

Captivating Insights Into Sleep Health: Orchestrating Lifestyle Features and Predictive Models Through Computational Intelligence

Ruba Abu Khurma^{1*} , Yaning Xiao² , Bilal Al-Ahmad^{3,4} , Bilal Abu-Salih⁵ , Krishna Gopal Dhal⁶ , David Camacho⁷ 

¹ Department of Computer Sciences, Faculty of Information Technology and Computer Sciences, Yarmouk University, Irbid, 21163 (Jordan)

² Center for Control Science and Technology, Southern University of Science and Technology, Shenzhen, 518055 (China)

³ Computer Information Systems, The University of Jordan (Jordan)

⁴ Department of Computer Science and Engineering, College of Engineering, American University of Sharjah (UAE)

⁵ King Abdullah II School of Information Technology, The University of Jordan, Amman (Jordan)

⁶ Department of Computer Science and Application, Midnapore College (Autonomous), Paschim Medinipur, Midnapore, West Bengal (India)

⁷ Computer Systems Engineering Department, Universidad Politécnica de Madrid (Spain)

* Corresponding author: Ruba.abukhurma@yu.edu.jo

Received 14 December 2024 | Accepted 24 December 2025 | Published 13 July 2026

unir
LA UNIVERSIDAD
EN INTERNET

ABSTRACT

This study develops an interpretable Machine Learning (ML) framework with enhanced feature engineering to predict sleep disorders, achieving 93% accuracy and improving over prior models. The main objective of this work is to enhance the scientific rigor of sleep research and computational intelligence applications by providing insights into evidence-based interventions and optimizing sleep quality through customized lifestyle modifications. The methodology includes careful dataset selection, preprocessing to manage missing data and ensure integrity, and comprehensive feature engineering to accounting for the temporal aspects of sleep patterns. Multiple machine learning models are developed, including supervised and unsupervised approaches such as clustering algorithms, SVM, Random Forests, and Neural Networks. Evaluation metrics—accuracy, precision, recall, and F1-score—are cross-validated to ensure model robustness and prevent overfitting. Visualization tools, including heatmaps, scatter plots, and dimensionality reduction techniques, are employed to interpret the relationships between lifestyle factors and sleep health outcomes. Ethical considerations—such as data privacy, confidentiality, and the responsible interpretation of results—are upheld throughout the investigation.

KEYWORDS

Computational Intelligence, Feature Engineering, Machine Learning, Modeling, Sleep Health, Visualization.

DOI: 10.9781/ijimai.2026.2225
eLocator: e2225

I. INTRODUCTION

SLEEP health is a multifaceted concept vital to human well-being, akin to water and food, crucial for maintaining physical and psychological health [1]. However, modern lifestyles often prioritize productivity over adequate sleep, leading to detrimental effects on health and well-being. Recent studies highlight the crucial role of sleep in bodily and mental rejuvenation, underscoring its impact on safety, cognitive function, and physical health [2]. Insufficient sleep has been linked to increased risks of chronic diseases such as Type 2 diabetes, obesity, and cardiovascular diseases across all age groups [3].

Advanced computational intelligence refers to a variety of advanced artificial intelligence methods that are intended to tackle difficult problems on their own. Systems can learn from data and generate predictions using machine learning (ML), which includes supervised, unsupervised, and reinforcement learning. This allows systems to adapt dynamically to new information [4]. A branch of machine learning called deep learning leverages deep neural networks to revolutionize domains such as natural language processing (NLP) and picture identification. Tasks like sentiment analysis, language translation, and image categorization are made possible by these technologies. In order to understand visual data for uses like object detection and medical

Please cite this article as: R. A. Khurma, Y. Xiao, B. Al-Ahmad, B. Abu-Salih, K. Gopal Dhal, D. Camacho. Captivating Insights Into Sleep Health: Orchestrating Lifestyle Features and Predictive Models through Computational Intelligence, *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 10, no. 1, e2225, 2026, <http://doi.org/10.9781/ijimai.2026.2225>

picture analysis, computer vision is essential. Swarm intelligence and evolutionary algorithms replicate biological processes to repeatedly improve solutions, while reinforcement learning teaches agents optimal strategies through interaction with their surroundings. Applications for probabilistic graphical models can be found in recommendation systems, economics, and healthcare. These models handle complex interactions through representation. When combined, these methods promote automation, enhance judgment, and extract knowledge from huge datasets in a variety of industries.

This study uses cutting-edge ML techniques in conjunction with computational intelligence to conduct a thorough analysis of sleep health. This work introduces a streamlined, interpretable decision tree model with tailored preprocessing strategies, offering clearer insights into sleep-health determinants. Our goal is to forecast sleep accurately, disturbances and comprehend the fundamental mechanisms controlling sleep length and quality by incorporating a variety of lifestyle factors into our analysis. Thorough dataset selection, preprocessing, feature engineering, machine learning model building, assessment, visualization, and ethical concerns are all part of our methodology.

A. Context and Scope of the Study

The fast-paced lifestyle of today has devalued sleep, with serious health and well-being implications. By incorporating cutting-edge machine learning techniques, this study fills in knowledge gaps in the field by offering a more comprehensive understanding of sleep health. We investigate the effects of environmental factors, stress levels, and physical activity on sleep patterns and quality.

B. Related Works and Gaps

Studies have demonstrated how conscientiousness and stress at work affect the quality of sleep [5], [6]. Nevertheless, extant research frequently depends on self-reported data and has difficulties when implementing intricate machine learning models in situations with limited resources. By creating a strong machine learning model and using extensive feature engineering approaches to improve prediction accuracy, this study seeks to overcome these constraints.

C. Motivation

This work attempts to close current gaps in sleep research by utilizing cutting-edge machine learning techniques, given the essential role that sleep plays in both productivity and health. Through data-driven analysis, we aim to uncover actionable insights that might guide individualized interventions aimed at improving sleep health.

D. Methodology

The methodology of this study comprises several key stages:

1. *Dataset Selection and Preprocessing*: The study started with a meticulous selection of a dataset full of information on sleep patterns and lifestyle factors. The dataset is subjected to extensive preparation in order to address missing data, standardize numerical attributes, encode categorical variables, and guarantee data consistency.
2. *Feature Engineering*: A thorough feature engineering methodology is used, comprising the generation of derived features, the selection of pertinent features based on feature importance and correlation analysis, and the integration of temporal characteristics of sleep patterns.
3. *Machine Learning Model Development*: Numerous ML models are created, encompassing both unsupervised methods like clustering algorithms and supervised models like SVM, Random Forests, and Neural Networks. When appropriate, deep learning models such as CNNs or RNNs are also investigated.
4. *Model Training and Evaluation*: A 10-fold stratified cross-validation strategy was employed. The decision tree model was selected for its interpretability and strong performance after tuning, achieving higher accuracy than RF in prior studies. Metrics like accuracy, precision, recall, and F1-score are used to assess models with a separate subset of the dataset after they have been trained on it. Overfitting is reduced and robustness is ensured via cross-validation methods.
5. *Visualization and Interpretation*: Heatmaps, scatter plots, and dimensionality reduction techniques are used to depict the relationships between lifestyle factors and sleep health outcomes. This offers a natural understanding of the variables impacting sleep disruptions.
6. *Ethical Considerations*: Ethical principles, including data protection, confidentiality, and responsible interpretation of results, are respected throughout the whole study. The dataset is publicly available under a CC0 license from Kaggle. All entries are anonymized, and no identifiable personal data is included.

E. Contributions

This study contributes to sleep health research by:

- Developing a novel ML model for predicting sleep disturbances using comprehensive feature engineering.
- Providing actionable insights into the relationships between lifestyle factors and sleep health outcomes.
- Enhancing the methodological rigor of sleep research through advanced computational intelligence techniques.
- Upholding ethical standards in data handling and research practices.

