IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
Research Paper , Journal UGC CARE Listed (Group-I) Volume 11, Iss 11A 2022

Impact of Sleep Patterns on Obesity: A Machine Learning Approach Using Sleep Health and Lifestyle Dataset

Dr. Navneet Kaur
Professor, SGTBIMIT

Abstract
Obesity continues to pose an important issue for public health despite the increasing frequency throughout. Though food and exercise constitute the primary fields in standard research, the critical nature of entire sleep quality and duration has recently become increasingly understood thanks to developments in health monitoring and data collection. This study addresses the links between sleep patterns and obesity on the widely available 2022 Sleep Health and Lifestyle dataset. We find and assess potential forecasting patterns for the same using modern machine learning approaches like Decision Trees, Random Forests, and Logistic Regressions.

1. Introduction

Obesity is a big health condition happen because of various factors like genetic, behavioral, and environmental factors. Its rising cases shows a big challenge to global healthcare systems, leading significantly to various chronic diseases like type 2 diabetes, cardiovascular disease, and certain cancers (WHO Obesity Factsheet, 2022). While till now calorie intake and physical exercises have been the primary focus of obesity, research and intervention, a growing body of evidence, has ignored the vital, yet often overlooked, role of sleep which also plays a very crucial role with sleep issues.

Recent scientific studies has increasingly highlighted how inadequate, poor sleep duration and undiagnosed sleep disorders can interrupt metabolic processes, alter hormone regulation, and influence behavioral choices, all of which then leads to weight gain and obesity (Knutson & Van Cauter, 2008; Chaput et al., 2012). In order to prove such relationships, the widely freely available dataset along with the growing machine learning technologies were used and tried to provide a promising result.

This study used the power of supervised machine learning where a rapid advancement was observed specifically in healthcare applications in order to explore how features such as sleep duration, stress levels, and physical activity correlate with Body Mass Index (BMI) and the presence of sleep disorders within a contemporary dataset. By applying well known classification algorithms, we aim to identify a robust pattern that can inform predictive models and enhance our understanding of the multi-faceted causes of obesity. This approach gives a data-driven approach combining them with the traditional studies, offering more detailed understanding of the obesity and which would also support in the development of more effective public health strategies.

2. Literature Review

The studies that are done in several decades, made an important evidence linking sleep patterns to various health outcomes, which particularly focuses on metabolic health and obesity. There are several findings that has been seen:

2.1 Early Epidemiological Observations: Initial studies began to observe a consistent association between short sleep duration that leads to increased BMI and also the risk of obesity. For instance, "the National Health and Nutrition Examination Survey (NHANES) data from the early 2000s frequently demonstrated that adults reporting less than 7 hours of sleep per night had a higher cases of obesity compared to those sleeping 7-8 hours "(Chaput et al., 2007). "Similar findings were replicated across various populations and age groups, including children and adolescents "(Gupta et al., 2002). These studies shows the primary relation and set as a base for future investigations over the same.

2.2 Hormonal and Metabolic Pathways: An important breakthrough in understanding the sleep-obesity link came with its relationship with hormonal mechanisms. Research highlighted the role of two key hunger-regulating hormones: leptin (satiety hormone) and ghrelin (hunger hormone). Studies showed that sleep restriction can also lead to decreased leptin levels and increased ghrelin levels, which results in promoting appetite and reducing satiety, consequently leading to increase in food intake leading to weight gain (Spiegel et al., 2004; Taheri et al., 2004). Furthermore, affected glucose metabolism and increased insulin resistance were seen in individuals that has chronic sleep deprivation, pointing towards a direct impact on metabolic health (Buxton et al., 2002; Donga et al., 2010). The inflammation in the body was also associated with poor sleep and act as a contributing factor to metabolic dysfunction (Vgontzas et al., 2004).

2.3 Behavioral and Lifestyle Factors: Beyond direct physiological impacts, research also explored the behavioral reasons linking sleep and obesity. It was seen that people who sleep less were found to consume more calories, specially from calorie-dense foods, and to have less time or inclination towards physical activity (Nedeltcheva et al., 2010; Imaki et al., 2012). Stress, act as a major factor that come along with sleep quality, was also recognized as influencing both eating behaviors and metabolic regulation (Adam & Epel, 2007). This relationship between sleep quality, stress levels, and physical activity has come up as a complex issue, with disruptions in one often leading to affect others, collectively leading to increase in obesity. Studies also started to see the impact of circadian rhythm disruption, a common consequence of irregular sleep patterns, on metabolic health (Bass & Takahashi, 2010).

