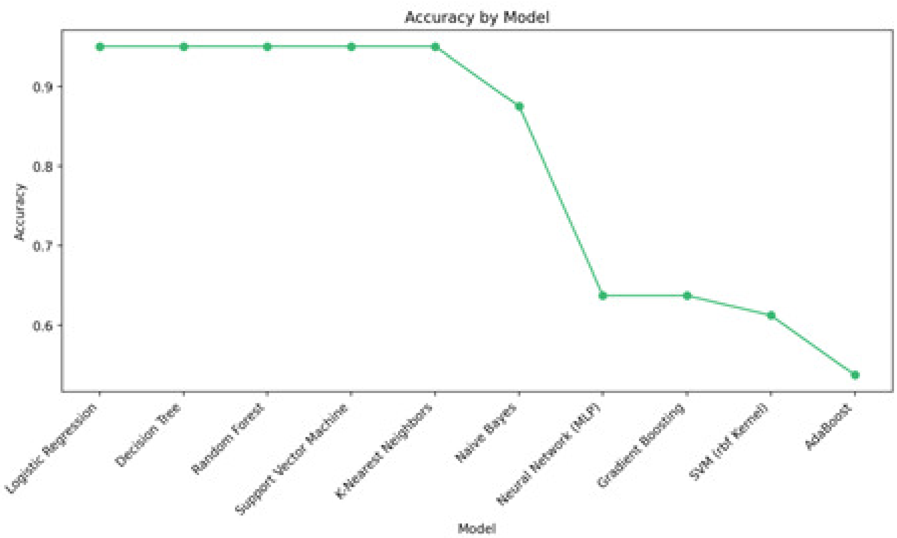

Figure 3. The Accuracy by Model

Table 1. Feature Selection Results

	bDTO	bGWO	bGWO_PSO	bPSO	bBA	bWAO	bBBO	bMVO	bSBO	bGWO_GA	bFA	bGA
AE	0.71933	0.73653	0.77583	0.77033	0.77993	0.77013	0.73853	0.74703	0.77863	0.75663	0.76873	0.75013
ASS	0.67213	0.87213	0.900543	0.87213	0.901153	0.803553	0.803593	0.96863	0.870424	0.79493	0.90663	0.81453
AF	0.78253	0.79873	0.80703	0.79713	0.82003	0.80493	0.80283	0.82683	0.83683	0.80483	0.84903	0.81013
BF	0.68433	0.71903	0.76053	0.77743	0.70973	0.76903	0.79253	0.75203	0.77993	0.78263	0.76773	0.71343
WF	0.78283	0.78593	0.87053	0.84513	0.81133	0.84513	0.87903	0.87003	0.85963	0.85883	0.86533	0.82853
SDF	0.60483	0.60953	0.62773	0.60893	0.61883	0.61113	0.65383	0.65963	0.66983	0.61013	0.64573	0.61113

Table 2. Machine Learning Models Results

Models	Accuracy	Sensitivity (TRP)	Specificity (TNP)	Pvalue PPV	Pvalue NPV	FScore
Logistic Regression	0.95	0.916666667	0.977272727	0.970588	0.934783	0.942857
Decision Tree	0.95	0.916666667	0.977272727	0.970588	0.934783	0.942857
Random Forest	0.95	0.916666667	0.977272727	0.970588	0.934783	0.942857
Support Vector Machine	0.95	0.916666667	0.977272727	0.970588	0.934783	0.942857
K-Nearest Neighbors	0.95	0.916666667	0.977272727	0.970588	0.934783	0.942857
Naive Bayes	0.875	0.916666667	0.840909091	0.825	0.925	0.868421
Neural Network (MLP)	0.6375	1	0.340909091	0.553846	1	0.712871
Gradient Boosting	0.6375	1	0.340909091	0.553846	1	0.712871
SVM (rbf Kernel)	0.6125	1	0.295454545	0.537313	1	0.699029
AdaBoost	0.5375	1	0.159090909	0.493151	1	0.66055

Figure 4 depicts the F1 scores attained by different machine learning models in determining sleep disorders. The F1 score, which is the combination of precision and recall, gives a balanced way of evaluating the model's performance in the case of the imbalanced class distribution. Through the F1 scores of each model, we can measure their potential to have both high precision and recall in the identification of people with sleep disorders. This number gives the audience an understanding of the overall effectiveness of each model, and it helps the decision-makers choose the best model for real-world implementation.