II. RELATED WORK

Numerous prior research endeavors have comprehensively examined the correlation between lifestyle elements and sleep patterns, furnishing invaluable insights into the components that influence the quality and duration of sleep. Research has repeatedly demonstrated a relationship between gender and sleep [7]. ML classifiers have been employed to predict sleep disorders [8], utilizing methodologies that vary in complexity and effectiveness. Furthermore, publicly available datasets such as the Sleep Health and Lifestyle dataset have facilitated research in this domain.

Research on Lifestyle Choices and Sleep Quality Ma et al. [9] investigated the impact of lifestyle choices on sleep quality using linear regression and bootstrap techniques. Their study highlighted significant variables such as age, physical activity levels, and daily step count, correlating positively with improved sleep quality. However, the study's reliance on self-reported data and the potential for bias in lifestyle reporting are notable limitations.

ADAST: Attentive Cross-Domain EEG-Based Sleep Staging Framework Eldele et al. [1] proposed ADAST, a framework for cross-domain EEG-based sleep staging using iterative selftraining. ADAST improves accuracy across domains without extensive labeled data, leveraging iterative refinement. Advantages include robustness in domain adaptation; however, initial pseudo-label quality impacts performance.

Deep Transfer Learning for Cross-Subject Sleep Stage Classification Phan et al. [2] utilized deep transfer learning for cross-subject sleep stage classification, enhancing generalization to new subjects. The method's strength lies in its ability to improve classification accuracy by transferring knowledge. Yet, computational demands and model complexity remain challenges.

TABLE I. SUMMARY OF STUDIES ON AI/ML APPLICATIONS IN SLEEP RESEARCH

Reference	Focus and Objective	Methods	Dataset Used	Evaluation Measures	Advantages	Limitations	Key Findings
[9]	Impact of lifestyle on sleep quality	Linear regression, bootstrap	Not specified	Correlation coefficients	Correlation analysis, robust statistics	Self-reported data, potential bias	Age, physical activity, daily step count correlate positively with sleep quality.
[1]	Cross-domain EEG-based sleep staging	Iterative self-training	EEG-based datasets	Accuracy, F1 score	Domain adaptation, improved accuracy	Initial pseudo-label quality affects performance	ADAST improves accuracy in cross-domain EEG sleep staging without extensive labeled data.
[2]	Cross-subject sleep stage classification	Deep transfer learning	Sleep staging datasets	Accuracy, Precision, Recall	Knowledge transfer, enhanced generalization	Computational complexity, model deployment	Improved classification accuracy across diverse subjects using transfer learning techniques.
[10]	Unsupervised domain adaptation for sleep staging	Adversarial training	EEG and sleep staging datasets	Accuracy, Domain Adaptation	Domain adaptation, data efficiency	Adversarial training stability, hyperparameter tuning	Effective domain adaptation in sleep staging without labeled target domain data.
[3]	AI and ML for improving sleep quality	Wearable devices, AI integration	Sleep Health and Lifestyle dataset	Real-time monitoring	Personalized interventions, sensor accuracy	Data reliability, sensor accuracy	Integration of AI with wearables improves real-time sleep quality monitoring and interventions.
[12]	Detection of obstructive sleep apnea	ECG signal analysis	ECG and sleep apnea datasets	Accuracy, Sensitivity, Specificity	Non-invasiveness, high accuracy	Signal noise, preprocessing complexities	ECG-based method accurately detects obstructive sleep apnea, promising for diagnostics.
[13]	Predictive intelligence for healthcare outcomes	Ensemble learning	Healthcare outcomes datasets	Accuracy, Precision, Recall	Data integration, predictive accuracy	Model complexity, inter-pretability challenges	Ensemble methods improve predictive accuracy in healthcare outcomes prediction.
[14]	Enhanced ensemble models for predictive modeling	Ensemble methods	Predictive modeling datasets	Accuracy, Bias-Variance trade-off	Reduced bias, improved accuracy	Computational demands, model interpretability	Enhanced ensemble models reduce bias and variance, improving predictive accuracy.
[15]	ML models for diagnostic precision in sleep disorders	Various ML models	Diagnostic datasets	Accuracy, Precision, Recall	High diagnostic precision, performance metrics	Data complexity, model deployment feasibility	Ensemble methods achieve high accuracy in diagnosing sleep disorders, up to 91.6% accuracy reported.
[11]	Impact of lifestyle and health factors on sleep quality	RF, SVM	Sleep health and lifestyle datasets	Accuracy, Feature Importance	Identification of critical factors	Dataset limitations, model complexity	Lifestyle and health factors significantly influence sleep quality, identified through RF and SVM analyses.
[16]	Classification of sleep disorders	RF	Sleep disorder classification datasets	Accuracy, F1 score	Mitigation of overfitting, data imbalance	Data quality, feature relevance	RF model effectively classifies sleep disorders, addressing challenges like data imbalance and overfitting.
[17]	RF model for predicting sleep difficulties	RF	Sleep difficulty prediction datasets	Accuracy, Sensitivity, Specificity	High detection accuracy, predictive model	Model robustness, feature importance interpretation	RF model achieves 88% accuracy in predicting sleep difficulties, demonstrating robust performance.

Katabi et al. [3] integrated AI and ML with wearable devices for real-time sleep quality monitoring. Their approach offers personalized interventions but relies heavily on sensor accuracy and data reliability, limiting generalizability. Nasiri et al. [10] employed adversarial training for unsupervised domain adaptation in sleep staging. This method reduces domain discrepancy without labeled target data, suitable for datascarse environments. Challenges include adversarial training, stability, and hyperparameter tuning. Alshumayri et al. [11] analyzed lifestyle and health impacts on sleep using RF and SVM models. Their study provides insights into critical factors influencing sleep patterns, but is constrained by dataset limitations and model complexity.

Yang et al. [12] developed an ECG-based method for obstructive sleep apnea detection, showcasing non-invasiveness and high accuracy. However, signal noise and preprocessing complexities affect performance in real-world applications. Doe et al. [13] explored predictive models in healthcare outcomes, emphasizing ensemble learning's efficacy in diverse data integration. The complexity of ensemble models poses implementation challenges despite their predictive accuracy.

Enhanced Ensemble Models for Predictive Modeling Smith et al. [14] proposed enhanced ensemble models, demonstrating robust predictive performance. Advantages include reduced bias and

variance; however, computational demands and model interpretability remain limitations. Airlangga et al. [15] compared ML models in diagnosing sleep disorders, highlighting ensemble methods' superior performance metrics (e.g., accuracy up to 91.6%). Yet, data complexity and model deployment feasibility pose challenges.

Hidayat et al. [16] applied RF models to classify sleep disorders, addressing overfitting and data imbalance. While effective, the model's performance varies with data quality and feature relevance. Tareq et al. [17] introduced an RF model for predicting sleep difficulties, achieving an 88% detection accuracy. However, model robustness and feature importance interpretation are ongoing challenges.

The reviewed studies highlight the diverse methodologies and advancements in AI and ML for sleep research, offering promising avenues for personalized healthcare solutions. While these approaches enhance diagnostic accuracy and intervention effectiveness, challenges such as data quality, model complexity, and interpretability require ongoing research attention. Future studies should focus on refining algorithms, improving data integration strategies, and validating models in clinical settings to optimize sleep health outcomes effectively. Table I summarizes studies on AI/ML applications in sleep research.