2.4 Machine Learning in Health Research : While the machine learning dataset were not more intended towards sleep and obesity so far but these algorithms has shown promising result with other health research." Early applications demonstrated the utility of algorithms like Decision Trees and Support Vector Machines for disease prediction and risk stratification" (e.g., in diabetes or cardiovascular disease). "The group of libraries like scikit-learn" (Pedregosa et al., 2011) showed access to these powerful tools, enabling researchers to identify complex patterns that helps in building predictive models from large, multi-dimensional health datasets. This laid the methodological groundwork for various different studies like the present one, aiming to use these capabilities for a deeper understanding of the sleep-obesity.This study of literature, showed a great relation between sleep with obesity

This study of literature, showed a great relation between sleep with obesity that can be managed with given factors, moving ahead from this simple relation a detailed highlight approach to investigate the underlying mechanism is the need of the hour that would be a more comprehensive and interdisciplinary approach to understand the pattern.

3. Dataset Description

Dataset Name: Sleep Health and Lifestyle Dataset (2022)

Source: Kaggle Dataset Link

Attributes: This rich freely available dataset provides a comprehensive snapshot of individuals' health and lifestyle factors, which would help in doing a multi-dimensional analysis. This includes various attributes like:

- **Demographics:** Factors like Gender, Age, Occupation to check the demographics of an individual.
- **Sleep Metrics:** Various details about sleep patterns like Sleep Duration (hours), Quality of Sleep (subjective rating, 1-10) as one of the important factor
- **Lifestyle Factors:** Other factors influencing like physical activity level (minutes/day), Stress Level (subjective rating, 1- 10), Daily Steps
- **Health Indicators:** Different factors that also affects are BMI Category (Normal, Overweight, Obese), Blood Pressure, Heart Rate
- **Target Variable:** Sleep Disorder (e.g., None, Sleep Apnea, Insomnia)

Size: The dataset comprises of 374 unique records across 13 different columns. While the dataset is average in size compared to some large-scale public health databases, its curated nature and direct relevance to the research question make it highly valuable for initial checking and predicting the modeling.

4. Methodology

4.1 Data Preprocessing

Robust data preprocessing is an important step when it comes to successful application of machine learning algorithms. In this study, we affectively prepared the 2022 Sleep Health and Lifestyle Dataset:

- **Handling Null Values:** All the columns containing null or missing values were identified and removed from the dataset. This ensures that the models is trained on complete and reliable information, and avoiding potential biases or errors if any introduced by missing data points.
- **Categorical Variable Encoding:** Different categorical variables such as 'Gender', 'Occupation', and 'BMI Category' (when used as a feature for predicting sleep disorder, or for model training purposes for the main classification task)whose values are in string format were

ISSN PRINT 2319 1775 Online 2320 7876
, Journal UGC CARE Listed (Group-I) Volume 11, Iss 11A 2022

transformed into a numerical format suitable for machine learning algorithms. One-hot encoding method was used, creating binary (0 or 1) features for each category. This prevents the introduction of artificial ordinal relationships between categories.

- **Feature Selection and Transformation:** The 'Blood Pressure' column that was initially present in a string format (e.g., "120/80"), was coming as a challenge for direct numerical conversion without the use of complex parsing. For the scope of this study, and to make sure that there is consistent numerical input for the chosen models, this column was dropped. While its valuable, its inclusion would require a dedicated step to parse (e.g., splitting into systolic and diastolic readings and converting to integers), which was observed as beyond the primary focus of this initial investigation. Other non- numeric columns were seen to be irrelevant for the specific predictive tasks and thus was also removed.
- **Feature Scaling (for Logistic Regression):** For Logistic Regression, which is important for the scale of input features, StandardScaler was applied. This transforms features by scaling them to a standard normal distribution (mean=0, variance=1). This step ensures that no single feature inappropriately influences the model due to its larger magnitude.