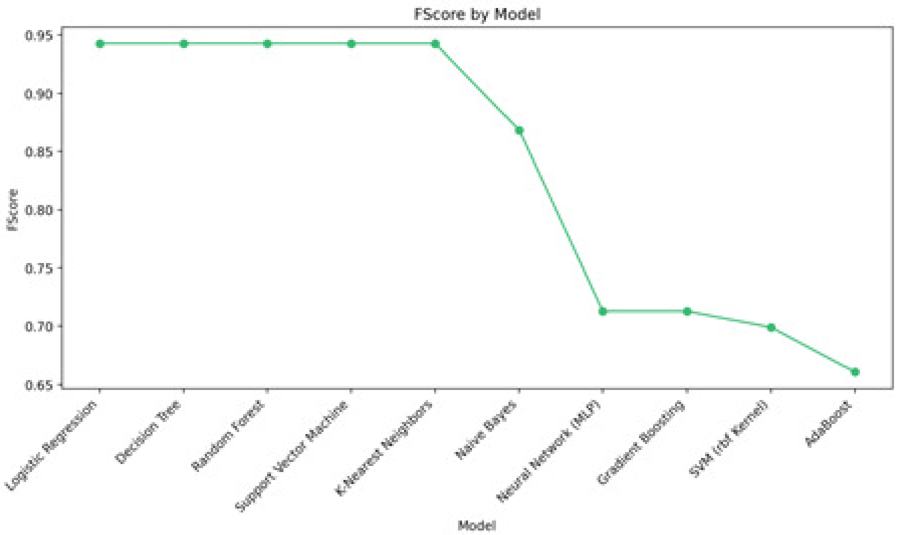

Figure 4. F Score by Model

The outcome proves the binary Dipper Throated Optimization Algorithm (bDTO) in the feature selection, because it achieved the lowest average error. This is the evidence that this can find the necessary data to establish the reliable relationship while at the same time error the prediction. Besides, the logistic regression grew to be the most accurate model for predicting something. sleep disorders, and an accuracy of 0.95. Thus, the above point stresses the significance of the use of the suitable feature selection techniques and machine learning algorithms for the creation of the predictive models in sleep health research that are reliable. More research on these findings would make possible the development of new techniques for the recognition and treatment of sleep disorders, and this in turn would have a positive impact on the general health of the population.

5. Conclusion

This research has enlightened us on the possibilities of machine learning approaches in the advancement of our knowledge of the sleep health and lifestyle factors. The analysis of the Sleep Health and Lifestyle Dataset led to the discovery of several important results which gave insights into both the algorithms of feature selection and the model of predictive for the classification of sleep disorder.

First of all, the study showed that the binary Dipper Throated Optimization Algorithm (bDTO) is a particular effective feature selection algorithm, for it could reach the lower average error among the algorithms tested. This indicates that bDTO has the potential to be used for the identification of the most necessary features from the dataset, thus, the outcome of the predictive models will be more reliable. Besides, the procedure of machine learning models for sleep disorders prediction also showed that logistic regression was the most accurate model, with an amazing accuracy of 0. 95. This proves the fact that logistic regression is capable of correctly classifying people according to their sleep health and lifestyle analysis, thus healthcare professionals can use this tool to detect and intervene in sleep problems early.