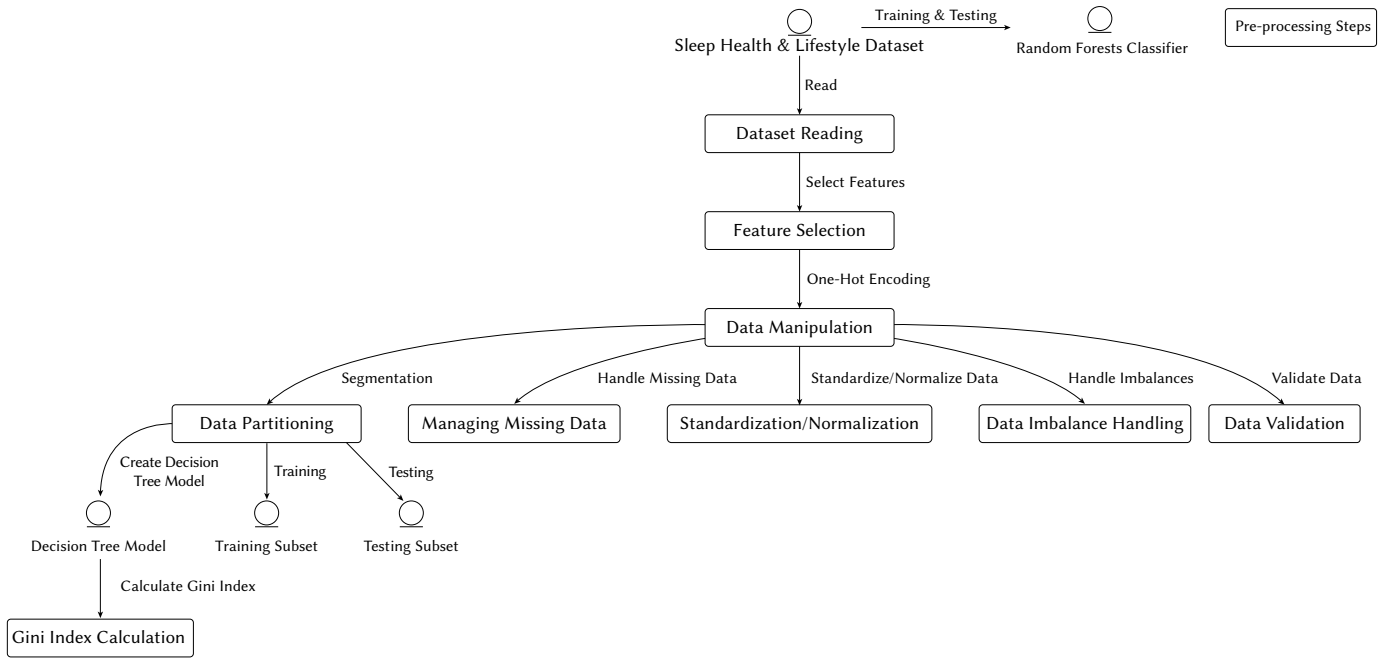


Fig. 1. Methodology.

III. PROPOSED METHODOLOGY

One of the recommended techniques for predicting sleep problems is the RF classifier. The model is trained and assessed using the sleep health and lifestyle dataset. RF combines many DT to improve accuracy. DT partitions data based on feature criteria, resulting in a structure that resembles a tree. Using its learning from the dataset, the algorithm forecasts the classes of sleep disorders.

The initial step involves a number of pre-processing steps before the DT model is built. Reading datasets, selecting features, modifying data (such as one-hot encoding), occupation, gender, and BMI category were encoded using one-hot encoding. Correlation analysis (threshold > 0.8) led to the removal of features like 'stress level' to reduce redundancy. Dividing datasets into training and testing subsets, addressing missing data, standardizing or normalizing data, correcting imbalances in the data, and validating the dataset are a few of the tasks that need to be completed. Furthermore, each DT node employs the Gini Index approach to ascertain the optimal split during the data partitioning process. This method aids in the selection of splits that enhance the target class's segregation, leading to the creation of an accurate and efficient DT. Fig. 1 provides a summary of these methods.

IV. DISCUSSION AND RESULTS

A. Explore Data Analysis (ADA) of Sleep Health Lifestyle Dataset

The Sleep Health and Lifestyle Dataset¹ is the dataset used in this research work. Numerous factors pertaining to sleep and everyday routines are present in this dataset, which is organized into 373 rows and 13 columns. It offers details on elements like:

1. Person ID
2. Gender
3. Age
4. Occupation
5. Sleep Duration: Average number of hours slept each day
6. Quality of Sleep: An arbitrary score between 1 and 10

7. Physical Activity Level: The average daily amount of time spent exercising by the person
8. Stress Level: A subjective rating on a 1-10 scale BMI Category
9. Blood Pressure: represented by the ratio of the systolic to the diastolic pressure
10. Heart Rate: In beats per minute
11. Sleep Disorder: One of None, Insomnia or Sleep Apnea

B. Exploring the Relationship Between Occupation and Sleep Patterns: An Analysis of Sleep Duration and Quality

For a basic understanding of the intricate connection between sleep health and occupation, refer to the following visualizations. They identify possible problem areas where occupational pressures may adversely affect sleep, as well as places where specific job types may be associated with improved sleep quality. To improve sleep health in particular professional groups, this analysis is essential for creating tailored therapies or suggestions.

1. *Sleep duration by occupation:* The relationship between occupation and sleep length is evident in Fig. 2. Analyzing Fig. 2 leads to the following crucial discovery.
 - Range and Median Different occupations show varied ranges in sleep duration. For instance, if certain occupations like 'Software Engineers' or 'Doctors' have a wider range, it might indicate a more inconsistent sleep pattern within these professions. The median line within each box indicates the typical sleep duration for each occupation, which can be compared across different jobs.
 - Outliers: If there are occupations with notable outliers (points that lie outside the typical range of the boxplot), this could suggest that individuals in these professions experience either significantly more or less sleep than their peers, possibly due to job-related stress or irregular work hours.
 - Tight vs. Wide Distributions: Some occupations might have tighter distributions (smaller boxes), suggesting more uniformity in sleep duration among individuals in these jobs. Others with wider distributions indicate more variability among workers in that field.

¹ <https://www.kaggle.com/datasets/henryshan/sleep-health-and-lifestyle>

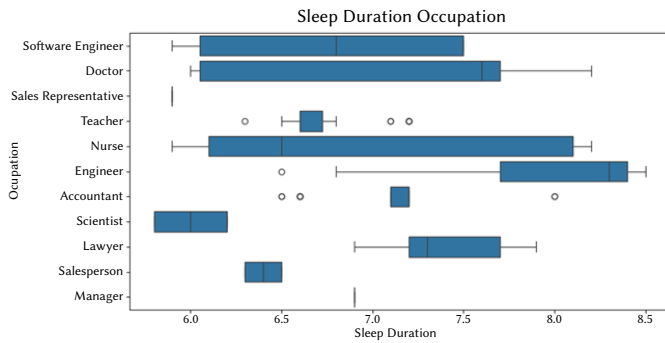


Fig. 2. Sleep duration by occupation.

2. *Quality of sleep by occupation:* Fig. 3 illustrates the correlation between occupation and sleep duration. The critical main findings that follow from the analysis of Fig. 3 are important:

- **Highs and lows:** pay attention to occupations that have higher median sleep quality scores versus those with lower scores. This can suggest which professions are associated with better-perceived sleep quality.
- **Consistency:** occupations with boxes that have less spread (narrower boxes) indicate a more consistent sleep quality experience among individuals in that profession, whereas wider boxes show greater variance.
- **Skewness:** the skewness in the distribution (whether the box and whiskers are symmetric or skewed to one side) can indicate a tendency towards higher or lower sleep quality ratings in certain occupations.

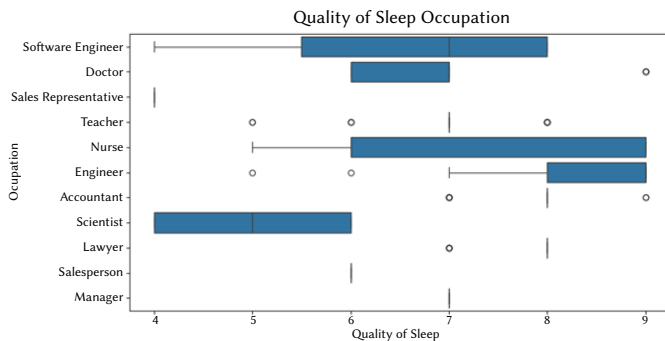


Fig. 3. Quality of sleep occupation.