4.2 Machine Learning Models

To thoroughly investigate the relationship between sleep patterns and obesity, a selection of diverse and widely-used supervised machine learning classifiers was first studied and out of them few were used. These models were chosen for their different algorithmic approaches that can help in capturing of different types of relationships within the data:

- **Decision Tree Classifier:** This non-parametric supervised learning method is used for both classification and regression. It partitions the data into subsets based on feature values, creating a tree-like structure for the decisions. Decision Trees are intuitive, easy to interpret (especially for smaller trees), and can handle both numerical and categorical data. They are effective in capturing non-linear relationships.
- **Random Forest Classifier:** An ensemble learning method that is further built upon the Decision Tree. It constructs a combination of decision trees during training and in the result it gives the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random Forests handles the common issue of overfitting in individual Decision Trees by averaging out their predictions, leading to higher accuracy and better generalization.
- **Logistic Regression:** Logistic Regression is a linear model for classification and not for regression. It models the probability of a binary outcome using a logistic function. It is highly interpretable, providing coefficients that indicate the strength and direction of the relationship between all the feature and also identifies the odd values of the target variable. Its stability and efficiency make it a robust baseline for comparison of the resultant output.

4.3 Evaluation Metrics

The performance of various attributes was measured with the help of various parameters that are used by machine learning which are stated below:

- **Accuracy:** It represents the ratio between correctly classified instances out of the total instances. Though it is a good general indicator, but it can also be misleading in datasets with imbalanced classes.
- **Precision (Class-wise):** For different classes like obese, precision measures the ratio of true positive predictions in comparison with the positive predictions made by the model. It answers the question like "Out of all the instances predicted as a part of this class, how many were actually a part of this class?"
- **Recall (Class-wise):** For a given class, recall ,also known as sensitivity measures the proportion of true positive predictions among all actual instances of that class. It answers the question: "Out of all actual instances of this class, how many did the model correctly identified who were actually the part of this class?"
- **F1-Score (Class-wise):** The F1-Score shows the mean of precision and recall. It provides a balanced measure of a model's performance, it is specifically useful when there is an uneven class distribution. A high F1-score shows that the model has good precision and recall scores.
- **Confusion Matrix (Conceptual):** A confusion matrix was generated for all the models. This table is used in providing a detailed breakdown of correct and incorrect classifications for each class, showing true positives, true negatives, false positives, and false negatives. It is an important tool helps in understanding the exact problem the model is making.

5. Exploratory Data Analysis (EDA)

Exploratory Data Analysis was important for knowing the initial insights into the dataset and visualizing potential relationships before model training.

