

Article

# Context-Aware Behavioral Tips to Improve Sleep Quality via Machine Learning and Large Language Models

Erica Corda , Silvia M. Massa \* and Daniele Riboni \*

Department of Mathematics and Computer Science, University of Cagliari, Via Ospedale 72, 09124 Cagliari, Italy; erica.corda@gmail.com

\* Correspondence: silviam.massa@unica.it (S.M.M.); riboni@unica.it (D.R.)

**Abstract:** As several studies demonstrate, good sleep quality is essential for individuals' well-being, as a lack of restoring sleep may disrupt different physical, mental, and social dimensions of health. For this reason, there is increasing interest in tools for the monitoring of sleep based on personal sensors. However, there are currently few context-aware methods to help individuals to improve their sleep quality through behavior change tips. In order to tackle this challenge, in this paper, we propose a system that couples machine learning algorithms and large language models to forecast the next night's sleep quality, and to provide context-aware behavior change tips to improve sleep. In order to encourage adherence and to increase trust, our system includes the use of large language models to describe the conditions that the machine learning algorithm finds harmful to sleep health, and to explain why the behavior change tips are generated as a consequence. We develop a prototype of our system, including a smartphone application, and perform experiments with a set of users. Results show that our system's forecast is correlated to the actual sleep quality. Moreover, a preliminary user study suggests that the use of large language models in our system is useful in increasing trust and engagement.

**Keywords:** pervasive healthcare; behavior change; sleep quality forecasting; machine learning; large language models



**Citation:** Corda, E.; Massa, S.M.; Riboni, D. Context-Aware Behavioral Tips to Improve Sleep Quality via Machine Learning and Large Language Models. *Future Internet* **2024**, *16*, 46. <https://doi.org/10.3390/fi16020046>

Academic Editor: Guan Gui

Received: 14 December 2023

Revised: 18 January 2024

Accepted: 29 January 2024

Published: 30 January 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Several studies show that sleep is a fundamental biological requirement for human life, and good sleep quality is crucial in maintaining physical, mental, and emotional health [1–3]. Indeed, adequate and restorative sleep contributes to enhanced cognitive function, including improved memory consolidation, problem-solving abilities, and overall mental alertness [4]. Moreover, it supports emotional resilience by regulating mood and reducing stress, anxiety, and irritability [5]. Good-quality sleep is also necessary in maintaining physical health, promoting immune system function, and aiding in the repair and recovery of tissues and organs [6]. Sleep plays a vital role in hormonal regulation, appetite, and metabolism [7]. Chronic sleep deprivation is associated with an increased risk of various health issues, including cardiovascular, respiratory, gastrointestinal, endocrine, and reproductive diseases [8]. In addition to physical and mental benefits, good-quality sleep has a positive influence on the quality and quantity of social relationships, as it contributes to emotional stability and empathy [9,10].

Since good-quality sleep is essential for physical, mental, and emotional well-being, there is increasing interest in innovative personal healthcare tools for the monitoring of sleep [11]. Wearable fitness trackers and smartwatches use accelerometers and heart rate monitors to recognize sleep patterns, including duration, efficiency, and sleep stages, such as deep and rapid eye movement (REM) sleep. Dedicated sleep tracking devices, possibly placed under the mattress and including microphones, provide additional data, including snoring and interruptions. These tools provide useful insights to inspect various aspects of

sleep. However, generally, they fall short in helping the user to improve their sleep quality, since they do not explain the reasons for the detected sleep quality and do not provide context-aware suggestions for improvement.

According to the scientific literature, several behavioral factors have an impact on sleep [12], including the intensity and timing of physical activities, sleep schedules, nighttime environment, exposure to noise and electronic screens, and social relationships and communication. Hence, a fine-grained analysis of the individual's behavior may indicate which specific factors may negatively influence sleep quality and enable the provision of context-aware behavior change tips.