The conclusions of these studies are important for healthcare professionals, scientists, and everyone else. Hence, the use of predictive modeling will help healthcare practitioners recognize those who are at risk of having sleep disorders in the initial stages, which will enable them to create customized interventions and treatment plans. Besides, the researchers can make use of these findings to deepen the investigation of the complex connection between the state of sleep, the way of life and health in general. Nevertheless, it is significant to recognize the shortcomings of this study. The machine learning models and feature selection algorithms chosen for the research have displayed the potential for excellent results, yet there is still the possibility for more improvement. Research in the future could be directed towards the addition of more variables or data sources to amplify the model performance even more. Besides, the study of the more advanced machine learning techniques might offer the researchers a more profound understanding of the sleep health and lifestyle interactions of the subjects.

Mainly, this study shows how machine learning can be of great help in the progress of sleep health research and the personalized prevention of sleep disorders in the people susceptible to this. Through the adoption of these cutting-edge methods, we can work towards a future where sleep disorders are detected early, and their management is easy and effective, which will result in the improvement of the quality of life of people around the world.

Funding: “This research received no external funding”

Conflicts of Interest: “The authors declare no conflict of interest.”

References

1. Anbarasi, L. J., Jawahar, M., Ravi, V., Cherian, S. M., Shreenidhi, S., & Sharen, H. (2022). Machine learning approach for anxiety and sleep disorders analysis during COVID-19 lockdown. Health and Technology, 12(4), 825–838. https://doi.org/10.1007/s12553-022-00674-7

2. Arora, A., Chakraborty, P., & Bhatia, M. P. S. (2020). Analysis of Data from Wearable Sensors for Sleep Quality Estimation and Prediction Using Deep Learning. Arabian Journal for Science and Engineering, 45(12), 10793–10812. https://doi.org/10.1007/s13369-020-04877-w

3. El Sherbini, A., Rosenson, R. S., Al Rifai, M., Virk, H. U. H., Wang, Z., Virani, S., Glicksberg, B. S., Lavie, C. J., & Krittanawong, C. (2024). Artificial intelligence in preventive cardiology. Progress in Cardiovascular Diseases. https://doi.org/10.1016/j.pcad.2024.03.002

4. Kamel Boulos, M. N., & Koh, K. (2021). Smart city lifestyle sensing, big data, geo-analytics and intelligence for smarter public health decision-making in overweight, obesity and type 2 diabetes prevention: The research we should be doing. International Journal of Health Geographics, 20(1), 12. https://doi.org/10.1186/s12942-021-00266-0

5. Kang, C., An, S., Kim, H. J., Devi, M., Cho, A., Hwang, S., & Lee, H. W. (2023). Age-integrated artificial intelligence framework for sleep stage classification and obstructive sleep apnea screening. Frontiers in Neuroscience, 17. https://doi.org/10.3389/fnins.2023.1059186

6. Kourou, K., Manikis, G., Poikonen-Saksela, P., Mazzocco, K., Pat-Horenczyk, R., Sousa, B., Oliveira-Maia, A. J., Mattson, J., Roziner, I., Pettini, G., Kondylakis, H., Marias, K., Karademas, E., Simos, P., & Fotiadis, D. I. (2021). A machine learning-based pipeline for modeling medical, socio-demographic, lifestyle and self-reported psychological traits as predictors of mental health outcomes after breast cancer diagnosis: An initial effort to define resilience effects. Computers in Biology and Medicine, 131, 104266. https://doi.org/10.1016/j.compbiomed.2021.104266

7. Manocha, A., Kumar, G., Bhatia, M., & Sharma, A. (2023). IoT-inspired machine learning-assisted sedentary behavior analysis in smart healthcare industry. Journal of Ambient Intelligence and Humanized Computing, 14(5), 5179–5192. https://doi.org/10.1007/s12652-021-03371-x

8. Oyebode, O., Fowles, J., Steeves, D., & Orji, R. (2023). Machine Learning Techniques in Adaptive and Personalized Systems for Health and Wellness. International Journal of Human–Computer Interaction, 39(9), 1938–1962. https://doi.org/10.1080/10447318.2022.2089085