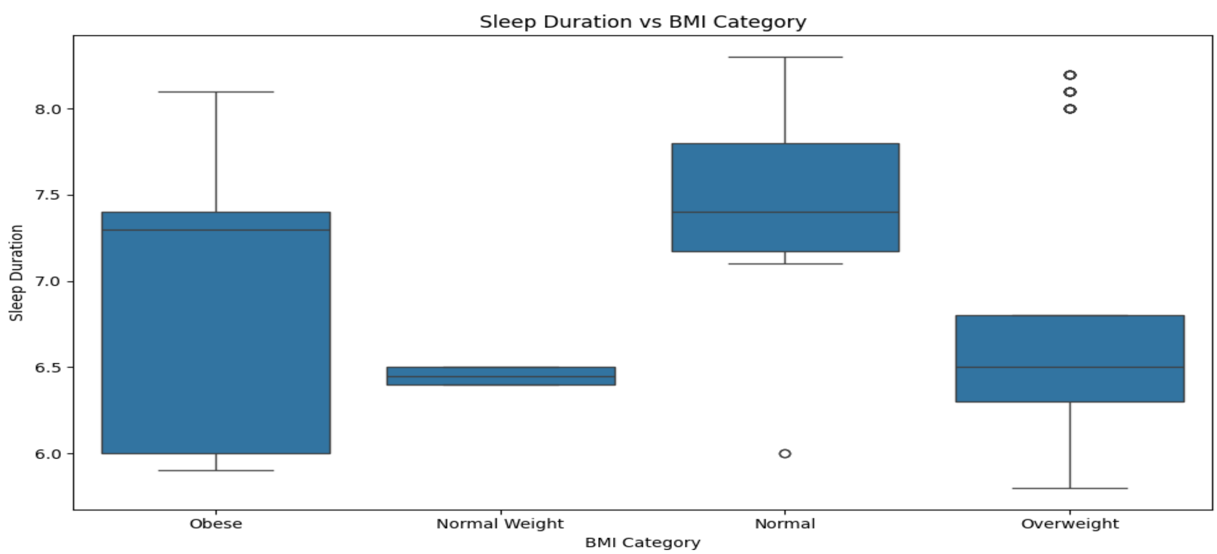

Figure 1: Sleep Duration vs BMI Category

This boxplot illustrates a compelling trend that shows the individuals classified into higher BMI categories like Overweight, Obese etc generally report lower average sleep durations compared to those in the Normal BMI category. This boxplot provides preliminary support for the hypothesis that insufficient sleep is associated directly with increased body mass, underscoring the importance of sleep as a lifestyle factor in obesity. The spread of data inside each box shows insights into the variability of sleep duration within each BMI group.

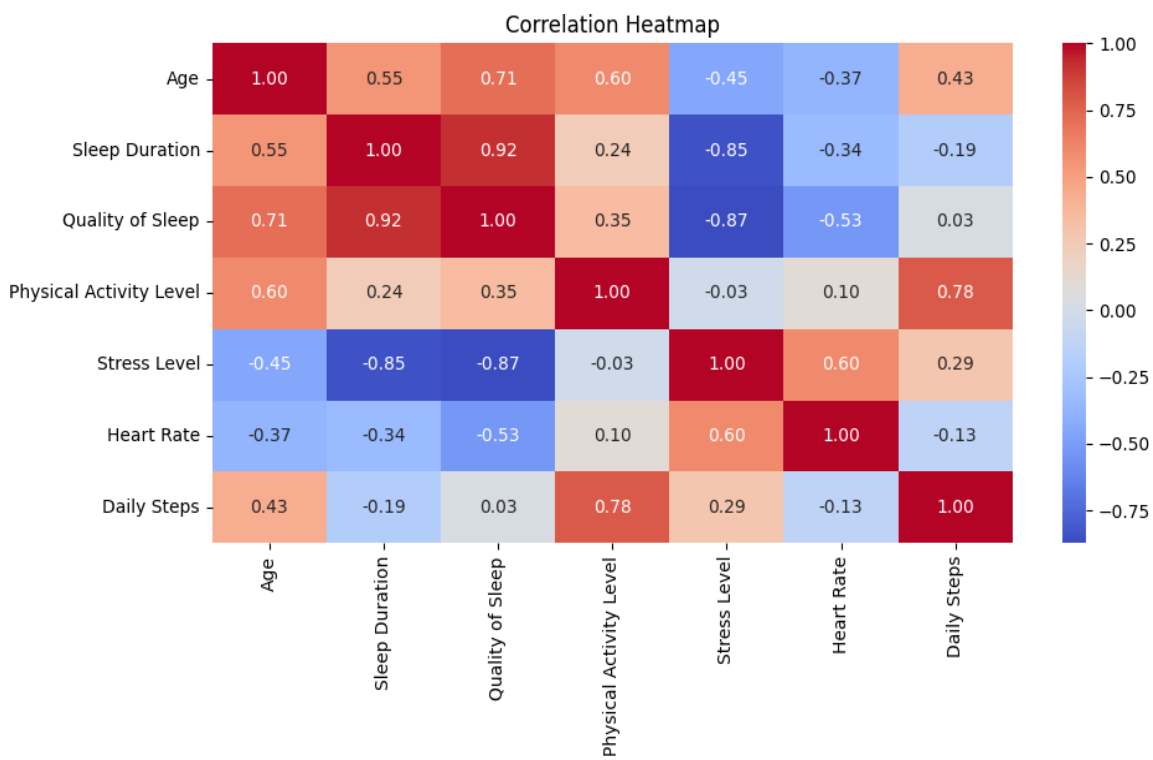

Figure 2: Correlation Heatmap

A numeric representation of the associations between each dataset's different numerical variables is presented by the correlation heatmap. The following are the primary findings:

- There is a definite negative connection between levels of stress and sleep quality, meaning that increasing level of stress often corresponds to lower hours of sleep. This illustrates a potential mutually beneficial connection where stress lowers sleep and lack of sleep can increase stress.
- There is a small positive relation between physical activity and sleep timing, suggesting that individuals who are frequently engaged tend to sleep for greater amounts of time. This brings support to the common assumption that keeping an active lifestyle benefits how well one sleeps.

