Predictive Modelling Approach for Sleep Disorder using Sleep Health and Lifestyle Properties

Dr. N. Satyavathi[1*], Kandagatla Karthik[2], Gummula Sai Vamshi[2], Bolla Shivasai[2]

[1]Associate Professor & Head, [2]UG Student, [1,2]Department of Computer Sciences and Engineering,

[1,2]Vaagdevi College of Engineering (UGC - Autonomous), Bollikunta, Warangal, Telangana.

[*]Corresponding author: Dr. N. Satyavathi (satyanadendla15@gmail.com)

ABSTRACT

The Sleep Health and Lifestyle Dataset is a comprehensive collection of sleep and lifestyle-related variables for 400 individuals, providing valuable insights into sleep patterns, daily habits, and potential sleep disorders. By analyzing physical activity levels, stress, and BMI categories, healthcare professionals can design personalized lifestyle interventions to improve overall health and well-being. The presence or absence of sleep disorders, such as Insomnia and Sleep Apnea, allows for the identification of individuals at risk and informs targeted diagnostic and therapeutic strategies. Previous research on sleep health and lifestyle factors often relies on self-reported data, which may introduce biases and inaccuracies. Subjective sleep quality assessments may not capture objective sleep measurements accurately. Additionally, existing studies lack comprehensive datasets, hindering the ability to analyze multiple factors simultaneously. To address the limitations of existing research, this work proposes a multi-faceted analysis approach. Develop a machine learning model to predict the presence of sleep disorders based on a combination of sleep-related and lifestyle variables. The model can assist in early detection and intervention. The dataset allows for the investigation of sleep duration, quality, and factors influencing sleep patterns, enabling researchers to identify trends and correlations related to sleep health.

Keywords: Sleep Disorders, Predictive Modeling, Machine Learning, Sleep Health, Lifestyle Factors, Insomnia

1. INTRODUCTION

The research topic, "Predictive Modeling Approach for Sleep Disorder using Sleep Health and Lifestyle Properties," represents a pioneering effort at the crossroads of healthcare, data science, and predictive analytics. Sleep disorders, including insomnia, sleep apnea, and restless legs syndrome, are prevalent and can have a profound impact on individuals' physical and mental well-being. This research embarks on a journey to develop predictive models that leverage sleep health data and lifestyle properties to proactively identify and manage sleep disorders [1]. The motivation for this research is deeply rooted in the widespread occurrence of sleep disorders and their far-reaching consequences [2]. In the modern world, factors such as stress, irregular work schedules, and lifestyle choices have contributed to an increase in sleep-related issues. These disorders not only affect an individual's daily functioning but are also linked to a range of health problems, including cardiovascular diseases, diabetes, and mental health disorders. Hence, there is a compelling need for advanced predictive tools that can enable early detection, personalized intervention, and improved sleep health outcomes [3]. To address this pressing need, the research delves into the development of predictive models using machine learning and data analytics techniques [4]. These models analyze a multitude of sleep health metrics, lifestyle properties, and contextual factors, which collectively offer a holistic view of an individual's sleep patterns and behaviors. The outcome is a predictive framework that can forecast the likelihood of sleep disorders, guiding healthcare providers and individuals in making informed decisions for preventive care and tailored interventions [5].

Furthermore, the research underscores the ethical dimension of technology deployment in healthcare. It emphasizes the importance of patient data privacy, ethical data usage, and responsible AI practices to ensure that the benefits of predictive modeling in sleep disorder management align with ethical standards and patient well-being [6]. In this introductory overview, we will delve into the key components and objectives of this research. We will explore the challenges posed by sleep disorders in the modern era, introduce the role of predictive modeling and data-driven insights, and highlight the transformative potential of this research in enhancing sleep health. Additionally, we will underscore the ethical considerations and real-world applications of this research, which extend across clinical sleep medicine, telehealth, and lifestyle interventions [7]. The "Predictive Modeling Approach for Sleep Disorder using Sleep Health and Lifestyle Properties" signifies a crucial initiative to harness the power of predictive analytics in addressing the global challenge of sleep disorders. By developing data-driven models for early detection and personalized intervention, this research aims to improve sleep health outcomes while upholding ethical standards and responsible technology use, ultimately contributing to better sleep health for individuals and communities worldwide [8].

2. LITERATURE SURVEY

Lee, et al. [11] proposed Machine learning-based predictive modelling of depression in hypertensive populations. This cross-sectional study included 8,628 adults with hypertension (11.3% with depression) from the National Health and Nutrition Examination Survey (2011–2020). We selected several significant features using feature selection methods to build the models. Data imbalance was managed with random down-sampling. Six different ML classification methods implemented in the R package caret—artificial neural network, random forest, AdaBoost, stochastic gradient boosting, XGBoost, and support vector machine were employed with 10-fold cross-validation for predictions. Byeon, et al. [12] proposed a predictive model for depressive disorders using a stacking ensemble and naive Bayesian nomogram. This study explored the main risk factors of depressive disorders using the stacking ensemble machine technique. Moreover, this study developed a nomogram that could help primary physicians easily interpret high-risk groups of depressive disorders in primary care settings based on the major predictors derived from machine learning.