9. Safaei, M., Sundararajan, E. A., Driss, M., Boulila, W., & Shapi'i, A. (2021). A systematic literature review on obesity: Understanding the causes & consequences of obesity and reviewing various machine learning approaches used to predict obesity. Computers in Biology and Medicine, 136, 104754. https://doi.org/10.1016/j.compbiomed.2021.104754

10. Schwartz, A. R., Cohen-Zion, M., Pham, L. V., Gal, A., Sowho, M., Sgambati, F. P., Klopfer, T., Guzman, M. A., Hawks, E. M., Etzioni, T., Glasner, L., Druckman, E., & Pillar, G. (2020). Brief digital sleep questionnaire powered by machine learning prediction models identifies common sleep disorders. Sleep Medicine, 71, 66–76. https://doi.org/10.1016/j.sleep.2020.03.005

11. Singh, B., Olds, T., Brinsley, J., Dumuid, D., Virgara, R., Matricciani, L., Watson, A., Szeto, K., Eglitis, E., Miatke, A., Simpson, C. E. M., Vandelanotte, C., & Maher, C. (2023). Systematic review and meta-analysis of the effectiveness of chatbots on lifestyle behaviours. Npj Digital Medicine, 6(1), 1–10. https://doi.org/10.1038/s41746-023-00856-1

12. Subramanian, M., Wojtusciszyn, A., Favre, L., Boughorbel, S., Shan, J., Letaief, K. B., Pitteloud, N., & Chouchane, L. (2020). Precision medicine in the era of artificial intelligence: Implications in chronic disease management. Journal of Translational Medicine, 18(**??**), 472. https://doi.org/10.1186/s12967-020-02658-5

13. Sujith, A. V. L. N., Sajja, G. S., Mahalakshmi, V., Nuhmani, S., & Prasanalakshmi, B. (2022). Systematic review of smart health monitoring using deep learning and Artificial intelligence. Neuroscience Informatics, 2(3), 100028. https://doi.org/10.1016/j.neuri.2021.100028

14. Tiwari, S., Chandra, R., & Agarwal, S. (2023). An Optimized Hybrid Solution for IoT Based Lifestyle Disease Classification Using Stress Data. In M. Tanveer, S. Agarwal, S. Ozawa, A. Ekbal, & A. Jatowt (Eds.), Neural Information Processing (pp. 433–445). Springer Nature. https://doi.org/10.1007/978-981-99-1648-1_36

15. Arora, A., Chakraborty, P., & Bhatia, M. P. S. (2022). Intervention of Wearables and Smartphones in Real Time Monitoring of Sleep and Behavioral Health: An Assessment Using Adaptive Neuro-Fuzzy Technique. Arabian Journal for Science and Engineering, 47(2), 1999–2024. https://doi.org/10.1007/s13369-021-06078-5

16. Bandyopadhyay, A., & Goldstein, C. (2023). Clinical applications of artificial intelligence

in sleep medicine: A sleep clinician's perspective. Sleep and Breathing, 27(1), 39–55. https://doi.org/10.1007/s11325-022-02592-4

17.Chattu, V. K., Sunil, T. S., Santaji, S., Desai, V. V., Garg, L., Nurtazina, A., Allahverdipour, H., & Pandi-Perumal, S. R. (2021). Precision Medicine Meets Sleep Medicine: How Can Sleep Health Aid to Reduce the Preventable Burden of Non-communicable Diseases? Sleep and Vigilance, 5(2), 179–188. https://doi.org/10.1007/s41782-021-00149-1

18.Cho, P. J., Singh, K., & Dunn, J. (2021). Chapter 9—Roles of artificial intelligence in wellness, healthy living, and healthy status sensing. In L. Xing, M. L. Giger, & J. K. Min (Eds.), Artificial Intelligence in Medicine (pp. 151–172). Academic Press. https://doi.org/10.1016/B978-0-12-821259-2.00009-0