In [13], the authors used the random forest algorithm to predict the occurrence of an apnea event according to polysomnography signal analysis. Next, they used a tool for model interpretability to generate predictions' explanations and create a graphic representation of the specific airflow signal and important feature points. Chan-Hen et al. [14] presented a long short-term memory (LSTM) model capable of predicting panic disorder-related events and providing suggestions for improved sleep quality to decrease panic attacks. The data analyzed were collected through the Diagnostic and Statistical Manual of Mental Disorders 5th ed., the Mini International Neuropsychiatric Interview, various clinical questionnaires, and smartwatch sensors capable of detecting information related to sleep, physical activity, and heart rate. The study inferred general suggestions to improve sleep considering parameters such as sleep duration, walking activity, and the duration of different sleep states. In [15], the authors presented an extreme gradient boosting (XGBoost) model capable of classifying the non-rapid eye movement (NREM) and REM sleep phases through the analysis of electroencephalography (EEG) signals. The authors used a tool for model interpretability to determine the most significant features and to explain the predictions. However, the use of large language models (LLMs) in the analysis of factors influencing sleep to provide suggestions is still limited. In this context, the work presented by Mira et al. [16] demonstrates the potential of LLMs in assisting healthcare providers in the diagnosis of obstructive sleep apnea (OSA) by comparing the responses of a generative pre-trained Transformer and those of experienced otolaryngologists.

In this paper, we present a system to support individuals in improving their sleep quality via the machine learning (ML)-based forecasting of future sleep quality and large language model (LLM)-based provisioning of context-aware behavior change tips through a smartphone application. In order to increase users' trust, our system is supported by an LLM-based algorithm to explain the reasons behind the predictions and tips; these explanations are provided in natural language to the user through the app interface. In fact, leveraging the language generation capabilities of LLM, our system can articulate the underlying factors and reasoning behind each behavioral tip in a clear and understandable manner, helping to bridge the gap between the complex ML models used by the system and the user's understanding and trust. The transparency provided by the cooperation of ML and LLM in our system may also enhance users' engagement, enabling them to make informed decisions and take proactive steps toward enhancing their sleep health.

Technically, our system continuously collects data from the user's smartphone about activities (physical activities and conversations), usage modalities, location, and environmental conditions (light level, noise, types of sound). The system integrates the raw data and extracts features of interest on a daily basis. The feature vectors are used by a random tree regressor to forecast the next night's sleep quality. The model of the random tree regressor is trained at the server side in order to avoid the disclosure of sensitive data to third parties. Considering the decision path of the ML model, an LLM-based algorithm is used to infer the contextual factors that lead to the prediction. If the system forecasts poor sleep quality, the smartphone application provides context-aware behavior change tips. Moreover, to increase the user's trust and to enhance their adherence to the provided recommendations, the LLM-based algorithm provides explanations for the predictions and tips.

We have developed a working prototype of our system and performed experiments regarding both sleep quality forecasting and the utility of context-aware tips together with their explanations. Experiments with real-world data show that the predictions of our sleep quality forecasting algorithms are correlated with the ground truth. Moreover, a user study indicates the potential utility of our system and the increase in trust given by the LLM-based module.

The rest of the paper is structured as follows. Section 2 examines the state of the art in pervasive healthcare systems for sleep. Section 3 presents the architecture and algorithms used in our system. Section 4 illustrates our experimental evaluation, the user study, and the achieved results. Section 5 discusses the results, limitations, and possible improvements of our work. Finally, Section 6 concludes the paper.

## 2. Related Work

Artificial intelligence (AI) applied to healthcare aims to develop computer systems capable of performing tasks that normally require human intelligence, allowing the extraction of deeply hidden information in medical big data that would otherwise be difficult to find. It was first employed in the medical field in the early 1970s as a diagnostic support tool [17]. Today, due to the increasing computing power of modern hardware architectures and the vast amount of digital data available for research, AI is increasingly used to guide clinical decisions [18], to detect or predict in advance a disease [19], and to support surgeons during surgery and patients during rehabilitation [20]. ML has been widely used to predict and diagnose mental disorders, monitor patients' moods and mental health, and improve communication between patients and mental health professionals [21].