Thong, et al. [13] proposed Telehealth Technology Application in Enhancing Continuous Positive Airway Pressure Adherence in Obstructive Sleep Apnea Patients. Continuous positive airway pressure (CPAP) is the first choice for moderate-severe OSA but poor compliance brings a great challenge to its effectiveness. Telehealth interventions ease the follow-up process and allow healthcare facilities to provide consistent care. Fifth-generation wireless transmission technology has also greatly rationalized the wide use of telemedicine. Alramadeen, et al. [14] proposed a sparse linear mixed model that combines the modified Cholesky decomposition with group lasso penalties to enable joint group selection of fixed effects and random effects. A novel Expectation Maximization (EM) algorithm integrated with an efficient Majorization Maximization (MM) algorithm is developed for model estimation of the proposed sparse linear mixed model with group variable selection. The proposed method was applied to the SHHS data for telemonitoring and diagnosis of sleep disorder and found that a few significant feature groups are consistent with prior medical studies on sleep disorder.

Kim, et al. [15] proposed a Prediction of metabolic and pre-metabolic syndromes using machine learning models with anthropometric, lifestyle, and biochemical factors from a middle-aged population in Korea. Early prediction of the risk of MetS in the middle-aged population provides greater benefits for cardiovascular disease-related health outcomes. This study aimed to apply the latest machine learning techniques to find the optimal MetS prediction model for the middle-aged Korean population. Tsai, et al. [16] proposed Machine learning approaches for screening the risk of obstructive sleep apnea in the Taiwan population based on body profile. This study developed new models for screening the risk of moderate-to-severe OSAS (apnea-hypopnea index, AHI ≥15) and severe OSAS (AHI ≥30) in various age groups and sexes by using anthropometric features in the Taiwan population. Patients' characteristics – namely age, sex, body mass index (BMI), neck circumference, and waist circumference – were obtained.

Arafa, et al. [17] proposed Sleep duration and atrial fibrillation risk in the context of predictive, preventive, and personalized medicine. The authors investigated the association between sleep duration and AF risk using a prospective cohort study and a meta-analysis of epidemiological evidence. The Cox regression was used to compute the hazard ratios (HRs) and 95% confidence intervals (CIs) of AF risk for daily sleep ≤ 6 (short sleep), ≥ 8 (long sleep), and irregular sleep, including night-shift work compared with 7 h (moderate sleep). Then, we combined our results with those from other eligible prospective cohort studies in two meta-analyses for the short and long sleep. Gomes, et al. [18] proposed Predicting depressive symptoms in middle-aged and elderly adults using sleep data and clinical health markers. We used the National Health and Nutrition Examination Survey (NHANES) 2015–2016 data. Eighteen variables on self-reported physical activity, self-reported sleep habits, sleep disturbance indicative, anthropometric measurements, sociodemographic characteristics and plasma biomarkers of obesity and diabetes were selected as predictors. A total of 2907 middle-aged and elderly subjects were eligible for the study. Supervised learning algorithms such as Lasso penalized Logistic Regression (LR), Random Forest (RF) and Extreme Gradient Boosting (XGBoost) were implemented.

Mehra, et al. [19] proposed Sleep-disordered breathing and cardiac arrhythmias in adults. Day-night patterning and circadian biology of SDB-induced pathophysiological sequelae collectively influence the structural and electrophysiological cardiac substrate, thereby creating an ideal milieu for arrhythmogenic propensity. Immediate consequences of SDB include autonomic nervous system fluctuations, recurrent hypoxia, alterations in carbon dioxide/acid-base status, disrupted sleep architecture, and accompanying increases in negative intrathoracic pressures directly affecting cardiac function. Tsai, et al. [20] proposed Screening for obstructive sleep apnea risk by using machine learning approaches and anthropometric features. They developed models based on the following machine learning approaches: logistic regression, k-nearest neighbours, naïve Bayes, random forest (RF), support vector machine, and XGBoost. Collected data were first independently split into two data sets (training and validation: 80%; testing: 20%). Thereafter, we adopted the model with the highest accuracy in the training and validation stage to predict the testing set.

Cai, et al. [21] proposed Associations of the cardiometabolic index with the risk of cardiovascular disease in patients with hypertension and obstructive sleep apnea. The authors explore the relationship between the cardiometabolic index (CMI) and cardiovascular disease (CVD) and its subtypes (coronary artery disease and stroke) in patients with hypertension and obstructive sleep apnea (OSA). Methods. We conducted a retrospective cohort study enrolling 2067 participants from the Urumqi Research on Sleep Apnea and Hypertension study. Hussain, et al. [22] proposed a Quantitative evaluation of EEG biomarkers for the prediction of sleep stages. This work aims to quantify the neurological EEG biomarkers and predict five-class sleep stages using sleep EEG data. We investigated the three-channel EEG sleep recordings of 154 individuals (mean age of 53.8 ± 15.4 years) from the Haaglanden Medisch Centrum (HMC, The Hague, The Netherlands) open-access public dataset of PhysioNet. The power of fast-wave alpha, beta, and gamma rhythms decreases; and the power of slow-wave delta and theta oscillations gradually increases as sleep becomes deeper.

