OXFORD JOURNAL
OF STUDENT
SCHOLARSHIP

Artificial Intelligence in Sleep Medicine: Predicting Sleep Quality Using Lifestyle and Demographic Indicators

Artur Yamanov
yamanovartur@gmail.com

ABSTRACT

Sleep disorders are increasingly common and are associated with impaired bodily functioning and quality of life. Early identification of individuals at risk is therefore essential but challenging as it depends on complex interactions between demographic and lifestyle factors. This study investigates whether machine learning (ML) methods can predict sleep quality from demographic indicators. Using a publicly available Sleep Health and Lifestyle dataset, we analysed demographic variables (age) and lifestyle factors (sleep duration, self-reported sleep quality, physical activity level) alongside labels indicating the presence of no disorder, insomnia or sleep apnea. We implemented and compared several regression-based machine learning models. Linear Regression, Decision Tree Regressor, Random Forest Regressor, K-Neighbors Regressor and an MLP Regressor neural network to estimate a continuous sleep quality score. To enhance accuracy, we further designed a weighted system inspired by the Multiplicative Weight Update (MWU) framework, allowing models with lower prediction error to contribute more strongly to the final output. Model performance was evaluated using Mean Squared Error (MSE), Mean Absolute Error (MAE).The study demonstrates the feasibility of using combined lifestyle and demographic metrics with machine learning techniques to estimate sleep quality, with the decision tree regressor showing great success. Overall, models showed success in predicting possible sleep quality according to the weighted systems. This work provides evidence that AI can be used as a tool for more personalised screening for insomnia and sleep apnea in everyday clinical situations.

INTRODUCTION

In this study, we focused on the two most common sleep disorders: Insomnia and sleep apnea.

Sleep disorders encompass a range of problems with the quality, timing and the amount of sleep a person gets, which result in daytime stress and impairment in functioning (American Psychiatric Association, 2022). Common sleep disorders include insomnia, sleep apnea, narcolepsy, parasomnias, restless leg syndrome and circadian rhythm disorders. These conditions can frequently lead to chronic fatigue, impaired cognition, hormonal imbalances as well as a weakened immune function and long term health consequences such as cardiovascular disease, diabetes and mental health conditions.

February 2026
Vol 4. No 1.

Insomnia: is the inability to stay asleep or fall asleep and is one of the most common sleep disorders in the world. Its key features include feeling unrefreshed in the morning and reduced cognitive ability as well as mind fog. Being a common sleep disorder numerous studies have demonstrated that insomnia is linked to elevated stress, heart disease and poor academic and work performance.

Sleep Apnea: is more clinically severe compared to insomnia in which people experience repeated interruptions during sleep. Typical signs include loud snoring, pauses in breath (apneas) as well as excessive daytime sleepiness. Sleep apnea is more medically significant as it increases risk of hypertension and cardiovascular disease- which can create further severe complications- diabetes and accidents due to daytime fatigue.

In recent years sleep disorders have become a major global health challenge and an area of concern. Modern lifestyle factors such as: consistent late night screen exposure, increased stress, irregular work schedules, diet changes and sedentary behavior, have contributed to a rapid rise in sleep related illnesses over the coming decades. The lack of proper sleep affects millions of people worldwide and is associated with reduced productivity, higher accident risk, and elevated rates of chronic disease and other factors listed above. However, early detection of sleep disorders can be used to prevent these long term complications, but identifying them accurately is very challenging as it requires understanding complex interactions among biological, behavioral, and environmental factors.

Sleep disorders continue to rise globally, resulting in researchers increasingly turning to artificial intelligence to better understand and predict these conditions. Artificial intelligence has already managed to transform several areas of healthcare as it is efficient in identifying patterns in clinical data that are too subtle or complex for traditional statistical methods or human observation (Faust et al., 2019). In the field of sleep research, machine learning (ML) models have shown strong potential to analyse physiological, behavioural and lifestyle data to classify sleep stages, detect abnormalities and predict disorder risk. For example, (Radha et al., 2019) demonstrated that ML based systems can accurately identify sleep stage transitions and nocturnal disturbances from electroencephalogram (EEG) and other biosignals, offering a faster as well as a more scalable alternative to manual scoring. Other studies have also primarily focused specifically on sleep apnea, where models such as the Random Forest (RF) , gradient boosting algorithms and neural networks have been able to predict apnea severity using biometric and lifestyle features (Alqurashi et al., 2021; Zhang et al., 2018).