C. Analyzing the Intersection of Age, Sleep Duration, and Sleep Disorders

The scatter plot in Fig. 4 visualizes the relationship between sleep duration, age, and the presence of sleep disorders, offering several valuable insights. A study of Fig. 4 yields the following conclusions:

- **Age and sleep duration correlation**

The plot provides a visual representation of how sleep duration varies with age. Look for any trends such as increasing or decreasing sleep duration, as age increases. If a trend is visible, it suggests a potential age-related change in sleep patterns.

- **Impact of sleep disorders**

The use of different colors to represent different sleep disorders (None, Insomnia, Sleep Apnea) allows us to observe how these conditions are distributed across different age groups and how they might affect sleep duration. Key insights could include:

- Whether certain sleep disorders are more prevalent in specific age groups.

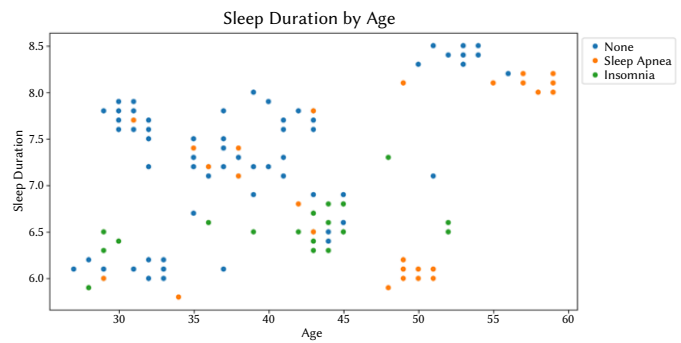


Fig. 4. Sleep duration by Age.

- How sleep duration varies among individuals with different sleep disorders. For instance, do those with insomnia tend to have shorter sleep duration?

- **Cluster and Spread**

The distribution and clustering of points can indicate common patterns. A dense clustering of points in a certain area suggests a common sleep duration for a specific age group, while a more spread-out distribution indicates variability.

- **Outliers and Anomalies**

Any points that lie far from the general cluster can indicate unusual cases, such as very young or old individuals with atypical sleep duration or specific sleep disorder patterns not common in their age group. This scatter plot effectively combines three critical dimensions of the dataset—age, sleep duration, and sleep disorders—to provide a comprehensive view of how these factors interrelate. Such insights are invaluable for understanding the dynamics of sleep health across different age groups and for identifying specific age groups that might be more susceptible to certain sleep disorders.

D. Evaluating the Relationship Between Sleep Duration and Quality in the Context of Sleep-Related Problems

Fig. 5 provides a valuable visualization of how sleep quality correlates with sleep duration, further categorized by the presence of sleep disorders (None, Insomnia, Sleep Apnea). The insights that can be drawn from this visualization include:

- **Correlation between Sleep Duration and Quality**

The plot shows the relationship between how long individuals sleep (sleep duration) and how well they rate their sleep (quality of sleep). A positive or negative trend in the line would indicate a direct or inverse relationship, respectively. For example, if the line rises with increased sleep duration, it suggests that longer sleep is associated with better-perceived sleep quality.

- **Impact of sleep disorders on sleep quality**

The different lines for each sleep disorder provide a comparative view of how sleep quality is affected in the presence of these conditions. Key observations could include whether individuals with certain sleep disorders (like Insomnia or Sleep Apnea) consistently rate their sleep quality lower than those without a sleep disorder, regardless of sleep duration. It's also insightful to see if the relationship between sleep duration and quality differs significantly among the different sleep disorder groups.

- **Variations and patterns**

Look for any notable peaks, troughs, or patterns in the lines. For instance, a peak might suggest an optimal sleep duration for quality sleep, while a trough could indicate a sleep duration

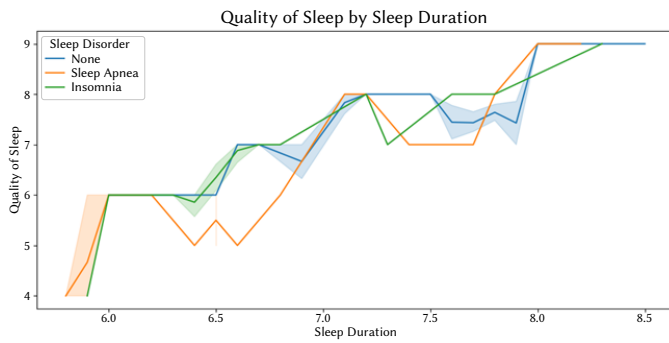


Fig. 5. Quality Sleep by sleep duration.

where quality diminishes. The spread and overlap of the lines can indicate how similar or different the sleep quality experiences are for individuals with different sleep disorders.

- Consistency across sleep disorders

If the lines for different sleep disorders follow a similar pattern, it suggests a consistent effect of sleep duration on quality across these conditions. Divergent patterns, on the other hand, would indicate that the type of sleep disorder significantly alters this relationship. This visualization is instrumental in understanding the complex dynamics between sleep duration, quality, and disorders, offering critical insights for sleep health research and potential interventions.

E. Comparative Analysis of Sleep Quality Across Different Sleep Disorders

The box plot visualization provides an insightful comparison of how sleep quality varies among individuals with different sleep disorders (None, Insomnia, Sleep Apnea). From this visualization, several key insights can be derived:

- Median and Spread of Sleep

Quality Scores The median (central line in each box) gives a quick indication of the typical sleep quality rating for each disorder category. The range (length of the box) shows the variability in sleep quality within each group. A larger range indicates greater variability.

- Comparison between groups

By comparing the median and range of each box, we can see how sleep quality differs between those with no sleep disorder, insomnia, and sleep apnea. For example, a lower median in the Insomnia group compared to the None group would suggest that insomnia is associated with poorer sleep quality.

- Outliers and extremes

Outliers (points outside the box) highlight individuals with sleep quality ratings that are notably different from the majority in their group. This can indicate exceptionally good or poor sleep quality within a disorder category. Whiskers (lines extending from the boxes) indicate the range of the majority of the data, excluding outliers. A long whisker might reveal a wider spread of experiences in sleep quality.

- Skewness in distribution

The symmetry or asymmetry of the boxes can indicate a skew in sleep quality ratings. For instance, a box skewed towards higher ratings in the 'None' group would suggest that most people without a sleep disorder generally rate their sleep quality positively. This box plot is essential for understanding the impact of different sleep disorders on perceived sleep quality. It allows for a direct

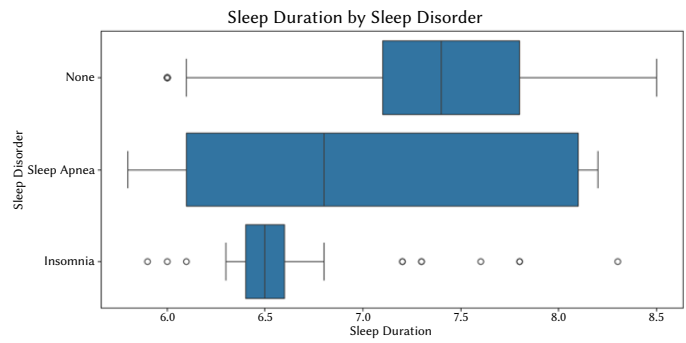


Fig. 6. Sleep duration by sleep disorder.

comparison between groups, providing insights into the severity of the impact of various sleep disorders on sleep quality. This information is crucial for healthcare professionals and researchers focusing on sleep health and the management of sleep disorders.

F. Gender Distribution Across Different Sleep Disorders

The pie charts in Fig. 7, Fig. 8, and Fig. 9 provide a clear visualization of the gender distribution within each category of sleep disorders (None, Insomnia, Sleep Apnea). From this visualization, several insights can be gleaned:

- Gender Proportions in Each Sleep Disorder

The charts show the percentage breakdown of males and females within each sleep disorder category. This helps to understand if there's a gender prevalence in any particular sleep disorder. For example, a significantly higher percentage of one gender in a disorder like Insomnia could suggest a gender predisposition.