19.Djanian, S., Bruun, A., & Nielsen, T. D. (2022). Sleep classification using Consumer Sleep Technologies and AI: A review of the current landscape. Sleep Medicine, 100, 390–403. https://doi.org/10.1016/j.sleep.2022.09.004

20.Huchaiah, M. D., & Kasubi, J. W. (2021). Predicting of Sleep Behaviour in Smart Homes Based on Multi-residents Using Machine Learning Techniques. SN Computer Science, 2(4), 244. https://doi.org/10.1007/s42979-021-00643-3

21.Kim, J., Mun, S., Lee, S., Jeong, K., & Baek, Y. (2022). Prediction of metabolic and pre-metabolic syndromes using machine learning models with anthropometric, lifestyle, and biochemical factors from a middle-aged population in Korea. BMC Public Health, 22(1), 664. https://doi.org/10.1186/s12889-022-13131-x

22.Perez-Pozuelo, I., Spathis, D., Clifton, E. A. D., & Mascolo, C. (2021). Chapter 3—Wearables, smartphones, and artificial intelligence for digital phenotyping and health. In S. Syed-Abdul, X. Zhu, & L. Fernandez-Luque (Eds.), Digital Health (pp. 33–54). Elsevier. https://doi.org/10.1016/B978-0-12-820077-3.00003-1

23.Perez-Pozuelo, I., Zhai, B., Palotti, J., Mall, R., Aupetit, M., Garcia-Gomez, J. M., Taheri, S., Guan, Y., & Fernandez-Luque, L. (2020). The future of sleep health: A data-driven revolution in sleep science and medicine. Npj Digital Medicine, 3(1), 1–15. https://doi.org/10.1038/s41746-020-0244-4

24.Watson, N. F., & Fernandez, C. R. (2021). Artificial intelligence and sleep: Advancing sleep medicine. Sleep Medicine Reviews, 59, 101512. https://doi.org/10.1016/j.smrv.2021.101512

25.Sleep Health and Lifestyle Dataset. (n.d.). Retrieved May 15, 2024, from https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset

26.Al-Tashi, Q., Md Rais, H., Abdulkadir, S. J., Mirjalili, S., & Alhussian, H. (2020). A Review of Grey Wolf Optimizer-Based Feature Selection Methods for Classification. In S. Mirjalili, H. Faris, & I. Aljarah (Eds.), Evolutionary Machine Learning Techniques: Algorithms and Applications (pp. 273–286). Springer. https://doi.org/10.1007/978-981-32-9990-0_13

27.Chen, R.-C., Dewi, C., Huang, S.-W., & Caraka, R. E. (2020). Selecting critical features for data classification based on machine learning methods. Journal of Big Data, 7(1), 52. https://doi.org/10.1186/s40537-020-00327-4

28.Dhal, P., & Azad, C. (2022). A comprehensive survey on feature selection in the various fields of machine learning. Applied Intelligence, 52(4), 4543–4581. https://doi.org/10.1007/s10489-021-02550-9

29.Odhiambo Omuya, E., Onyango Okeyo, G., & Waema Kimwele, M. (2021). Feature Selection for Classification using Principal Component Analysis and Information Gain. Expert Systems with Applications, 174, 114765. https://doi.org/10.1016/j.eswa.2021.114765

30.Toğaçar, M., Ergen, B., Cömert, Z., & Özyurt, F. (2020). A Deep Feature Learning Model for Pneumonia Detection Applying a Combination of mRMR Feature Selection and Machine Learning Models. IRBM, 41(4), 212–222. https://doi.org/10.1016/j.irbm.2019.10.006

31.Zhao, H., Liu, Z., Yao, X., & Yang, Q. (2021). A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach. Information Processing & Management, 58(5), 102656. https://doi.org/10.1016/j.ipm.2021.102656

SASAR
Open Access Journal