Streckenbach, et al. [23] proposed Validating Discriminative Signatures for Obstructive Sleep Apnea in Exhaled Breath. The authors validated this diagnostic approach and the proposed marker compounds, as well as their potential to reliably diagnose OSA. In this cross-sectional observational study, exhaled breath was analyzed using secondary electrospray ionization high-resolution mass spectrometry. The study cohort included untreated OSA patients, OSA patients treated with continuous positive airway pressure and healthy subjects. Peltz, et al. [24] proposed psychological processes linking problematic smartphone use to sleep disturbance in young adults. One potential factor underlying these problems is problematic smartphone use, which is defined as excessive phone use, impulse control problems related to the use, and negative consequences stemming from these behaviors. The 2-wave (baseline and 2-month follow-up) online sample consisted of 385 undergraduates (81% female; M = 20.0, SD = 1.6), who reported problematic smartphone use, psychological flexibility, anxiety symptoms, and sleep disturbance. Wang, et al. [25] proposed That obesity-related dietary pattern is associated with higher risk of sleep disorders. Obesity-related dietary patterns explaining most variance in waist circumference and BMI simultaneously were extracted from twenty-six food groups by the using partial least squares method. Sleep disorder and sleep duration, which were defined by self-reported questions, were the primary and the secondary outcome, respectively. Generalized linear models were performed to estimate the association of sleep disorders and sleep duration with dietary patterns.

3. PROPOSED METHODOLOGY

The project's procedure starts with data preprocessing to prepare the dataset for modeling, including data collection, cleaning, feature engineering, and data splitting. It then proceeds to compare two different machine learning models, the existing Random Forest Classifier (RFC) and the proposed Decision Tree Classifier (DTC), for their effectiveness in predicting sleep disorders based on sleep health and lifestyle properties. The choice between these models aims to identify the most suitable algorithm for accurate sleep disorder prediction. Figure 4.1 shows the proposed system model. The detailed operation illustrated as follows:

Step 1: Preprocessing

The project begins with data preprocessing, a crucial step in data analysis and modeling. This involves several sub-steps, starting with data collection, where relevant data related to sleep health and lifestyle properties is gathered. This may include variables such as sleep duration, sleep quality, exercise habits, diet, stress levels, and more. Following this, data cleaning is performed to identify and handle missing values, outliers, or errors in the dataset, ensuring the data is accurate and reliable for modeling. Feature engineering is then applied to select and construct variables that are likely to be informative for predicting sleep disorders, which may involve transforming or creating new features. Data transformation is also carried out, including scaling or normalizing features to bring them to a common scale suitable for modeling. Finally, the dataset is split into training and testing subsets to facilitate effective model training and evaluation.

Step 2: Random Forest Classifier

In this step, an existing Random Forest Classifier (RFC) model is used for predictive modeling of sleep disorders. RFC is a popular machine learning algorithm known for its ensemble learning capabilities. The existing RFC model is trained on the preprocessed dataset using the selected features. During training, the model learns patterns and relationships between sleep health and lifestyle properties and sleep disorder outcomes. Model performance is evaluated using appropriate metrics such as accuracy, precision, recall, and F1-score to assess its predictive capabilities.

Step 3: Proposed Decision Tree Classifier

In contrast to the existing RFC model, a Decision Tree Classifier (DTC) is proposed for predictive modeling in this step. DTC is a different machine learning algorithm that builds a tree-like structure to make decisions. The proposed DTC model is trained on the same preprocessed dataset, using the same features as the RFC model. The goal is to compare the performance of the DTC model with the existing

RFC model. Similar to the RFC model, the performance of the DTC model is evaluated using appropriate evaluation metrics to determine its effectiveness in predicting sleep disorders based on sleep health and lifestyle properties.

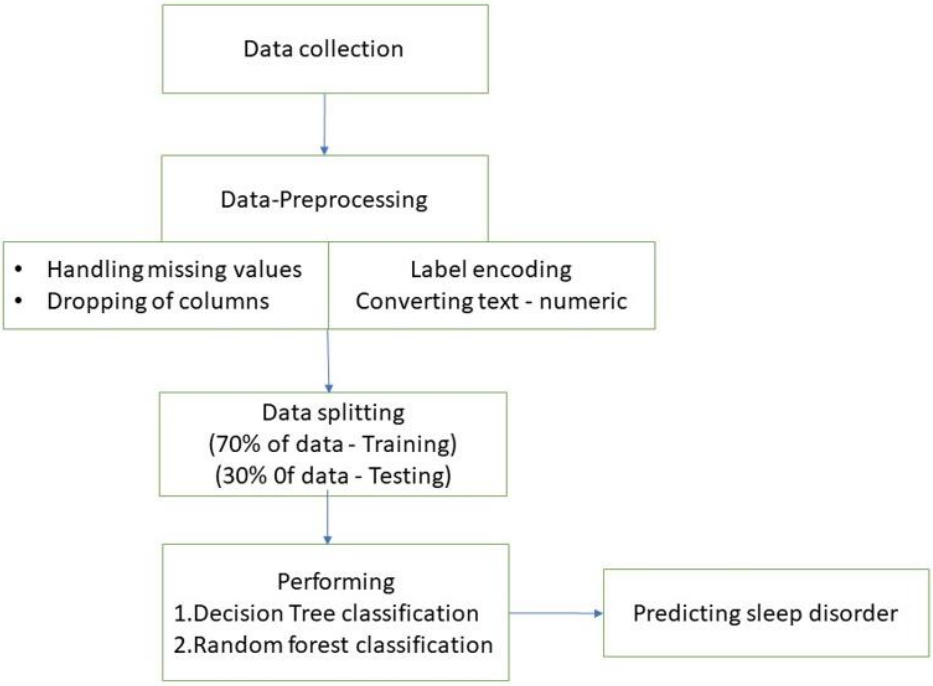


Fig. 1: Block diagram of proposed system.