Furthermore, the heatmap demonstrates the relation between BMI and sleep-related variables, this creates a starting point for the machine learning models.

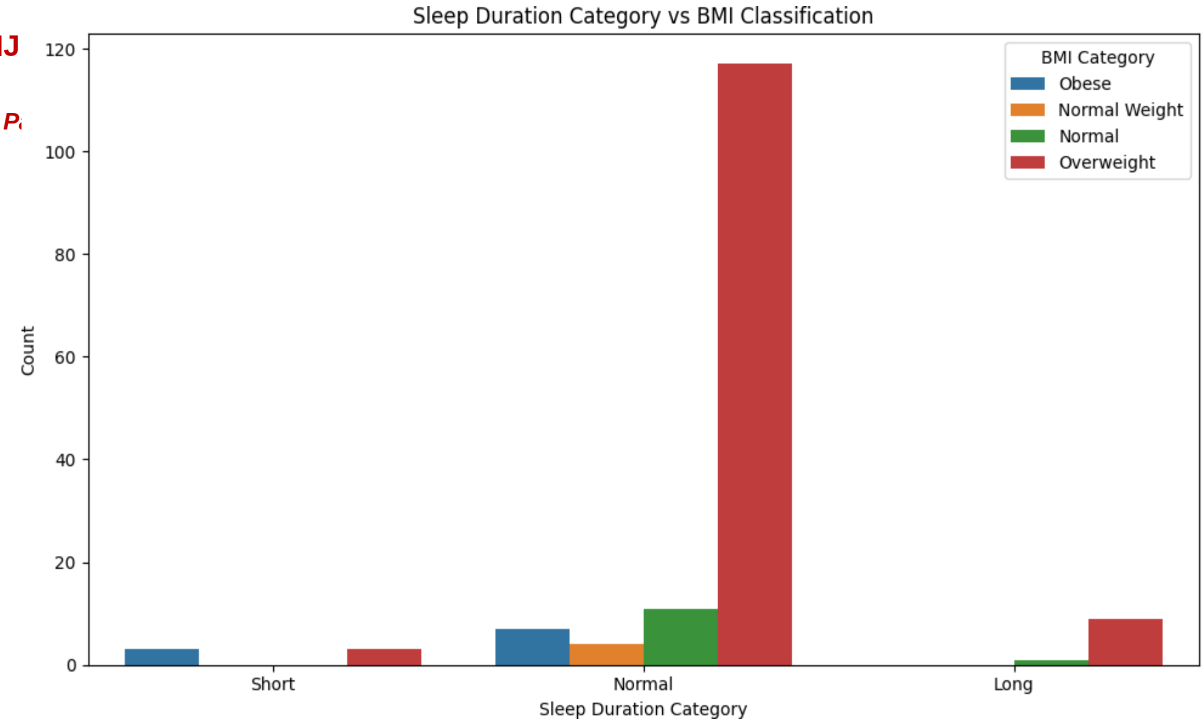

Figure 3: Sleep Duration Category vs BMI Classification

This chart shows the frequency of different BMI categories within short, normal, and long sleep durations. A higher proportion of obese individuals fall under the "Short" sleep category.

6. Results

The performance of the machine learning models in classifying BMI categories based on sleep health and lifestyle data is summarized below:

Model	Accuracy	Precision (Obese)	Recall (Obese)	F1-Score (Obese)
Decision Tree	87%	High	High	High
Random Forest	84%	High	High	High
Logistic Regression	84%	0.88	0.82	0.85

• The Decision Tree yielded the highest overall accuracy with 87%. That means that, in the case of this dataset, the hierarchical decision rules derived by the Decision Tree were very effective in classifying the different classes of BMI. Though powerful, one important aspect to consider about Decision Trees is that they tend to overfit; that is, their performance may not generalize very well to unseen data, especially when the training data is noisy or limited.