An important means to prevent different mental disorders is to improve sleep quality. In fact, people who suffer from sleep deficiencies are more prone to depression, anxiety, stress, and other illnesses affecting the psyche. Furthermore, the study conducted by Scott et al. [22] states that improved sleep quality can have positive effects on composite mental health, depression, rumination, stress, and psychosis symptoms. For this reason, several studies have been carried out concerning the prediction of sleep quality through the use of sensors and ML algorithms. The problem of sleep quality prediction is mainly treated as a classification problem rather than a regression problem [23].

In [24], the close correlation between physical activity and sleep quality is verified. Specifically, sleep efficiency is predicted through the use of the wearable device ActiGraph Gt3X+, capable of acquiring data on physical activity during awake time, and several deep learning algorithms. Arora et al. [25] analyze data from commercial smartwatches and clinical actigraphy for sleep quality estimation and prediction through a convolutional neural network (CNN) and multilayer perceptron (MLP). In [26], the authors propose to monitor the sleep quality of caregivers of people suffering from dementia to assess their stress and support clinical decisions. The E4 wristband is used to capture heart rate variability, electrodermal activity, body movement, and skin temperature. Subsequently, several ML algorithms are tested.

Sound analysis can be beneficially used to monitor sleep quality. The iSleep system presented by Chang et al. [27] uses the smartphone's built-in microphone to detect events that may influence sleep, such as body movement, coughing, and snoring, and provide information on overall sleep efficiency on a daily basis in an unobtrusive manner. Sleep-related events are detected by means of a decision tree-based classifier. In addition to the microphone, other smartphone sensors employed to detect sleep information in the literature are accelerometers, screen on/off events, light sensors, stationary modes, batteries, and touch screen events [28]. For instance, the accelerometer can be used to measure body movement and to understand sleep phases, and the time when the screen is switched off, the touch screen is not used, or darkness is detected can be used as an indicator of the sleep duration.

A few recent works propose to generate user-friendly textual explanations of AI predictions using generative pre-trained Transformers (GPT) such as ChatGPT, the popular

large language model (LLM)-based natural language processing (NLP) system developed by Open AI [29]. In particular, Susnik [30] proposes a framework that takes advantage of ML techniques to understand whether a student is at risk of leaving university, and ChatGPT to convert candidate prescriptive feedback into a form of natural language that students can understand. In this paper, we investigate a similar approach. To the best of our knowledge, our work is the first attempt to provide a system to forecast sleep quality and provide context-aware tips to improve sleep based on ML and LLM.

### 3. Methodology

Figure 1 illustrates our system architecture.