3.2 Data Preprocessing

Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always a case that we come across clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put it in a formatted way. So, for this, we use data pre-processing tasks. Real-world data generally contains noises, missing values, and maybe in an unusable format which cannot be directly used for machine learning models. Data pre-processing requires tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model.

3.2.1 Standardization

Standard Scaler is applied to scale numeric features, ensuring that they have a mean of 0 and a standard deviation of 1. The StandardScaler from scikit-learn is used to standardize specific numeric features. Standardization is a common preprocessing step that brings features to a similar scale, which can enhance the performance of many machine learning algorithms. This transformation is important for several reasons. First, it ensures equal scaling, which is crucial for algorithms sensitive to feature magnitude, such as gradient-based optimization methods used in neural networks and distance-based algorithms like k-means clustering. Second, by subtracting the mean from each data point, it centers the data around zero, helping models converge faster and perform better during training. Third, scaling by the standard deviation normalizes the data, allowing features to have comparable variances and preventing any single feature from dominating the model. Lastly, standardized data improves interpretability, as placing all features on a common scale makes it easier to compare their relative importance in the modeling process.

3.3 Model Building

DTC is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "DTC is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the DTC takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

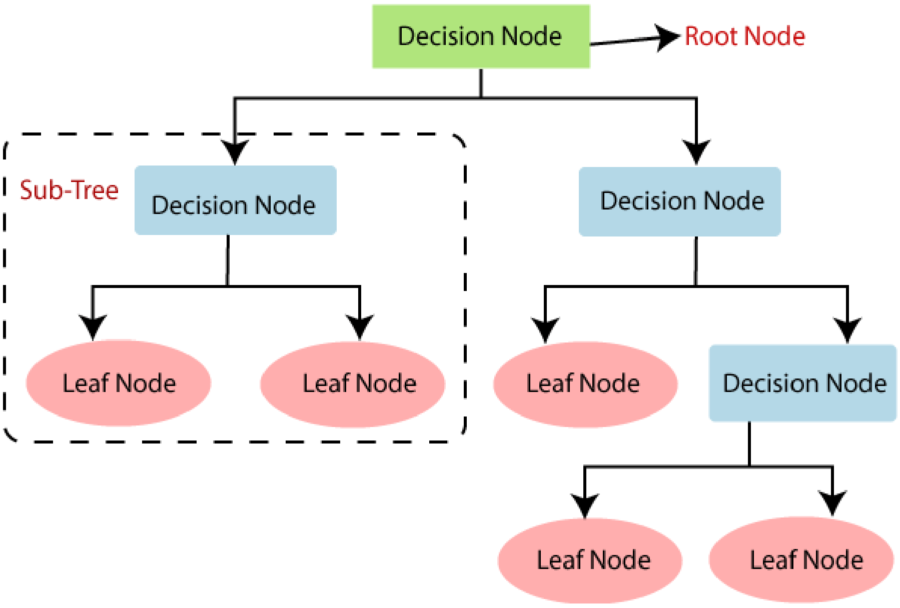


Fig. 2: DTC algorithm.

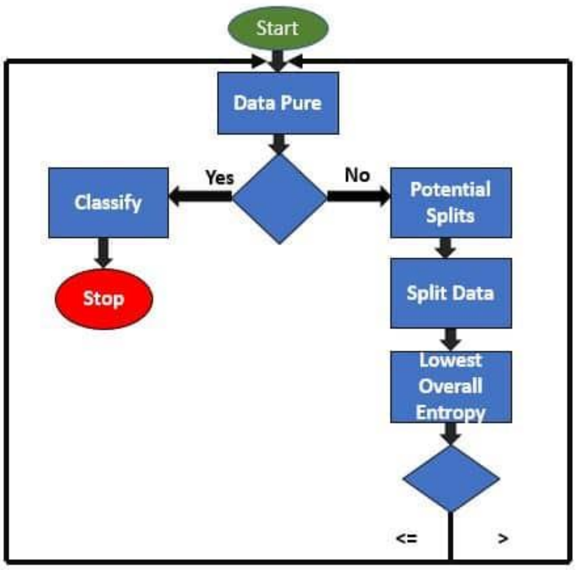


Fig. 3: Internal Operation of DTC.

One of the key strengths of the Decision Tree Classifier (DTC) lies in its diversity. Not all attributes or features are considered while constructing an individual tree, making each tree different and contributing to the overall robustness of the model. Additionally, DTC is relatively immune to the curse of dimensionality, as each tree only considers a subset of features, thereby reducing the feature space and improving efficiency. Another advantage is its support for parallelization. Since each tree is built independently using different subsets of data and features, the model can fully utilize CPU capabilities for faster training.