These findings highlight a broader trend that AI methods are increasingly capable of capturing the complex interactions among biological, behavioural and environmental factors that contribute to sleep disorders. As a result, we can conclude that machine learning approaches have the ability to provide an important framework for developing early screening tools that may improve diagnosis, reduce long term health risks and support personalised interventions, aligning closely with the objectives of the present study.

Sleep disorders such as insomnia and sleep apnea are well known to have a substantial impact on overall sleep quality. While clinical diagnosis of these disorders typically requires specialized assessment, self-reported sleep quality can be a meaningful and practical indicator of sleep health that reflects lifestyle factors and sleep disturbances. Thus, modelling sleep quality offers a meaningful way to study sleep health and its links with lifestyle and demographic indicators even when specific clinical diagnoses are not the primary prediction target.

In this study, we aim to predict the self-reported quality of sleep score, which is also a continuous variable provided in the dataset, using regression-based machine learning models. Rather than predicting specific diagnostic categories or clinical probabilities we focused on estimating sleep quality from the lifestyle and demographic indicators.

MATERIALS AND METHODS

Dataset Description

https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset

The dataset used in this study contains detailed information about individuals' gender, age, occupation, sleep duration, quality of sleep, physical activity level, stress levels, BMI category, blood pressure, heart rate, daily steps, and the presence or absence of sleep disorders. It is composed of 400 rows and 13 columns, which cover a wide range of variables related to sleep and daily habits.

The dataset is suitable for AI analysis for a variety of reasons. It contains clinical relevance by having core features such as heart rate, BMI, stress and sleep duration which are medically recognized predictors of sleep disorders. The dataset offers multidimensionality as it integrates demographic, behavioural and physiological factors, offering a holistic view of factors associated with sleep quality. The data quality of the dataset is also up to standard as it provides clean, structured numerical and categorical features suitable for ML modeling.

Artificial intelligence (AI) and machine learning (ML) have become powerful tools in healthcare. As they have the ability to train on data to be able to uncover subtle, non linear patterns in data, patterns that traditional statistical methods cannot detect.

Our approach focused on using many different types of artificial intelligence models such as Linear Regression (LR), Decision Tree Regressor (DTR), Random Forest Regressor (RFR), MLPRegressor (MLP), K Neighbors Regressor (KNR) to accurately model and predict if a person is likely to be facing reduced sleep quality.

In this study the feature set was intentionally limited to only three variables, physical activity level, sleep duration, and age, all of these are already provided as numerical values. Thus in our case no categorical encoding was required as gender, occupation, BMI category, and blood pressure were not included in the model input and therefore did not require any specific preprocessing. Furthermore, as all selected features were already numeric, no additional cleaning, scaling, or transformation was applied prior to model training.

Linear Model:

A linear model tries to find a straight line relationship between inputs and outputs. It assumes that changes in the input lead to proportional changes in the output. It is used as a starting point in data analysis. The role of the Linear Regression model is to capture basic linear trends.

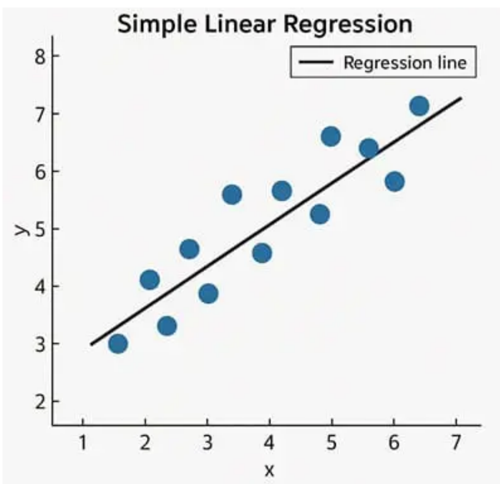

Figure 1: Simple linear regression illustration (Source: Six Sigma DSI, n.d.)