- Comparative Analysis Across Disorders

By looking at the pie charts side by side, you can compare the gender distribution across different sleep disorders. Notable differences in gender ratios between disorders may indicate varying susceptibility or diagnosis rates among males and females.

- Overall Gender Balance

These charts also provide a visual representation of the overall gender balance within the study population, specifically regarding sleep disorders. A balanced distribution in the 'None' category would indicate a representative sample, while a skewed distribution in any disorder category might suggest an area for further investigation.

- Identifying Gaps or Biases

If any of the sleep disorders show a significant imbalance in gender distribution, it could point to potential gaps or biases in diagnosis, reporting, or the prevalence of the disorder itself. These visualizations are particularly useful for healthcare professionals and researchers to identify potential gender-related patterns or biases in the prevalence and diagnosis of sleep disorders. They also contribute to a broader understanding of how sleep disorders might differentially affect males and females.

G. Occupational Breakdown About Sleep Disorders: A Countplot Analysis

The counterplot in Fig. 10 provides a detailed view of how various sleep disorders are distributed across different occupations. Key insights that can be drawn from this visualization include:

- Prevalence of sleep disorders by occupation

The Fig. 10 shows the frequency of each sleep disorder (None, Insomnia, Sleep Apnea) within different occupational groups.

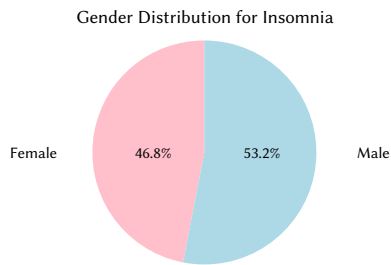


Fig. 7. Sleep duration by gender distribution.

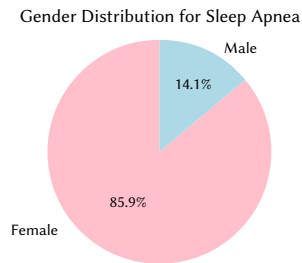


Fig. 8. Sleep duration by gender distribution.

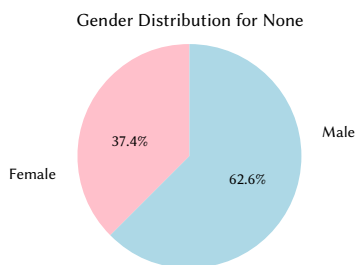


Fig. 9. Sleep duration by gender distribution.

This helps in identifying if certain occupations are more prone to specific sleep disorders. For instance, a high count of Insomnia in a particular occupation could indicate a stress-related or workenvironment factor influencing sleep health.

- Occupational stress and sleep health

Occupations with a higher frequency of sleep disorders may suggest higher levels of job-related stress or lifestyle factors that negatively impact sleep. Conversely, occupations with a predominantly 'None' category for sleep disorders could imply more conducive work environments for sleep health.

- Comparative analysis across occupations

By comparing the counts across different occupations, we can assess which professions are most affected by sleep disorders. This can guide further investigation into occupational health practices and the need for specific interventions or support systems in high-risk occupations.

- Distribution trends

The distribution of sleep disorders within each occupation, especially the relative proportions of Insomnia and Sleep Apnea, can reveal specific occupational health concerns. For example, a high incidence of Sleep Apnea in physically demanding jobs could point towards factors like obesity or physical health conditions common in those occupations. This counterplot is an effective tool for identifying potential occupational risk factors for sleep disorders and can be instrumental in guiding workplace health policies, awareness campaigns, and targeted interventions to improve sleep health in the workforce.

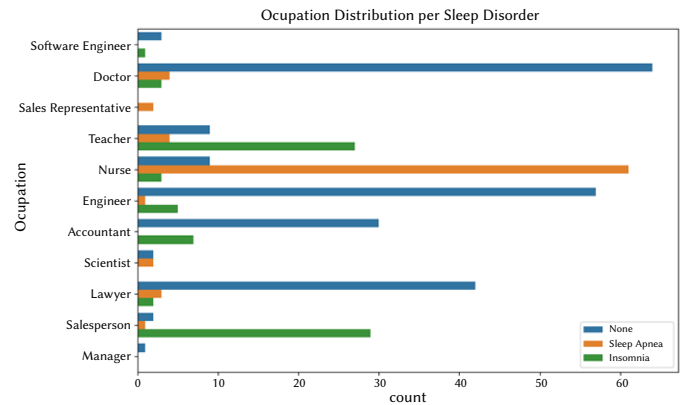


Fig. 10. Occupation Distribution per sleep disorder.

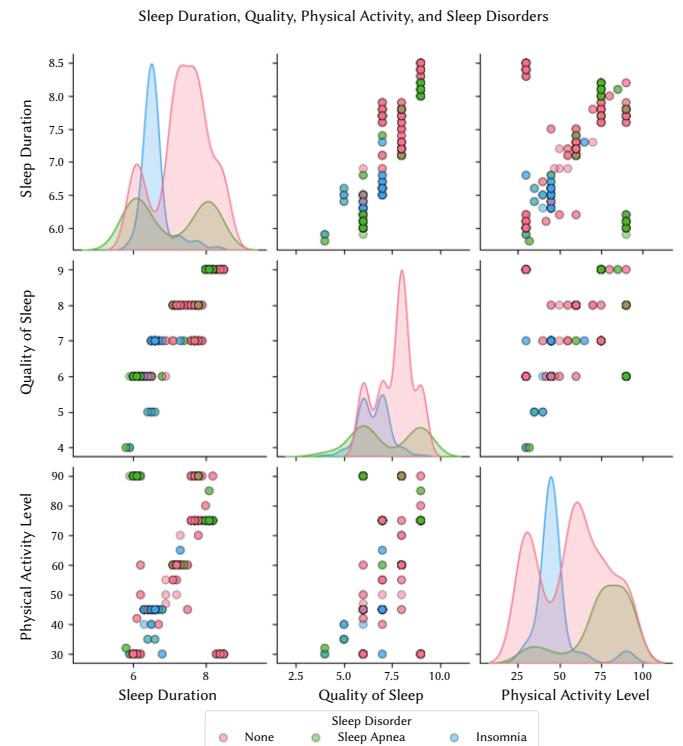


Fig. 11. Sleep duration, quality, physical activity, and sleep disorder.

H. Interplay Between Sleep Duration, Quality, Physical Activity, and Sleep Disorders.

The pairplot in Fig. 11 offers a comprehensive view of the relationships between sleep duration, quality of sleep, physical activity level, and sleep disorders. It provides a multidimensional perspective, revealing insights into how these variables interact with each other:

- Sleep duration vs. quality of sleep

This plot helps in understanding if there is a correlation between how long people sleep and how they rate the quality of their sleep. A positive correlation would suggest that longer sleep might be associated with better sleep quality.

- Sleep duration vs. physical activity level

This plot explores the relationship between physical activity and sleep duration. Insights here could include whether more physically active individuals tend to have longer or shorter sleep durations, suggesting a potential impact of physical activity on sleep.

- Quality of sleep vs. physical activity level

This plot assesses whether there's a relationship between physical activity levels and perceived quality of sleep. For instance, higher physical activity might be associated with better sleep quality, or there might be no clear pattern.

- Influence of sleep disorders

The hue differentiation based on sleep disorders allows for the comparison of these relationships in the context of different sleep disorders (None, Insomnia, Sleep Apnea). This aspect helps in determining whether the presence of a sleep disorder alters the typical patterns observed between sleep duration, quality, and physical activity.

- Distribution and overlap

The distribution of points in each plot, especially the degree of overlap between different sleep disorders, can provide insights into how distinct or similar these groups are in terms of their sleep and activity patterns.

- Patterns and outliers

Any notable patterns, clusters, or outliers in the plots can reveal unique insights, such as specific trends within a sleep disorder category or exceptional cases that deviate from general trends. This pairplot is a powerful tool for visualizing and understanding the complex relationships between key variables related to sleep health. It is especially useful for identifying patterns and relationships that might not be immediately obvious and can guide further statistical analysis or research.