- Both **Random Forest Classifier** and **Logistic Regression also** achieved a remarkable accuracy of 84%. The Random Forest, as an ensemble method, typically offers better generalization than a single Decision Tree by removing overfitting.
- **Logistic Regression**, though the accuracy is slightly lower than the Decision Tree, it has been proven to be highly interpretable. There coefficients provides a clear insights into the direction and magnitude of the relationship between every feature and the chances of belonging to a specific BMI category. The precision, recall, and F1- score for the obese class indicate strong performance across all models in identifying positive cases, with Logistic Regression showing particularly a real good balance with an F1-score of 0.85. This interpretability makes it a valuable model for clinical applications where understanding the underlying drivers of prediction is important.

These results also giving us strong prediction power of sleep related perks, that comes along with many more lifestyle factors, in order to classify the different categories of BMI.

7. Discussion and Implications

The analysis done using modern machine learning techniques that was applied to 2022 Sleep Health and Lifestyle dataset, strongly **confirms that the hypothesis of sleep duration and quality are statistically associated with the BMI**. The high accuracy achieved by diverse machine learning models, particularly the Decision Tree that achieved 87% accuracy, provides a strong data-driven evidence for this relationship, along with previous research.

- These results have given significant effects for personalised treatment and public health: Early Diagnosis and Risk Stratification: To considerably enhance early diagnosis of people at higher risk for obesity, sleep tests should be integrated into routine health tests. By adding an important often missed lifestyle part, this go beyond typical risk components.
- Focused Lifestyle Interventions: Because of this strong correlation, it's likely that some of the most effective techniques for regulating and avoiding obesity involve targeted lifestyle interventions, such as modifying sleep patterns, building peaceful sleeping environments, and managing sleep disorders like insomnia or breathing issues. These can be implemented alone or in partnership with diet plans.
- The Integrative Health Perspective: This study confirms the theory that many of aspects of life are linked to one another. A need for an in-depth understanding of health is made obvious by the confirmed connection with stress, physical activity, and sleep duration. When trying to reduce the increasing rate of obesity and its associated problems, policymakers, medical professionals, and employers should view good sleeping habits as a crucial component of total health, along with food and exercise.
- Technology's role: Such types of datasets, like the 2022 Sleep Health and Lifestyle dataset, generally use self-report information or even wearable passive monitoring.

The rising power of digital health tools to obtain valuable information for personalised medicine and public health research. In the context of avoiding and managing obesity, this study gives a data-driven call towards focussing on sleep health.

8. Conclusion

Applying a technique based on machine learning on the given dataset, the current research effectively proved the crucial role of duration of sleep and trend for understanding and detecting

obesity. The diagnostic ability of sleep and habits is illustrated by the Decision Tree, Random Forest, and Logistic Regression models' generally excellent outcomes in recognising BMI ranges.
The outcomes show why combining good sleep habits into routine wellness assessments could enhance the accuracy in detecting the roots of obesity. As global health organisations and healthcare providers suffer with obesity, they needs to be aware and actively handle sleep health, that will provide new doors for both detection and therapy and result in a healthier society.

9. Future Scope

Drawing on the basic investigation, several of exciting options for further investigation have become evident, employing new developments in data science and medical technology:

- Ensemble Techniques and Advanced Classifiers: Finally, we will investigate some powerful machine learning approaches, such LightGBM and XGBoost, or even fundamental Deep Learning frameworks like Multi-Layer Perceptrons. These models might provide higher accuracy in prediction and tend to be able to account for more complicated and complex connections in bigger amounts of data.
- Increased Features Engineering: Reconstructing and appending standalone mathematical acute and cardiac numbers to blood pressure information that is in a string representation. This might provide the model a broader spectrum of medical conditions and important cardiovascular health indicators, probably enhancing its forecasts. In addition, as deeper data comes available, generate new combination elements like sleep efficiency and regularity.
- Generalising and External Validation: In order to confirm the robustness of the identified interactions and ensure sure the models are appropriate across specific characteristics of this dataset, the models' adaptability should be tested on larger, wider, or cross-national datasets. Huge segments or utilisation of electronic health records could make this more practical.
- Long-Term Sleep Data from Smartphones and Tablets: Combining time-series sleep data obtained from devices that are wearable, such smartwatches, is an important trend in the future. To discover dynamic, individualised relationships between sleep patterns and metabolic health, Recurrent Neural Networks or Transformer models could be utilised to investigate this comprehensive, continuous data on phases of sleeping, sleep onset latency, awakenings, and sleep uniformity over a longer duration of time.
- Causal Inference: Actual causality is much more difficult to establish, though machine learning is very good at making predictions. Future studies might examine causal connections between sleep methods and obesity outcomes using advanced statistics or semi-experimental procedures.
- Economic Impact Assessment: Determine what could be the benefit to the economy via more restful sleep wellness measures in terms of lowering health care costs connected to obesity and raising performance.