Decision Tree:

A decision tree splits data into branches based on various conditions, like asking a series of yes/no questions. It's easy to visualize and understand how the model makes decisions. However, it can become complex if not carefully controlled. The role of the Decision Tree Regressor model is to model sharp thresholds and rule based decisions. By changing the max depth, which essentially changes the amount of conditions our model has or the amount of "branches" our tree has, we managed to get different values provided by our model. As seen in the data table above.

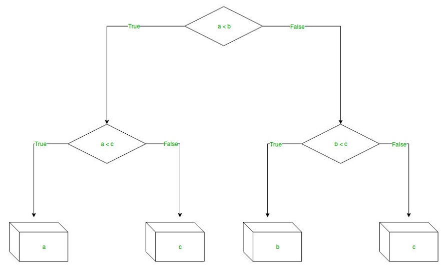

Figure 2: Decision tree regression illustration (Source: Chugh, n.d.)

Random Forest:

Uses many decision trees to make more precise predictions. Uses the outcome of each decision tree then finds an average of the decisions to get a more accurate answer. The role of the Random Forest Regressor model is to enhance stability and accuracy by aggregating many trees. In this model we also looked at the parameter "n_estimators" which is a parameter that is responsible for the amount of decision trees in our forest. We looked at a variety of values from 10-100 however we noticed that the default value of 100 gave us better, more accurate predictions, thus we decided to keep it at that for the rest of our research.

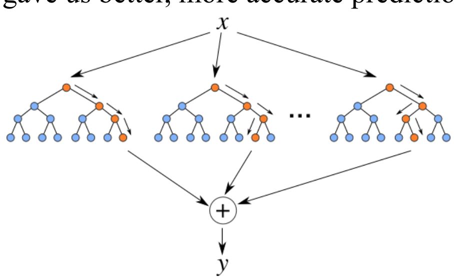

Figure 3: Random forest regression ensemble illustration (Source: gitconnected, n.d.)

MLP (Multilayer Perceptron - Neural network):

Neural networks are inspired by how the human brain works, by using layers of connected nodes called neurons. They can learn complex patterns from large amounts of data. Although proven to be very powerful they often need a lot of data and computing power to work well. An MLP is a type of neural network with multiple layers of neurons between the input and output. Each layer helps the model learn more abstract patterns in the data. It's great for tasks like classification and regression when relationships are not simple. In our case an MLP is a simple neural network. The MLP served as the deep learning model capable of learning nonlinear relationships for comparison. We altered the amount of max iterations our MLP model had during training, essentially how many times did data pass through the model, this was done to achieve a variety of comparative results.

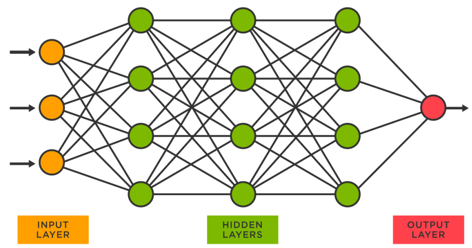

Figure 4: Neural network regression using MLPRegressor (Source: VitalFlux, n.d.)

K-Nearest Neighbours (KNN):

KNN predicts an outcome based on how similar it is to nearby examples. It looks at the "k" closest data points and uses them to make a prediction. It's simple and intuitive but can be slow when working with large datasets as each datapoint has to have its proximity calculated. The role of the KNN Regressor model is to leverage similarity among individual data points to generate predictions. We altered the value of "k", the value that is responsible for how many neighbouring points are allowed to influence a prediction, this was done to achieve a variety of comparative results.

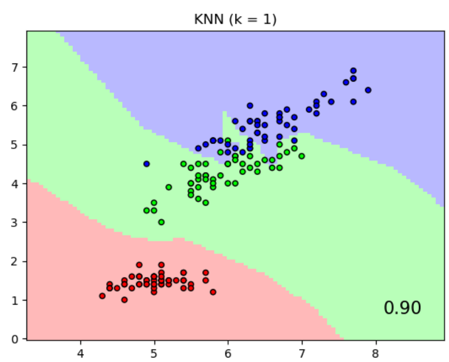

Figure 5: K-Nearest Neighbors regression illustration (Source: scikit-learn, n.d.)