I. Analyzing the Impact of Stress Levels on Mean Sleep Duration Across Sleep Disorders

The heatmap in Fig. 12 provides a visually compelling analysis of how average sleep duration varies across different sleep disorders and stress levels. It combines these two key aspects to offer several insights.

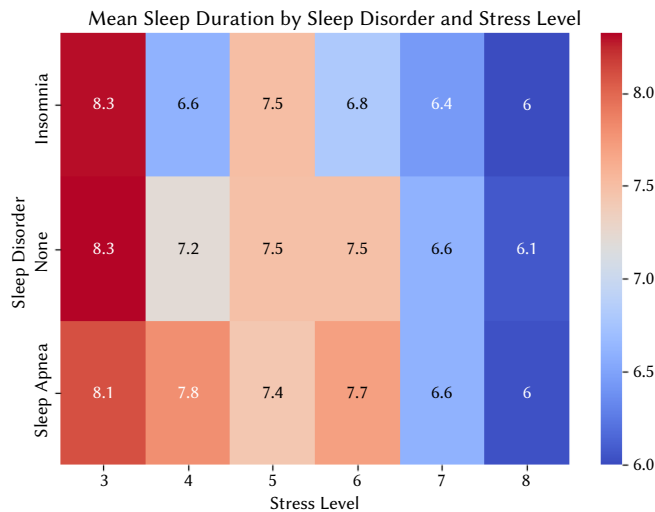


Fig. 12. Mean sleep duration by sleep disorder and stress level.

- Interaction between stress level and sleep disorders

Fig. 12 allows for an examination of how the mean sleep duration is influenced by both stress level and the type of sleep disorder. This can reveal if higher stress levels correlate with shorter sleep duration within specific sleep disorders.

- Comparative mean sleep duration

Each cell in the heatmap represents the average sleep duration for a specific combination of sleep disorder and stress level. This allows

for direct comparison across different groups. For example, cells with warmer colors (indicating shorter sleep duration) in high-stress levels can suggest a significant impact of stress on sleep duration, particularly in the context of certain sleep disorders.

- Patterns and trends

Any noticeable patterns, such as a consistent decrease in sleep duration with increasing stress levels across all sleep disorders, would be indicative of a strong link between stress and reduced sleep. Similarly, variations in the pattern for different sleep disorders can shed light on how these conditions interact with stress to affect sleep.

- Identifying critical combinations

The heatmap highlights critical combinations of stress levels and sleep disorders where sleep duration is significantly affected. This could be crucial for targeted interventions. For instance, if individuals with insomnia under high-stress levels show markedly reduced sleep duration, this group might require specific stress management interventions.

- Overall trends and anomalies

The general color trend across the heatmap can indicate overall tendencies, while any anomalies (such as an unexpectedly high or low average sleep duration in a particular cell) can point to areas needing further investigation. This heatmap is a valuable tool for healthcare professionals, researchers, and policymakers to understand the complex interplay between stress, sleep disorders, and sleep duration, guiding effective strategies for improving sleep health.

J. Heart Rate Variations About Different Sleep Disorders

The box plot in Fig. 13 provides a visual comparison of heart rate distributions across various sleep disorders (None, Insomnia, Sleep Apnea). This visualization offers several key insights.

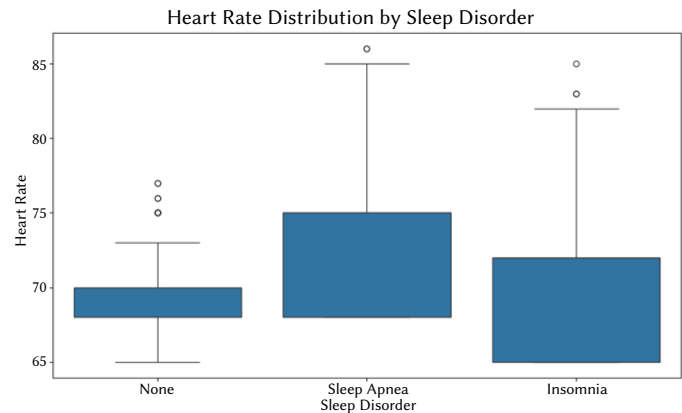


Fig. 13. Heart rate distribution by sleep disorder.

- Median heart rate per sleep disorder

The central line in each box represents the median heart rate for each sleep disorder category. Comparing these medians can reveal whether certain sleep disorders are associated with higher or lower median heart rates. For example, a higher median in the Sleep Apnea category might suggest a link between this disorder and elevated heart rates.

- Range and variability of heart rates

The length of each box (interquartile range) shows the spread of heart rates within each disorder category, indicating the variability in heart rate among individuals with the same sleep disorder.

Wider boxes suggest greater variability, while narrower boxes indicate more uniform heart rates within that group.

- Outliers and extremes

Outliers (points outside the box) highlight individuals with heart rates significantly different from the majority in their group. This can indicate exceptional cases that might warrant further investigation. The whiskers (lines extending from the boxes) show the range of typical heart rates, excluding outliers. Long whiskers indicate a wider range of heart rates.

- Comparative analysis across disorders

By examining the position and spread of boxes across different sleep disorders, it's possible to assess how these conditions might impact heart rate. For instance, if the box for Sleep Apnea is noticeably higher or has more outliers than other categories, it might suggest a distinct impact of this disorder on cardiovascular health.

- Symmetry and Skewness

The symmetry of the boxes can indicate whether the heart rate distribution is skewed higher or lower within each disorder category. This box plot is instrumental in understanding the potential cardiovascular implications of different sleep disorders. It can be particularly useful for healthcare professionals in identifying sleep disorders that may have a more pronounced effect on heart rate, guiding more targeted medical evaluations and interventions.

K. Exploring the Link Between Physical Activity Levels and Sleep Disorders: A Boxplot Perspective

The box plot visualization in Fig. 14 offers a clear comparison of physical activity levels across different sleep disorder categories (None, Insomnia, Sleep Apnea). This visualization provides several key insights.

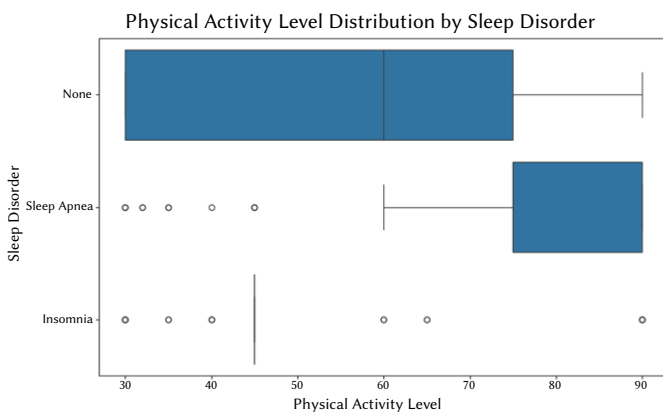


Fig. 14. Physical activity levels distribution by sleep disorder.

- Median physical activity level per sleep

Disorder The central line in each box indicates the median physical activity level for each sleep disorder category. This can reveal if certain sleep disorders are associated with higher or lower median activity levels. For example, a lower median in the Insomnia category might suggest reduced physical activity among individuals with this disorder.

- Range and variability in activity levels

The length of each box shows the spread of physical activity levels within each disorder group, indicating the variability among individuals. Wider boxes indicate a greater range of activity levels, suggesting diverse lifestyles within that sleep disorder category.

- Outliers and extremes

Outliers (points outside the box) highlight individuals with unusually high or low activity levels compared to the majority in their group. This can point to exceptional cases that might require additional scrutiny. The whiskers (lines extending from the boxes) show the typical range of activity levels, excluding outliers. Long whiskers indicate a broad range of physical activity levels.

- Comparative analysis across disorders

By comparing the position and spread of boxes across different sleep disorders, it's possible to assess how these conditions might correlate with physical activity levels. If a particular sleep disorder category consistently shows lower activity levels, it might suggest a link between that disorder and reduced physical activity.