Step 1: In DTC n number of random records are taken from the data set having k number of records.

Step 2: Individual decision trees are constructed for each sample.

Step 3: Each decision tree will generate an output.

Step 4: Final output is considered based on Majority Voting or Averaging for Classification and regression respectively.

Unlike some other models, DTC inherently handles the train-test split, as approximately 30% of the data typically remains unseen by individual trees, providing a form of internal validation. Furthermore, DTC offers stability in its predictions due to majority voting or averaging across multiple decision trees, which reduces the variance and improves generalization. There are also specific assumptions that help enhance the performance of a DTC. Since the model aggregates the predictions of multiple trees, it's expected that while some individual trees might misclassify, the ensemble as a whole should accurately predict the output. For optimal performance, it is assumed that there should be meaningful values in the feature variables, enabling the classifier to make informed decisions rather than random guesses. Moreover, the predictions from each tree should be weakly correlated to ensure that the ensemble benefits from diverse perspectives. The DTC algorithm is preferred for various reasons. It requires relatively less training time compared to other complex models, delivers high accuracy, scales efficiently even for large datasets, and maintains accuracy even when a significant portion of data is missing.

4. RESULTS AND DISCUSSION

4.1 Dataset description

The dataset is used to explore trends in sleep patterns, assess the impact of occupation on health, and investigate associations between variables like BMI, blood pressure, and sleep disorders. It includes several key columns. The Gender column records the gender of each individual, typically categorized as "Male" or "Female." The Age column represents the age of each individual in years, providing information about the age distribution of the dataset. The Occupation column records the occupation of each individual, which varies widely and offers insights into the diversity of professions within the dataset. Common occupations include "Software Engineer," "Doctor," and "Nurse". The Sleep Duration column indicates how many hours of sleep each individual gets, offering insights into the average sleep duration across the dataset. Quality of Sleep is a subjective measure of how well an individual perceives their sleep, often rated on a scale, with higher values indicating better sleep quality. The Physical Activity Level column records the level of physical activity for each individual, which can vary in unit and scale, and reflects how active people are in the dataset. Stress Level measures an individual’s perceived stress, usually on a scale from low to high, providing insights into the stress levels of the participants. The BMI Category column categorizes individuals into BMI groups such as "Overweight," "Normal," and "Obese" based on their BMI values. Blood Pressure records systolic and diastolic measurements, typically expressed as a fraction (e.g., 126/83). The Heart Rate column represents the resting heart rate of individuals in beats per minute (BPM). Daily Steps records the number of steps individuals take daily, helping assess their physical activity levels. Finally, the Sleep Disorder column indicates whether an individual has a sleep disorder, including categories like "Sleep Apnea" and "None," offering insights into the prevalence of sleep disorders in the dataset.

4.2 Results analysis

The figure 5 visualizes the target column of the data frame after SMOTE. It represents the distribution of classes or categories in the target variable by balancing the count of each class related to sleep disorders. The figure 6 presents the confusion matrix for the Random Forest classifier. It provides a detailed breakdown of the model's predictions, including true positives, true negatives, false positives, and false negatives.

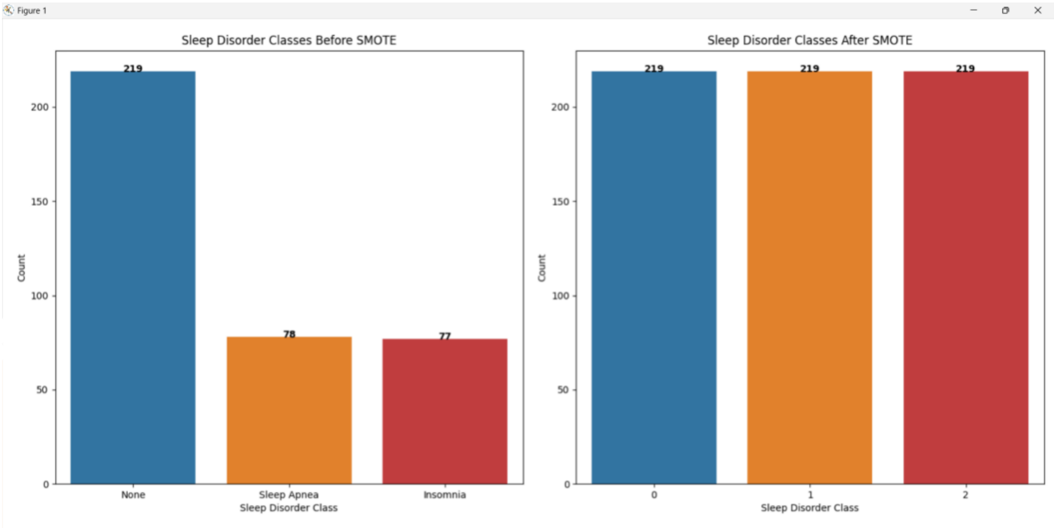


Fig. 5: Count plot of the Target column of data frame before and after SMOTE.