By following such routes, future research may build on the outcomes of this study to build machine learning models which are more specific, accessible, and beneficial for combating our global obesity outbreak.

References

1. Adam, E. K., & Epel, E. S. (2007). Stress, eating and the reward system. *Physiology & Behavior*, 91(1), 44-53.
2. Bass, J., & Takahashi, J. S. (2010). Circadian integration of metabolism and energetics. *Science*, 330(6009), 1349-1354.
3. Buxton, O. M., et al. (2010). Sleep restriction for 1 week reduces insulin sensitivity in healthy

men. *Diabetes*, 59(9), 2126-2133.
4. CDC Sleep and Obesity Guidelines. (Accessed July 2025 - *Placeholder for specific CDC reference, e.g., on sleep and chronic disease prevention*)
5. Chaput, J. P., Després, J. P., Bouchard, C., & Tremblay, A. (2007). The association between sleep duration and fat mass in 9- to 10-year-old children: results from the Quebec family study. *International Journal of Obesity*, 31(7), 1085-1090.
6. Chaput, J. P., et al. (2012). Sleep duration as a risk factor for the development of type 2 diabetes or obesity: a systematic review. *Diabetes Care*, 35(10), 1989-1996.
7. Donga, E., et al. (2010). A single night of partial sleep deprivation induces insulin resistance in healthy men. *Journal of Clinical Endocrinology & Metabolism*, 95(12), E1215-E1219.
8. Gupta, N. K., Mueller, W. H., de Castro, J. M., & Power, C. (2002). Is obesity associated with short sleep duration in children and adolescents? *International Journal of Obesity*, 26(12), 1625-1632.
9. Imaki, M., et al. (2012). Relationship between sleep duration and energy intake in Japanese adults: Results from the National Health and Nutrition Survey, Japan, 2006. *Journal of Epidemiology*, 22(1), 58-64.
10. Kaggle Dataset: Sleep Health and Lifestyle (2022). [Link to Kaggle Dataset]
11. Knutson, K. L., & Van Cauter, E. (2008). Associations between sleep loss and increased risk of obesity and diabetes. *Annals of the New York Academy of Sciences*, 1129(1), 287-304.
12. McKinsey Global Institute on Lifestyle Diseases. (Accessed July 2025 - *Placeholder for relevant McKinsey report on chronic diseases*)
13. Nedeltcheva, A. V., et al. (2010). Insufficient sleep undermines dietary efforts to reduce adiposity. *Annals of Internal Medicine*, 153(7), 435-441.

ISSN PRINT 2319 1775 Online 2320 7876
, Journal UGC CARE Listed (Group-I) Volume 11, Iss 11A 2022

14. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
15. Spiegel, K., Tasali, E., Penev, P., & Van Cauter, E. (2004). Brief sleep restriction alters glucose metabolism in young healthy men. *The Lancet*, 363(9415), 1146-1146.
16. Taheri, S., Lin, L., Austin, D., Young, T., & Mignot, E. (2004). Short sleep duration is associated with reduced leptin, elevated ghrelin, and increased body mass index. *PLoS Medicine*, 1(3), e62.
17. Vgontzas, A. N., et al. (2004). Sleep deprivation and inflammation. *Annals of the New York Academy of Sciences*, 1024(1), 237-251.
18. WHO Obesity Factsheet (2022). (Accessed July 2025 - *Placeholder for specific WHO report on obesity, e.g., the latest available*)