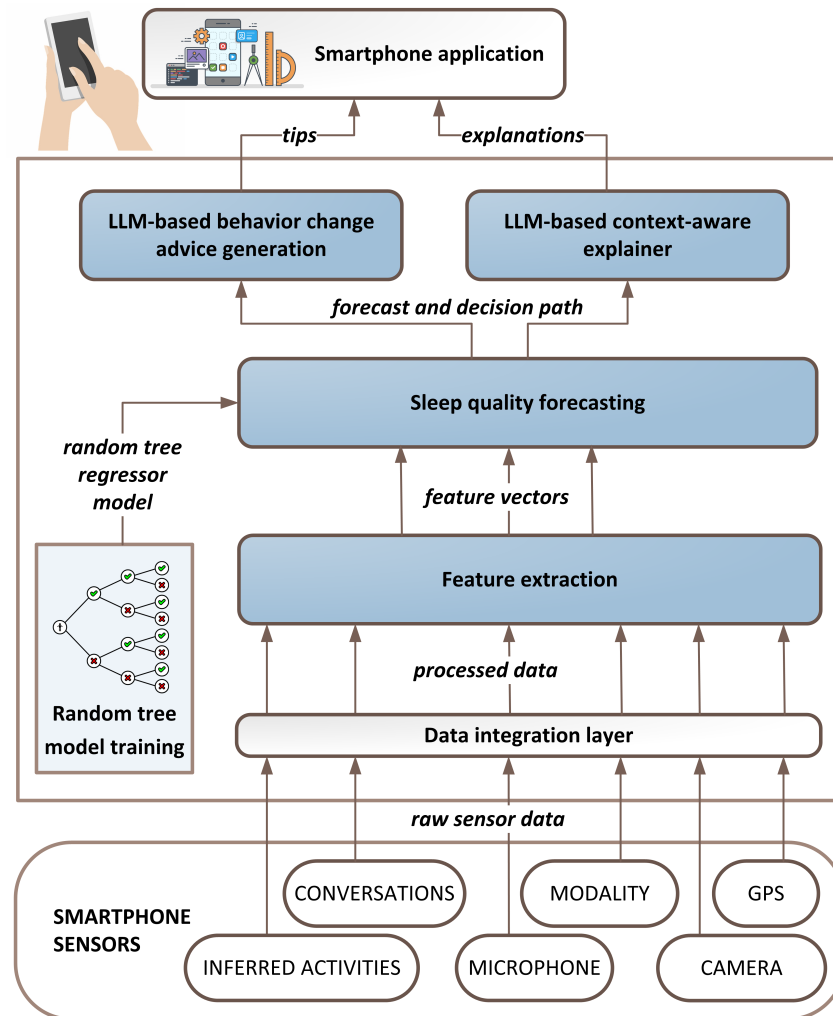


Figure 1. System architecture.

A large set of raw data is continuously acquired from smartphone sensors, including current activity, location, movement data, sleep patterns, and other contextual information. A data integration layer is responsible for preprocessing the raw data to represent them in a common format, dealing with noisy and missing values. Then, the most relevant features are extracted from the preprocessed data by the feature extraction module, which creates the feature vectors used by the sleep quality forecasting module. The latter relies on a random tree regressor model to predict the user’s sleep quality for the next night. Finally, an LLM architecture is exploited to generate recommendations to improve sleep quality based on the obtained predictions, and to provide detailed explanations of the reasons behind the predictions. Recommendations and explanations are provided to the individual through a user-friendly smartphone app.

### 3.1. Raw Data Acquisition

Several heterogeneous data are continuously collected via the microphone, accelerometer, light sensor, GPS, and Bluetooth of the smartphone to acquire a comprehensive view of the user's context, including physical activity, conversation, sleep patterns, location, and co-location. In particular, the system acquires the following raw context data:

- The type of performed activity;
- The types of sounds recorded by the smartphone's microphone.;
- The smartphone modality, i.e., locked or unlocked;
- Light information, i.e., whether the smartphone is in a dark or in an illuminated environment;
- Conversation activities, e.g., detected through phone calls or messaging sessions;
- The distance covered by the user.

Each of the above pieces of information is annotated with the start and end timestamp of occurrence.

For the sake of experimentation, and to train the random tree regressor model, ground truth data about the number of slept hours and perceived sleep quality are acquired with the help of ecological momentary assessment (EMA) questionnaires, i.e., short surveys periodically administered to a selected set of volunteers.

### 3.2. Data Integration

Data cleaning and normalization are applied to the acquired raw data to make them consistent and homogeneous.

- *Data cleaning* includes the detection of outliers, missing values, and noise reduction. This step is particularly important when dealing with raw sensor data, since they are typically characterized by uncertainty and the presence of noise. In this context, outlier detection has the goal of recognizing noisy sensor data and removing them. Indeed, these data may confuse the ML algorithm and damage the recognition performance. Outlier detection is performed using a statistical method [31]. Missing data are replaced by means of interpolation. Noise reduction is achieved by applying low-pass filters to inertial sensor data to eliminate high-frequency noise.
- Data provided by different smartphone sensors may use different reference scales. *Data normalization* is applied to ensure that all the acquired data have the same scale. For this reason, data acquired from sound and light sensors are normalized in a [0, 1] range, while the reported distance is represented in a metric scale.

### 3.3. Feature Extraction

Extracting representative features from sensor data is essential to improve the prediction capabilities of the ML algorithm. In this work, we use a sliding-window-based technique to extract statistical features from preprocessed data provided by the data integration module. We partition each day into 12 time slots of two hours each. For each time slot, we compute the following features.