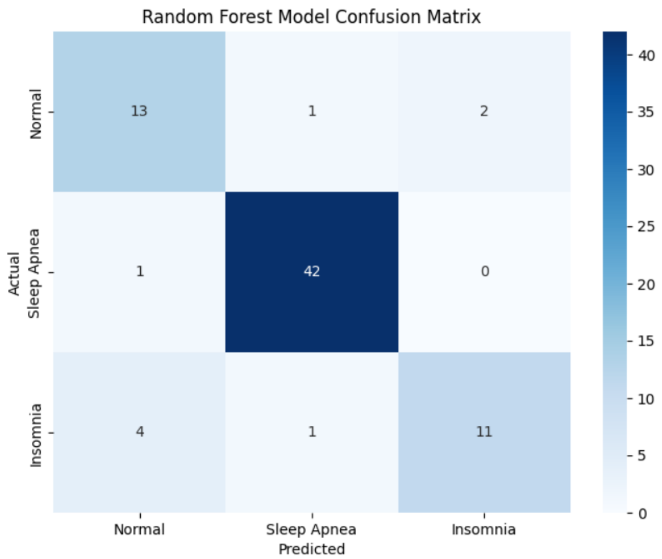


Fig. 6: Confusion matrix of Random Forest classifier.

The figure 7 displays the confusion matrix, but for the Decision Tree classifier. It provides insights into the model's performance in terms of classifying instances related to sleep disorders.

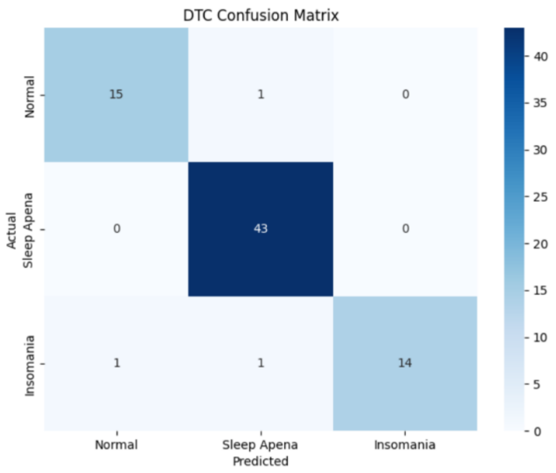
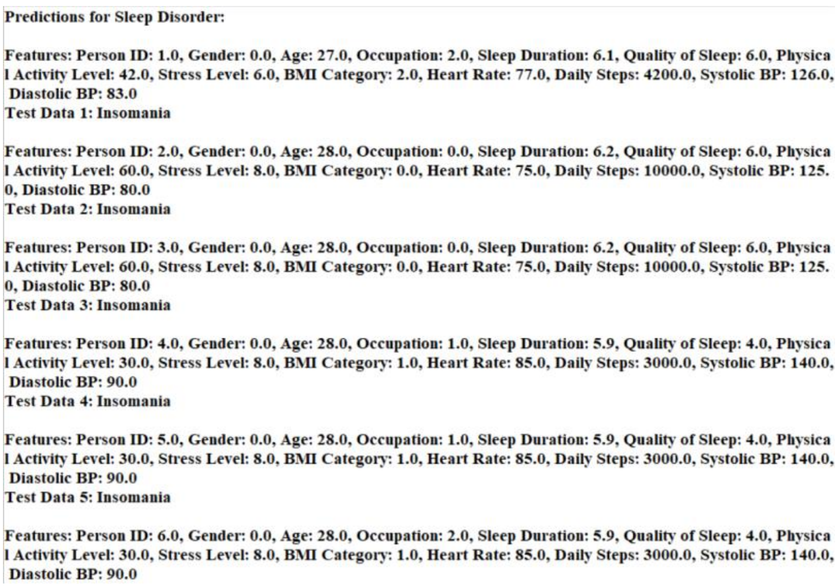


Fig. 7: Confusion matrix for Decision tree classifier.

The table compares the performance metrics (Accuracy, Precision, Recall, F1 Score) between two classification models Decision Tree classifier and Random Forest Classifier. It provides a side-by-side analysis of their effectiveness in handling sleep disorder-related predictions.

Table 1: Performance comparison of quality metrics obtained using Decision tree classifier & Random forest Classifier

Model	Accuracy	Precision	Recall	F1score
Decision tree classifier	96	96	96	96
Random forest Classifier	88	88	88	88



Predictions for Sleep Disorder:

Features: Person ID: 1.0, Gender: 0.0, Age: 27.0, Occupation: 2.0, Sleep Duration: 6.1, Quality of Sleep: 6.0, Physical Activity Level: 42.0, Stress Level: 6.0, BMI Category: 2.0, Heart Rate: 77.0, Daily Steps: 4200.0, Systolic BP: 126.0, Diastolic BP: 83.0
Test Data 1: Insomania

Features: Person ID: 2.0, Gender: 0.0, Age: 28.0, Occupation: 0.0, Sleep Duration: 6.2, Quality of Sleep: 6.0, Physical Activity Level: 60.0, Stress Level: 8.0, BMI Category: 0.0, Heart Rate: 75.0, Daily Steps: 10000.0, Systolic BP: 125.0, Diastolic BP: 80.0
Test Data 2: Insomania