To achieve more reliable and consistent predictions than any individual model could provide, we developed a weighted ensemble system based on the Multiplicative Weight Update (MWU) framework. In our approach we assigned all models to have equal weights initially, and their influence was adjusted based on performance. After each model was evaluated, those with a lower prediction error were assigned higher weights while those with higher error got their weights reduced. The final prediction is then generated as a weighted average of all the model outputs, this ensured that contribution was proportional to accuracy.

This method has several important advantages to offer. It prevented weaker models from disproportionately affecting the final results and naturally amplified the weights of the stronger more accurate models, allocating more significance to them. As the weights can adjust dynamically in response to performance, the system remains adaptive and responsive to changing conditions. Overall, this ensemble strategy increased robustness, stability, and reliability by incorporating the strengths of multiple modeling approaches, thereby reducing the risk of relying on a single model making our system more reliable.

During our model development, we divided the dataset into training and testing subsets for each model available. To ensure reliable evaluation, We split the dataset in a 66/34 train-test ratio. This was done to prevent overfitting by ensuring the model is evaluated on data it has not seen before. Random shuffling was applied with a fixed random seed (random_state = 23) to ensure reproducibility while avoiding bias introduced by fixed data ordering. This ensured that all reported results are directly reproducible across runs.

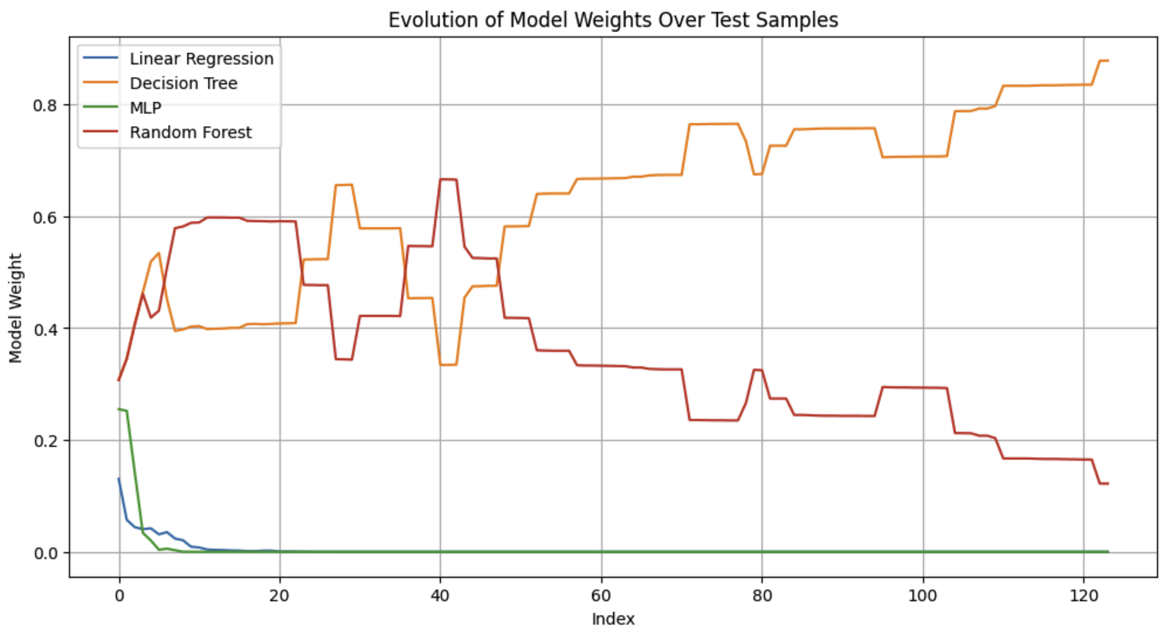

Figure X

Figure X shows the MWU framework in use, in this case a random fixed seed of 23 was assigned. Figure X illustrates how model weights evolve across test samples under the MWU framework using a fixed train-test split, highlighting how model influence changes based on predictive performance and not due to variation across runs.