- The time duration spent performing each activity type, i.e., *stationary*, *walking*, *running*, or *unknown*.
- The time duration of each detected sound type, i.e., *silence*, *voice*, *noise*, or *unknown*.
- The time duration during which the smartphone was *locked*.
- The time duration during which the smartphone was in a *dark* environment.
- The time duration during which the user was engaged in a *conversation*.

In the above features, all the time durations are measured in minutes. We also collect the number of hours slept during the previous night, as well as features regarding the distance traveled during the morning (7 a.m.–1 p.m.), afternoon (1 p.m.–8 p.m.), and night (8 p.m.–7 a.m.), respectively. In these features, the traveled distance is measured in kilometers.



The above features represent the context information extracted during the day from the smartphone user's activities. Hence, a user's day is characterized by 135 contextual features.

Finally, the ground truth regarding *sleep quality* is represented by a real number in  $[1, 4]$ , where 1 represents very good sleep quality, and 4 represents very bad sleep quality. As explained before, the ground truth is collected from a selected set of users by means of questionnaires.

### 3.4. Sleep Quality Forecasting

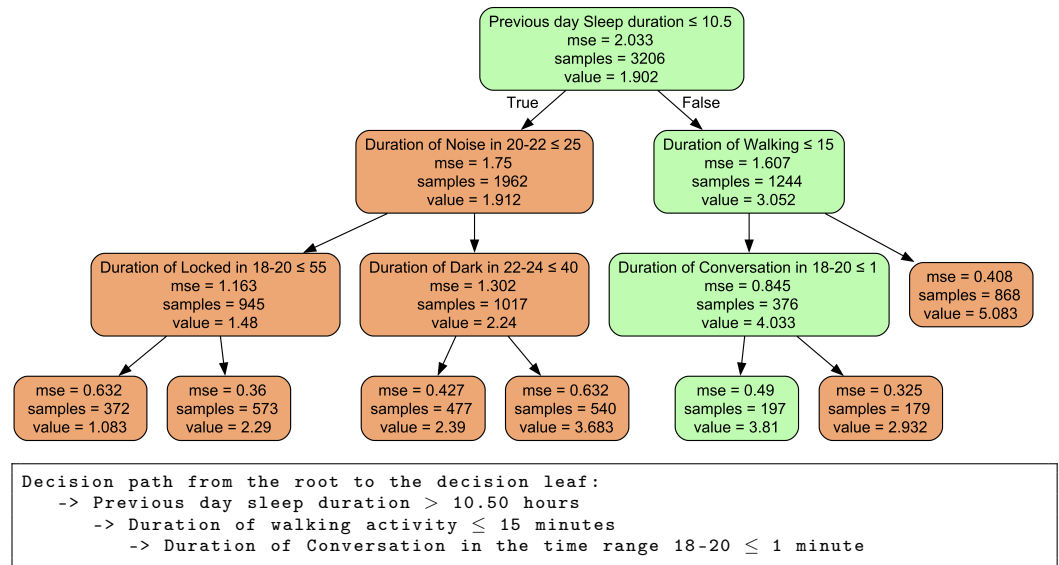
This module is executed on a daily basis. It is responsible for processing the feature vectors received during the day to forecast the sleep quality that the user will experience in the next night. The forecast is obtained by executing a random tree regressor model. As explained in Section 3.1, the model is trained thanks to a set of labeled data obtained from volunteers.

The model of the random tree regressor corresponds to a decision tree, where the feature space is progressively split into subsets based on certain features. Each internal node of the tree, including the root, represents a decision based on the value of a particular feature, and each leaf node represents a possible forecast value. In order to provide context-aware tips to improve sleep, we are interested in understanding the reasons for which the ML algorithm predicts a given value. To this aim, we need to inspect the *decision path* of the decision tree, i.e., the sequence of feature conditions evaluated along the branches from the root to the specific leaf node that led to the actual forecast. We derive the decision path by leveraging the machine learning library employed in the experimental evaluation of our work, namely scikit-learn <https://scikit-learn.org/> (accessed on 29 January 2024).