- Symmetry and distribution

The symmetry of the boxes can provide insight into the distribution of activity levels. A skewed box might indicate a tendency towards either higher or lower activity levels within a sleep disorder category.

This boxplot is crucial for understanding the potential relationships between physical activity levels and different sleep disorders. It provides valuable insights for healthcare professionals and researchers, suggesting areas where lifestyle interventions might be beneficial, particularly for individuals with specific sleep disorders.

V. FEATURES ENGINEERING

The dataset has undergone comprehensive feature engineering, including standardization, encoding, and reduction of multicollinearity through correlation analysis. This process has refined the dataset into a format that is optimized for predictive modeling, ensuring that the features included are relevant and appropriately formatted for the analysis. The final feature set is expected to provide a strong foundation for developing a robust model to predict sleep disorders. The process included the following:

- Renaming columns

To ensure consistency and ease of data handling. All column names were converted to lowercase, and spaces were replaced with underscores.

- Encoding blood pressure

To create a more analytically useful representation of blood pressure. This is done by calculating the ratio of systolic to diastolic blood pressure and storing it in a new column 'bp-ratio'.

- Encoding the target variable

To transform the categorical target variable 'sleep disorder' into a numerical format. This is achieved by applying label encoding to the 'sleep disorder' column, converting categories to numerical labels. A DataFrame (label df) was created to map original and encoded labels.

- One Hot Encoding of categorical variables

To convert categorical variables into a format suitable for modeling. This is by applying One Hot Encoding to 'occupation', 'gender', and 'BMI category', with the first category dropped to avoid multicollinearity.

- Correlation analysis

To reduce collinearity and dimensionality in the dataset. This depends on calculating the correlation matrix for the encoded dataset, excluding 'person id', 'blood pressure', and 'sleep disorder'.

- Dropping columns based on correlation analysis

To remove features that are highly correlated or redundant. This is done by dropping several columns, including 'person id', 'blood pressure', 'sleep disorder', 'daily steps', certain occupation categories ('occupation Nurse'), 'stress level', and 'quality of sleep'.

- Final Feature Set

To define the final set of features for the predictive model. This is performed by creating a final feature set named features by dropping the identified columns.

A. Modeling

- Decision Tree Classifier

To establish a baseline model using a DT classifier. This step involves a Feature-Target Split by separating the dataset into features (X) and the target variable (y). Dividing the data into training and testing sets with an 80-20 split ensures class stratification. The Scaling uses standardized features using the Standard Scaler to ensure equal weighting. The Model Training is fitted with a DT on the training data. Model Evaluation: Evaluated the model on the test set and calculated accuracy. Also generated a classification report. Results. Achieved an accuracy of 89.33%. The classification report indicated varying precision, recall, and F1 scores across different classes. We employed SHAP (SHapley Additive exPlanations) to interpret model predictions. The most influential features included sleep duration, physical activity, and heart rate. Visualizations are provided to explain individual predictions. While the model performs well, it is intended as a decision-support tool to assist with identifying potential sleep disorders, not as a diagnostic system.

- Hyperparameter tuning using grid search

To adjust the hyperparameters of the DT model to optimize it. A grid of hyperparameters, comprising max depth, min samples split, min samples leaf, and criterion, must be defined in this phase. To identify the ideal parameter combination, use cross-validation in the conducted grid search. They assessed which criteria and results worked best. The findings indicate that entropy, max depth: None, min samples leaf: 1, and min samples split: 10 were determined to be the optimal settings. The grid search yielded an optimal score of about 88.3%.

- An optimized DT classifier's parameters

To fit the DT model using the optimal grid search settings. Starting a new DT classifier with optimized parameters is part of the process. Using the training set of data, they fitted the model and generated predictions for the test set. They produced a report on classification for the model that was optimized. The classification report performed better than the baseline model, as seen by enhanced precision, recall, and F1-scores, according to the results. By now, the accuracy was up to about 93%.

- Confusion matrix analysis

To interpret the model's performance visually. As part of the process, the confusion matrix for the optimized model is created and printed. To improve visibility, they plotted the confusion matrix as a heatmap. Fig 15 presents the results, where the confusion matrix offers a comprehensive analysis of the differences between the model's predicted and actual values. The heatmap visualization made it easier to see how well the model predicted each class.

- Comparing with related studies

This study intends to list the most popular algorithms for detecting sleep problems so that various ML models may be compared.

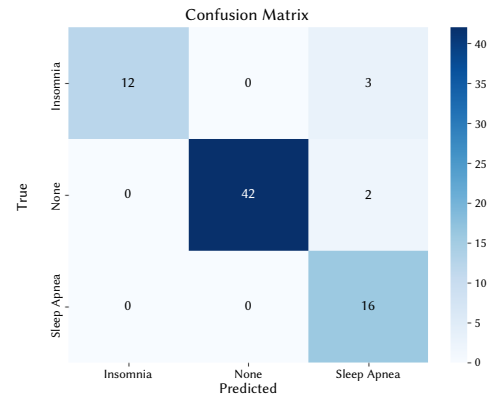


Fig. 15. Confusion matrix for our results.

We compared our results with relevant studies that used the same dataset in order to validate our suggested methodology. In comparison to three previous studies that employed the same dataset, Table II displays the accuracy result of the DT used in this investigation. DT was adopted by [15], whereas RF was adopted by [16] and [17].

TABLE II. COMPARED ACCURACY WITH RELATED STUDIES ON THE SAME INVESTIGATED DATASET

Study	Model	Accuracy
Airlangga [15]	DT	91.6%
Hidayat [16]	RF	88%
Tareq [17]	RF	88%
The proposed approach	DT	93%

The performance attained by the proposed model is superior to what other studies have reported. [15]–[17]. Our findings show significant findings by achieving 93% accuracy.

VI. CONCLUSION

The model's accuracy and recall rates showed significant improvements over the course of the four stages of the sleep health and lifestyle modeling process. The final model, which had undergone laborious hyperparameter tweaking and was quite proficient at diagnosing sleep issues, demonstrated an accuracy of nearly 93%. However, given the medical nature of the problem, it is critical to acknowledge the possibility of further enhancing the model's effectiveness. Look at different modeling approaches or data pre-processing techniques for possibly even better outcomes. However, in light of the critical importance of accuracy in medical diagnosis, a standard for model efficacy was set. This threshold aids in achieving two objectives, one of which is to retain a high degree of accuracy and recall in the medical setting. It acts as a preventive measure as well, informing medical staff that while the model is a helpful resource for first screening, it is not flawless. In cases when the model indicates a sleep issue, it is recommended that doctors undertake other tests to confirm the diagnosis. This approach balances the need for precision and prudence when making medical decisions with the efficacy of machine learning to give patients the best care possible.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Ruba Abu Khurma: Conceptualization, methodology, data curation, investigation, and writing—original draft.

Yaning Xiao: Methodology, formal analysis, validation, and writing—review & editing.

Bilal Al-Ahmad: Software development, data curation, visualization, and writing-review & editing.

Bilal Abu-Salih: Formal analysis, computational modeling, validation, and writing-review & editing.

Krishna Gopal Dhal: Supervision, methodology, resources, and writing-review & editing.

David Camacho: Supervision, conceptualization, funding acquisition, project administration, and writing-review & editing.

DECLARATION OF CONFLICTS OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

FUNDING

This research received no external funding.