To evaluate model performance, we considered the relationship between the input features (X) and the corresponding target output values (Y). For a given observation (Xi), the model produces a prediction f(Xi), which ideally should be in close proximity to the true value (Yi). The error for each prediction was therefore calculated as:

$(f(X_i)-Y_i)^2$

To calculate the overall loss, we took the average of all of the squared errors. This gives the Mean Squared Error (MSE), defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (f(X_i)\text{-}Y_i)^2$$

We also derive the Mean Absolute Error (MAE), using:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |f(X_i)\text{-}Y_i|$$

Minimising both MSE and MAE loss functions during training allows the model to adjust its parameters, reduce prediction error, and better grasp the relationship between the features provided and the target variable.

We then predicted the outcomes of both the testing subset and the training subset evaluating how well the model works if the values used are similar or different compared to what was used during the training stage. The evaluations were also based on the Mean Squared Error (MSE) and Mean Absolute Error (MAE).

Evaluation of used metrics

We used two main metrics, Mean Squared Error (MSE) and Mean Absolute Error (MAE) to assess how well our models predict given values. We used MSE as it quantifies the difference between predicted and actual values, penalizing larger errors more heavily because the error is squared. It works by finding the difference between the predicted value and the actual value and then squaring the difference. It is useful because it offers a clear measure on the prediction accuracy of models.

Mean Absolute Error (MAE) is more straightforward and works by simply finding the difference between the predicted value and the actual value by measuring the average magnitude of prediction errors without squaring them. This makes it easier to interpret the error, however, makes it less sensitive to outliers. MAE is complementary and works well with MSE as the two provide a more complete evaluation. As the prediction target is a continuous sleep quality score, model performance was evaluated using regression-based metrics rather than classification metrics.

RESULTS

A wide range of predictive results were achieved by the different regression models used in this study, each one demonstrated varying degrees of ability to estimate sleep quality from the lifestyle and biometric features provided. Evaluation was conducted using Mean Squared Error (MSE) and Mean Absolute Error

(MAE), as the objective was to predict a continuous sleep quality score rather than a performing classification. These metrics were used as they effectively measure the magnitude of the prediction error.

Model	Hyperparameters	Test MSE	Test MAE	Train MSE	Train MAE
Linear Regression	–	**0.25**	**0.39**	0.28	0.43
Decision Tree Regressor	max_depth = 6	**0.04**	**0.05**	0.02	0.03
Decision Tree Regressor	max_depth = 10	**0.03**	**0.03**	0.00	0.01
Random Forest Regressor	max_depth = 6	**0.03**	**0.07**	0.01	0.04
Random Forest Regressor	max_depth = 10	**0.03**	**0.07**	0.01	0.04
MLP Regressor	max_iter = 500	**3.42**	**1.35**	3.32	1.31
MLP Regressor	max_iter = 50	**3.42**	**1.35**	3.32	1.31
KNN Regressor	k = 1	**0.22**	**0.10**	0.00	0.00
KNN Regressor	k = 10	**0.22**	**0.10**	0.00	0.00

Lower values indicate better performance for both Mean Squared Error (MSE) and Mean Absolute Error (MAE).

Based on our data, the MLP Regressor produced the highest overall error (Test MSE = 3.42, MAE = 1.35), which suggests that the relationship between the inputted data and the sleep quality was not captured effectively by this type of model. This indicates that the neural network struggled to generalize from the available data. A likely reason is the limited dataset size or insufficient architectural tuning done to the regressor. On the contrary, tree-based models such as the Decision Tree Regressor demonstrated stronger predictive capability. The Decision Tree Regressor with a max depth of 6 achieved a Test MSE of 0.04, while increasing the depth to 10 yielded one of the best results in the entire study, with a Test MSE of 0.03 and a Test MAE of 0.03. These improvements highlight the importance and capability of using such modelling approaches that capture non linear interactions.