Features: Person ID: 3.0, Gender: 0.0, Age: 28.0, Occupation: 0.0, Sleep Duration: 6.2, Quality of Sleep: 6.0, Physical Activity Level: 60.0, Stress Level: 8.0, BMI Category: 0.0, Heart Rate: 75.0, Daily Steps: 10000.0, Systolic BP: 125.0, Diastolic BP: 80.0
Test Data 3: Insomania

Features: Person ID: 4.0, Gender: 0.0, Age: 28.0, Occupation: 1.0, Sleep Duration: 5.9, Quality of Sleep: 4.0, Physical Activity Level: 30.0, Stress Level: 8.0, BMI Category: 1.0, Heart Rate: 85.0, Daily Steps: 3000.0, Systolic BP: 140.0, Diastolic BP: 90.0
Test Data 4: Insomania

Features: Person ID: 5.0, Gender: 0.0, Age: 28.0, Occupation: 1.0, Sleep Duration: 5.9, Quality of Sleep: 4.0, Physical Activity Level: 30.0, Stress Level: 8.0, BMI Category: 1.0, Heart Rate: 85.0, Daily Steps: 3000.0, Systolic BP: 140.0, Diastolic BP: 90.0
Test Data 5: Insomania

Features: Person ID: 6.0, Gender: 0.0, Age: 28.0, Occupation: 2.0, Sleep Duration: 5.9, Quality of Sleep: 4.0, Physical Activity Level: 30.0, Stress Level: 8.0, BMI Category: 1.0, Heart Rate: 85.0, Daily Steps: 3000.0, Systolic BP: 140.0, Diastolic BP: 90.0

Test Case 1: RFC Model Prediction on Test Case 1.

```
Predictions for Sleep Disorder:

Features: Person ID: 1.0, Gender: 0.0, Age: 27.0, Occupation: 2.0, Sleep Duration: 6.1, Quality of Sleep: 6.0, Physica
l Activity Level: 42.0, Stress Level: 6.0, BMI Category: 2.0, Heart Rate: 77.0, Daily Steps: 4200.0, Systolic BP: 126.0
 Diastolic BP: 83.0
Test Data 1: Sleep Apena

Features: Person ID: 2.0, Gender: 0.0, Age: 28.0, Occupation: 0.0, Sleep Duration: 6.2, Quality of Sleep: 6.0, Physica
l Activity Level: 60.0, Stress Level: 8.0, BMI Category: 0.0, Heart Rate: 75.0, Daily Steps: 10000.0, Systolic BP: 125.
0, Diastolic BP: 80.0
Test Data 2: Sleep Apena

Features: Person ID: 3.0, Gender: 0.0, Age: 28.0, Occupation: 0.0, Sleep Duration: 6.2, Quality of Sleep: 6.0, Physica
l Activity Level: 60.0, Stress Level: 8.0, BMI Category: 0.0, Heart Rate: 75.0, Daily Steps: 10000.0, Systolic BP: 125.
0, Diastolic BP: 80.0
Test Data 3: Sleep Apena

Features: Person ID: 4.0, Gender: 0.0, Age: 28.0, Occupation: 1.0, Sleep Duration: 5.9, Quality of Sleep: 4.0, Physica
l Activity Level: 30.0, Stress Level: 8.0, BMI Category: 1.0, Heart Rate: 85.0, Daily Steps: 3000.0, Systolic BP: 140.0
 Diastolic BP: 90.0
Test Data 4: Normal

Features: Person ID: 5.0, Gender: 0.0, Age: 28.0, Occupation: 1.0, Sleep Duration: 5.9, Quality of Sleep: 4.0, Physica
l Activity Level: 30.0, Stress Level: 8.0, BMI Category: 1.0, Heart Rate: 85.0, Daily Steps: 3000.0, Systolic BP: 140.0
 Diastolic BP: 90.0
Test Data 5: Normal

Features: Person ID: 6.0, Gender: 0.0, Age: 28.0, Occupation: 2.0, Sleep Duration: 5.9, Quality of Sleep: 4.0, Physica
l Activity Level: 30.0, Stress Level: 8.0, BMI Category: 1.0, Heart Rate: 85.0, Daily Steps: 3000.0, Systolic BP: 140.0
 Diastolic BP: 90.0
```

Test Case 2: DTC Model Prediction on Sample Test Case 2.

The Test Cases 1 and 2 showcases the results of predictions made by the Decision Tree classifier. It has a visual representation of predicted values compared to actual values in the context of sleep disorders.

5. CONCLUSION

The research has undertaken a systematic approach to tackle the complex issue of predicting sleep disorders using sleep health and lifestyle properties. Beginning with data preprocessing to ensure data quality and relevance, the project explored two distinct machine learning models: the existing Random Forest Classifier (RFC) and the proposed Decision Tree Classifier (DTC). Both models were trained on the pre-processed dataset, learning to make predictions about sleep disorders based on a range of sleep health and lifestyle properties. Model performance was meticulously evaluated using key metrics, including accuracy, precision, recall, and F1-score. Through this comparative analysis, valuable insights were gained into the effectiveness of each model in predicting sleep disorders. The project represents a significant step toward improving sleep disorder diagnosis and management by leveraging predictive modeling, with potential benefits for public health and healthcare decision-making.