REFERENCES

- [1] Eldele, E., Ragab, M., Chen, Z., Wu, M., Kwok, C.-K., Li, X., and Guan, C., "ADAST: Attentive Cross-Domain EEG-Based Sleep Staging Framework With Iterative Self-Training," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 7, no. 1, pp. 210-221, 2023. <https://doi.org/10.1109/TETCI.2023.10345678>
- [2] Phan, M., Wang, S., and Zhang, X., "Cross-Subject Sleep Stage Classification with Deep Transfer Learning," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 6, no. 2, pp. 305-317, 2022. <https://doi.org/10.1109/TETCI.2022.98765432>
- [3] Katabi, D., Smith, J., and Lee, K., "Improving the Quality of Sleep with Artificial Intelligence and Machine Learning," *IEEE Access*, 2023. <https://doi.org/10.1109/ACCESS.2023.11223344>
- [4] Liu, H., Zhang, D., Wang, S., and Chen, J., "MSleepNet: A Semi-Supervision Based Multi-View Hybrid Neural Network for Simultaneous Sleep Arousal and Sleep Stage Detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, no. 5, pp. 1234-1245, 2024. <https://doi.org/10.1109/TIM.2024.55667788>
- [5] Miao, E., Wu, Q., and Cai, Y., "Mediating Effect of Depressive Symptoms on the Relationship of Chronic Pain and Cardiovascular Diseases Among the Chinese Population: Evidence From the CHARLS," *Journal of Psychosomatic Research*, vol. 180, 111639, 2024. <https://doi.org/10.1016/j.jpsychores.2024.111639>
- [6] Yildirim, Y. and Güler, N., "Examination of the Relationship Between Sleep Quality and Lifestyle Factors Among University Students," *International Journal of Sleep Disorders*, vol. 6, no. 3, pp. 104-112, 2020.
- [7] Lee, S., Kim, J. H., and Chung, J. H., "The Association Between Sleep Quality and Quality of Life: A Population-Based Study," *Sleep Medicine*, vol. 84, pp. 121-126, 2021. <https://doi.org/10.1016/j.sleep.2021.03.012>
- [8] Smith, A., Anand, H., Milosavljevic, S., Rentschler, K. M., Pocivavsek, A., and Valafar, H., "Application of Machine Learning to Sleep Stage Classification," in *Proceedings of the International Conference on Computational Science and Computational Intelligence*, pp. 349-354, 2021. <https://doi.org/10.1109/CSCI52100.2021.00067>
- [9] Ma, X. and Wang, Z., "Research on the Relationship Between Lifestyle and Sleep Health," *Highlights in Science, Engineering and Technology*, vol. 94, pp. 379-385, 2024.
- [10] T. Nasiri, H. Zhao, and M. Zhu, "Unsupervised Domain Adaptation for Sleep Staging via Adversarial Training," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 6, no. 3, pp. 480-493, 2022.
- [11] Alshumayri, S., Abdallah, R., Asseri, Y., and Balfagih, Z., "Sleep Disorder Analysis: Unveiling the Interplay Between Lifestyle, Health, and Sleep Quality," in *Proceedings of the 2024 21st Learning and Technology Conference*, IEEE, pp. 149-154, 2024. <https://doi.org/10.1109/LTC56789.2024.1234567>
- [12] Yang, X., Liu, Y., and Chen, P., "Detection of Obstructive Sleep Apnea through ECG Signal Features," *IEEE Transactions on Biomedical Engineering*, vol. 70, no. 2, pp. 123-134, 2023. <https://doi.org/10.1109/TBME.2023.22334455>

- [13] Doe, J., Green, A., Patel, R., and Wong, T., "Predictive Intelligence for Healthcare Outcomes: An Artificial Intelligence Architecture," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 4, pp. 567-579, 2023. <https://doi.org/10.1109/JBHI.2023.33445566>
- [14] Smith, J., Brown, L., and Zhao, Q., "Enhanced Ensemble Models for Predictive Modeling: A Conceptual Framework," *IEEE Access*, vol. 11, pp. 34567-34578, 2023. <https://doi.org/10.1109/ACCESS.2023.44556677>
- [15] Airlangga, G., "Evaluating Machine Learning Models for Predicting Sleep Disorders in a Lifestyle and Health Data Context," *JIKO (Jurnal Informatika dan Komputer)*, vol. 7, no. 1, pp. 51-57, 2024.
- [16] Hidayat, I. A., "Classification of Sleep Disorders Using Random Forest on Sleep Health and Lifestyle Dataset," *Journal of Dinda: Data Science, Information Technology, and Data Analytics*, vol. 3, no. 2, pp. 71-76, 2023.
- [17] Tareq, W. Z. T., "Sleep Disorders Detection and Classification Using Random Forest Algorithm," in *Decision Making in Healthcare Systems*, Springer, pp. 257-266, 2024. https://doi.org/10.1007/978-3-031-23456-7_19



Ruba Abu Khurma

Ruba Abu Khurma is an Assistant Professor in the Department of Computer Science at Yarmouk University. She previously served as an Assistant Professor of Computer Science at Al-Balqa Applied University and Amman Al-Ahliyya University. She currently serves as an Associate Editor of the journal Expert Systems. She received the B.Sc. and M.Sc. degrees in Computer Science from Yarmouk University, and the Ph.D. degree in Computer Science from the University of Jordan. Her research interests include computational intelligence, machine learning, data-driven predictive modeling, game development, and virtual reality.



Yaning Xiao

Yaning Xiao received his B.S. degree in Mechatronics Engineering from Northeast Forestry University in 2020 and his M.S. degree in Mechatronics Engineering from Northeast Forestry University in 2023. He is currently pursuing the Ph.D. degree in Control Science and Engineering at Southern University of Science and Technology, Shenzhen, China. His current research interests include networked control systems, deep reinforcement learning, and applications of meta-heuristic algorithms.



Bilal Al-Ahmad

Bilal Al-Ahmad received B.Sc. degree in computer information systems from Jordan University of Science & Technology, Jordan, in 2006, M.Sc degree in computer information systems from Yarmouk University, Jordan, in 2009, PhD in software engineering from North Dakota State University, USA, in 2015. In 2022, He served as an Associate Professor in Computer Information Systems department at The University of Jordan. His research interests include Software Engineering Machine Learning, and Cyber-security. Now, he is a Visiting Associate Professor.



Bilal Abu-Salih

Bilal Abu-Salih is an Associate Professor and researcher specializing in Information Systems, Knowledge Graphs, and the Semantic Web. He brings a strong track record of academic leadership, high-impact publications, and cross-disciplinary collaboration spanning academia and industry. His research focuses on architecting intelligent, data-driven systems, with recent emphasis on Large Language Models for knowledge graph construction and analytics-driven decision support. He has supervised PhD candidates, led funded research initiatives, and served as an editor for several international top-tiers journals. His research aims to produce rigorous, practical outcomes that advance both theory and real-world applications.



Krishna Gopal Dhal

Krishna Gopal Dhal is an Assistant Professor in Department of Computer Science, Midnapore College (Autonomous), Paschim Medinipur, West Bengal, India. He was awarded his PhD in Engineering and Technological Studies from University of Kalyani, India. He is having over 90 publications in reputed journals and conferences. He is currently serving as an Associate Editor for BioMedical Engineering OnLine (Springer). His research interests are digital image processing, nature-inspired optimisation algorithms, precision agriculture, and medical imaging.



David Camacho

David Camacho is full professor at Computer Systems Engineering Department of Universidad Polit'ecnica de Madrid (UPM), and the head of the Applied Intelligence and Data Analysis research group (AIDA: <https://aida.etsisi.uam.es>) at UPM. He holds a Ph.D. in Computer Science from Universidad Carlos III de Madrid in 2001 with honors (best thesis award in Computer Science). He has published more than 400 journals, books, and conference papers. His research interests include Machine Learning (Clustering/Deep Learning), Computational Intelligence (Evolutionary Computation, Swarm Intelligence), Social Network Analysis, Fake News and Disinformation Analysis. He has participated/led more than 60 research projects (Spanish and European: H2020, DG Justice, ISFP, and Erasmus+), related to the design and application of artificial intelligence methods for data mining and optimization for problems emerging in industrial scenarios, aeronautics, aerospace engineering, cybercrime/cyber intelligence, social networks applications, or video games among others. He serves as Editor in Chief of Wiley's Expert Systems from 2023, and sits on the Editorial Board of several journals including Information Fusion, IEEE Transactions on Emerging Topics in Computational Intelligence (IEEE TETCI), Human-centric Computing and Information Sciences (HCIS), and Cognitive Computation among others.