The Random Forest Regressor also performed well, achieving strong results across a variety of different depth values, with a Test MSE of 0.03 for both max depth = 6 and max depth = 10. The identical scores across depth settings can be explained by the ensemble nature of Random Forests. As the model averages predictions across many decision trees, by increasing the maximum depth we do not necessarily improve accuracy once the dominant patterns in the data have already been acquired. Therefore the model becomes stable beyond a certain level of depth, as additional complexity does not provide new useful information. A similar effect is observed in the deeper Decision Tree model, once the tree reaches sufficient depth to closely fit the training data (Train MSE = 0.00 at max depth = 10), further increases in capacity do not meaningfully change prediction performance.

The Linear Regression model showed weaker generalization performance compared to the tree-based approaches (Test MSE = 0.25, MAE = 0.39), indicating that the relationship between the input features and sleep quality is not well captured by a purely linear function. The results suggest the presence of non linear interactions within the data that linear models are unable to represent effectively.

Finally, the K-Nearest Neighbors models demonstrated a strong dependence on the choice of k. With k = 1, the model achieved perfect training performance (Train MSE = 0.00) but noticeably weaker generalization (Test MSE = 0.22), the results indicate overfitting. Increasing the number of neighbors to k = 10 did not improve test performance (Test MSE = 0.22), suggesting that the distance-based approach is not well suited for this dataset, likely due to the feature space not being sufficiently smooth.

Overall, the best performing models in this study were the Decision Tree Regressor (max depth = 10) and both Random Forest models, which produced the lowest and most stable error values according to MSE and MAE metrics. Their success indicates the presence of meaningful non linear relationships within the dataset and demonstrates that structured tree-based approaches are well suited for predicting sleep quality. Linear and neural models underperformed, suggesting that they were either too simple (Linear Regression) or insufficiently optimized (MLP) to capture the patterns present in our data.

DISCUSSION

This study demonstrates that machine learning techniques, when combined with routinely collected lifestyle and demographic metrics, can effectively estimate an individual's likelihood for reduced sleep quality which can occur as a result of insomnia and sleep apnea. By evaluating multiple regression-based models and implementing an adaptive weighted ensemble inspired by the Multiplicative Weight Update framework, we showed that prediction accuracy can be improved by allowing stronger models to contribute more heavily to the final output. Our tree models showed exceptional promise by capturing non-linear relationships unlike other models. These findings highlight the potential of AI as a tool to support medical staff in early screening and decision making within sleep medicine.

However, several important questions remain. The generalizability of these models to be used in larger, clinically verified datasets is still uncertain, and further research is required to identify which features consistently provide accurate prediction performance across various diverse populations. Future work should also explore using measurements collected over time from wearable technologies, enhancing model interpretability, and evaluating real world applicability in clinical environments. As well as addressing ethical and privacy considerations surrounding sensitive health data will also be essential for responsible deployment. Overall, this study provides a promising foundation for AI assisted sleep quality assessment, while highlighting the need for continued refinement and validation.

REFERENCES

Alqurashi, Y.D., Alzahrani, A., Alshamrani, N. and Alfarraj, O. (2021) 'Machine learning approaches for sleep apnea classification using lifestyle and biometric data', *Sensors*, 21(23), p. 7973. Available at: https://doi.org/10.3390/s21237973

Cappuccio, F.P., Taggart, F.M., Kandala, N., Currie, A., Peile, E., Stranges, S. and Miller, M.A. (2010) 'Sleep duration and all-cause mortality: A systematic review and meta-analysis', *Sleep*, 33(5), pp. 585–592. Available at: https://doi.org/10.1093/sleep/33.5.585

Chugh, K. (n.d.) 'Decision tree regression in machine learning'. Available at: https://kashishchugh.medium.com/decision-tree-regression-in-machine-learning-4f117158cdcc (Accessed: 20 December 2025).