Figure 2 shows a sample of the random tree model. As exemplified in the figure, the green nodes represent the decision path for a certain feature vector. In our case, the decision path provides insights into how the algorithm arrived at a specific sleep quality forecast based on a subset of the user's feature vector values, which encode information about the user's past behavior and contextual conditions. The listing reported in the box below the tree in Figure 2 shows the notation that we use to represent the decision path illustrated by the green nodes. Note that, in our notation, we represent the true conditions that determined the path from the root to the decision leaf. For instance, since the condition of the root node ('previous day sleep duration  $\leq 10.5$ ') was evaluated as false in the example, in our notation, we report the corresponding true condition, 'previous day sleep duration  $> 10.5$ '. We use this notation to communicate the decision path to the LLM-based modules. Indeed, as explained in Section 3.5, we do not craft the prompt given to the LLM modules using the feature vector; instead, we use the decision path, since it precisely describes the reasons and conditions that determine the forecast for a given individual. Otherwise, the whole feature vector would contain several pieces of irrelevant information, i.e., features that did not contribute to the actual forecast.

The random tree regressor predicts a real number  $r$  between negative infinity and positive infinity, while we are interested in providing the user with a qualitative forecast of his/her future sleep quality. For the sake of this work, we consider four levels of sleep quality, ranging from very good sleep quality (level 1) to very bad sleep quality (level 4). Hence, we approximate the predicted number  $r$  to the nearest integer in  $[1, 4]$  to compute the corresponding qualitative sleep quality level. In particular,

- values of  $r$  in  $(-\infty, 1.5)$  are approximated to 1, which represents very good sleep quality;
- values of  $r$  in  $[1.5, 2.5)$  are approximated to 2, which represents fairly good sleep quality;
- values of  $r$  in  $[2.5, 3.5)$  are approximated to 3, which represents fairly bad sleep quality;
- values of  $r$  in  $[3.5, \infty)$  are approximated to 4, which represents very bad sleep quality.



**Figure 2.** A sample of the random tree model. Considering the feature vector values, the decision tree follows a path from the root node to a leaf node that represents the model’s forecast. The *decision path* is the sequence of feature conditions evaluated along the branches from the root to the specific leaf node, determining the prediction for a particular feature vector. Consider, as an example, a feature vector including the following values: ‘Previous Day Sleep Duration’ = 11 h; ‘Duration of Noise in 20–22’ = 10 min; ‘Duration of Walking’ = 5 min; ‘Duration of Locked in 18–20’ = 40 min; ‘Duration of Dark in 22–24’ = 85 min; ‘Duration of Conversation in 18–20’ = 0 min. In this case, the decision path is composed of the nodes highlighted in green, and the random tree sleep quality forecast value is 3.81. In each node, *samples* represents the number of instances of the training set that match the conditions of the node, the *value* represents the most probable value, and *mse* represents the corresponding mean squared error. The listing in the box below the tree shows the notation that we use to represent the decision path. We use this notation to communicate the decision path to the LLM-based modules.

The sleep quality forecasting module communicates its approximated forecast, together with the corresponding decision path, to the module for LLM-based behavior change tip generation and to the LLM-based context-aware explainer module.

### 3.5. LLM-Based Behavior Change Tip Generation

The objective of this module is to produce context-aware natural language tips to help users to improve their sleep quality through healthier behaviors. Indeed, several behaviors, such as inconsistent sleep schedules, inadequate physical activity, and the use of electronic devices, may disrupt the natural sleep–wake cycle and hinder the ability to achieve restful and deep sleep. However, the provision of generic tips may have little effect on behavior change, since they cannot capture the exact situation of the individual and the reasons for his/her poor sleep quality.