REFERENCES

[1] Hoyniak, Caroline P., et al. "The physical home environment and sleep: What matters most for sleep in early childhood." Journal of Family Psychology 36.5 (2022): 757.

[2] Sutherland, Kate, et al. "Comparative associations of oximetry patterns in Obstructive Sleep Apnea with incident cardiovascular disease." Sleep 45.12 (2022): zsac179.

[3] Yang, Gee Su, et al. "Longitudinal analysis of sleep disturbance in breast cancer survivors." Nursing Research 71.3 (2022): 177-188.

[4] You, Yanwei, et al. "Muscle quality index is associated with trouble sleeping: a cross-sectional population based study." BMC Public Health 23.1 (2023): 489.

[5] Yıldırım, Murat, Akif Öztürk, and Fatma Solmaz. "Fear of COVID-19 and sleep problems in Turkish young adults: Mediating roles of happiness and problematic social networking sites use." Psihologija 00 (2023): 27-27.

[6] Guo, Ningyuan, et al. "Between-and within-person associations of mobile gaming time and total screen time with sleep problems in young adults: Daily assessment study." Addictive Behaviors 134 (2022): 107408.

[7] Sun, Jie, et al. "Artificial intelligence in psychiatry research, diagnosis, and therapy." Asian Journal of Psychiatry (2023): 103705.

[8] Tang, Nicole KY, Ptolemy DW Banks, and Adam N. Sanborn. "Judgement of sleep quality of the previous night changes as the day unfolds: A prospective experience sampling study." Journal of Sleep Research 32.3 (2023): e13764.
[9] Salfi, Federico, et al. "The fall of vulnerability to sleep disturbances in evening chronotypes when working from home and its implications for depression." Scientific Reports 12.1 (2022): 12249.
[10] Bitkina, Olga Vl, Jaehyun Park, and Jungyoon Kim. "Modeling sleep quality depending on objective actigraphic indicators based on machine learning methods." International Journal of Environmental Research and Public Health 19.16 (2022): 9890.
[11] Lee, Chiyoung, and Heewon Kim. "Machine learning-based predictive modeling of depression in hypertensive populations." PLoS One 17.7 (2022): e0272330.
[12] Byeon, Haewon. "Developing a predictive model for depressive disorders using stacking ensemble and naive Bayesian nomogram: using samples representing South Korea." Frontiers in Psychiatry 12 (2022): 773290.
[13] Thong, Benjamin Ka Seng, et al. "Telehealth Technology Application in Enhancing Continuous Positive Airway Pressure Adherence in Obstructive Sleep Apnea Patients: A Review of Current Evidence." Frontiers in Medicine 9 (2022): 877765.
[14] Alramadeen, Wesam, et al. "A novel sparse linear mixed model for multi-source mixed-frequency data fusion in telemedicine." IISE Transactions on Healthcare Systems Engineering (2023): 1-11.
[15] Kim, Junho, et al. "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 (2022): 664.
[16] Tsai, Cheng-Yu, et al. "Machine learning approaches for screening the risk of obstructive sleep apnea in the Taiwan population based on body profile." Informatics for Health and Social Care 47.4 (2022): 373-388.
[17] Arafa, Ahmed, et al. "Sleep duration and atrial fibrillation risk in the context of predictive, preventive, and personalized medicine: the Suita Study and meta-analysis of prospective cohort studies." EPMA Journal 13.1 (2022): 77-86.
[18] Gomes, Stephania Ruth Basilio Silva, Malcolm von Schantz, and Mario Leocadio-Miguel. "Predicting depressive symptoms in middle-aged and elderly adults using sleep data and clinical health markers: A machine learning approach." Sleep Medicine 102 (2023): 123-131.
[19] Mehra, Reena, et al. "Sleep-disordered breathing and cardiac arrhythmias in adults: mechanistic insights and clinical implications: a scientific statement from the American Heart Association." Circulation 146.9 (2022): e119-e136.
[20] Tsai, Cheng-Yu, et al. "Screening for obstructive sleep apnea risk by using machine learning approaches and anthropometric features." Sensors 22.22 (2022): 8630.
[21] Cai, Xintian, et al. "Associations of the cardiometabolic index with the risk of cardiovascular disease in patients with hypertension and obstructive sleep apnea: Results of a longitudinal cohort study." Oxidative Medicine and Cellular Longevity 2022 (2022).
[22] Hussain, Iqram, et al. "Quantitative evaluation of EEG-biomarkers for prediction of sleep stages." Sensors 22.8 (2022): 3079.
[23] Streckenbach, Bettina, et al. "Validating Discriminative Signatures for Obstructive Sleep Apnea in Exhaled Breath." Cells 11.19 (2022): 2982.
[24] Peltz, Jack S., et al. "Psychological processes linking problematic smartphone use to sleep disturbance in young adults." Sleep Health 9.4 (2023): 524-531.
[25] Wang, Shanze, et al. "The obesity-related dietary pattern is associated with higher risk of sleep disorders: a cross-sectional study from NHANES." Nutrients 14.19 (2022): 3987.