Colten, H.R. and Altevogt, B.M. (2006) 'Functional and economic impact of sleep loss and sleep disorders'. In: Institute of Medicine (US) Committee on Sleep Medicine and Research, *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*. Washington, DC: National Academies Press. Available at: https://nap.nationalacademies.org/read/11617/chapter/4

Dietterich, T.G. (2000) 'Ensemble methods in machine learning'. In: *International Workshop on Multiple Classifier Systems*. Berlin, Heidelberg: Springer, pp. 1–15. Available at: https://doi.org/10.1007/3-540-45014-9_1

Faust, O., Hagiwara, Y., Hong, T.J., Lih, O.S. and Acharya, U.R. (2019) 'Deep learning for healthcare applications based on physiological signals: A review', *Computer Methods and Programs in Biomedicine*, 161, pp. 1–13. Available at: https://doi.org/10.1016/j.cmpb.2018.04.005

gitconnected (n.d.) 'Random forest regression'. Available at: https://levelup.gitconnected.com/random-forest-regression-209c0f354c84 (Accessed: 2 January 2026).

Grandner, M.A. (2017) 'Sleep, health, and society', *Sleep Medicine Clinics*, 12(1), pp. 1–22. Available at: https://doi.org/10.1016/j.jsmc.2016.10.012

Hublin, C., Partinen, M., Koskenvuo, M. and Kaprio, J. (2011) 'Sleep and mortality: A population-based 22-year follow-up study', *Sleep*, 34(10), pp. 1241–1248. Available at: https://doi.org/10.5665/sleep.1270

Institute of Medicine (US) Committee on Sleep Medicine and Research (2006) *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*. Washington, DC: National Academies Press. Available at:

https://nap.nationalacademies.org/catalog/11617/sleep-disorders-and-sleep-deprivation-an-unmet-public-health-problem

Kearns, M. and Valiant, L. (1994) 'Cryptographic limitations on learning Boolean formulae and finite automata', *Journal of the ACM*, 41(1), pp. 67–95. Available at: https://doi.org/10.1145/174644.174647

Khalighi, S., Sousa, T., Santos, J. and Nunes, U. (2014) 'Automatic sleep stage scoring using a single-channel EEG and random forest classifier', *Artificial Intelligence in Medicine*, 62(1), pp. 13–21. Available at: https://doi.org/10.1016/j.artmed.2014.08.002

Kim, J.H., Kim, K.R., Kim, M., Park, J.H. and Kim, H. (2017) 'Lifestyle factors and sleep health in adults: A population study', *Preventive Medicine*, 100, pp. 243–247. Available at: https://doi.org/10.1016/j.ypmed.2017.04.023

Pandey, A., Xiang, J. and Mei, Z. (2020) 'Sleep duration and health outcomes: A systematic review and meta-analysis', *Journal of Clinical Sleep Medicine*. Available at: https://doi.org/10.5664/jcsm.8612

Radha, M., Fonseca, P., Moreau, A., Ross, M., Cerny, A., Anderer, P. and Long, X. (2019) 'Sleep stage classification using machine learning: A systematic review', *IEEE Reviews in Biomedical Engineering*, 12, pp. 199–214. Available at: https://doi.org/10.1109/RBME.2018.2886670

scikit-learn (n.d.) 'Neighbors'. Available at: https://scikit-learn.org/stable/modules/neighbors.html (Accessed: 7 January 2026).

Six Sigma DSI (n.d.) 'Simple linear regression'. Available at: https://sixsigmadsi.com/glossary/simple-linear-regression/ (Accessed: 5 January 2026).

VitalFlux (n.d.) 'Sklearn neural network regression example – MLPRegressor'. Available at: https://vitalflux.com/sklearn-neural-network-regression-example-mlpregressor/ (Accessed: 2 January 2026).

Zhang, Q., Zhou, Y., Zhang, B. and Lin, G. (2018) 'A deep learning model for detecting sleep apnea from wearable sensor data'. In: *2018 IEEE EMBC*, pp. 2578–2581. Available at: https://doi.org/10.1109/EMBC.2018.8512862

Zhou, Z.-H. (2012) *Ensemble Methods: Foundations and Algorithms*. Boca Raton, FL: Chapman & Hall/CRC. Available at: https://www.routledge.com/Ensemble-Methods/