In order to provide context-aware tips, the behavior change tip generation module considers the decision path that leads to the forecast of poor sleep quality. The module dynamically creates a prompt to ask a generative pre-trained Transformer (GPT) model to provide suggestions to improve the user’s sleep quality considering the decision path conditions. A sample of the prompt is shown in Figure 3.

```

We used a decision tree to forecast the sleep quality of a person based on his/her
past behavior. The resulting decision path is the following:

Decision path from the root to the decision leaf:
-> Previous day sleep duration ≤ 6.50 hours,
  -> Duration of noise in the time range 20-22 > 29 minutes,
    -> Time spent in a dark room in the time range 22-24 ≤ 38 minutes
      -> Rating of sleep quality of the previous day > 3

The resulting sleep forecast is: Fairly bad sleep quality

Considering the decision path and the decision tree prediction, please provide
personalized natural language suggestions for the person to improve sleep quality.

```

**Figure 3.** A sample of the prompt issued to the generative pre-trained Transformer.

For the sake of this study, we used GPT 3.5 to generate context-aware tips according to the prompt request. Figure 4 shows the response generated to the sample prompt shown in Figure 3.

```

Based on the decision tree analysis and the identified factors influencing sleep
quality, here are personalized suggestions to improve your sleep:

Increase Sleep Duration:
  Ensure you get at least 6.50 hours of sleep each night. Consider adjusting your
  bedtime or creating a relaxing pre-sleep routine to help you wind down.

Manage Noise Exposure:
  Limit noise in the evening, especially between 20:00 and 22:00. This might involve
  using earplugs, closing windows, or finding a quieter environment during this
  time.

Enhance Sleep Environment Darkness:
  Spend more time in a dark room between 22:00 and 24:00. Consider using blackout
  curtains or an eye mask to create a conducive sleep environment.

Improve Sleep Quality Ratings:
  Work on activities or habits that contribute to a better sleep quality rating.
  This could include establishing a relaxing bedtime routine, practicing
  mindfulness or relaxation exercises, and addressing any stressors that may
  impact your sleep.

```

**Figure 4.** The generative pre-trained Transformer response to the prompt shown in Figure 3.

The provided tips strongly depend on the person's behavior, which determines his/her feature vectors and, consequently, the decision path. Figures 5 and 6 show the prompt and the corresponding response, respectively, in a different use case, which corresponds to the scenario illustrated in Figure 2.

```

We used a decision tree to forecast the sleep quality of a person based on his/her
past behavior. The resulting decision path is the following:

Decision path from the root to the decision leaf:
-> Previous day sleep duration > 10.50 hours
  -> Duration of walking activity ≤ 15 minutes
    -> Duration of Conversation in the time range 18-20 ≤ 1 minute

The resulting sleep forecast is: Very bad sleep quality

Considering the decision path and the decision tree prediction, please provide
personalized natural language suggestions for the person to improve sleep quality.

```

**Figure 5.** A different sample of the prompt issued to the generative pre-trained Transformer.



```

Based on the decision tree analysis and the identified factors influencing sleep
quality, here are personalized suggestions to improve your sleep:

Optimize Sleep Duration:
    While it's important to get enough sleep, sleeping more than 10.50 hours might be
    excessive and could contribute to feelings of grogginess. Aim for a more
    balanced sleep duration, typically between 7 to 9 hours per night.

Increase Physical Activity:
    Since the decision tree indicates a low duration of walking activity, consider
    incorporating more physical activity into your daily routine. Regular exercise
    has been shown to improve sleep quality. Aim for at least 30 minutes of
    moderate-intensity exercise most days of the week.

Extend Conversation Time in the Evening:
    Increase the duration of your social interactions during the early evening hours
    (18:00 to 20:00). Meaningful conversations can help relax your mind and
    promote a sense of connection, which may positively impact your sleep.

```

**Figure 6.** The generative pre-trained Transformer response to the prompt shown in Figure 5.

Considering the above examples, it is evident that the provided suggestions are contextualized considering the past behavior of the individual. In the first example (Figures 3 and 4), the individual slept for less than 6.50 h, was exposed to noise for more than 29 min in the time range 20–22, spent less than 38 min in a dark place from 22 to 24, and experienced poor sleep quality during the previous day. Hence, the module suggests sleeping longer, limiting noise exposure in the evening, enhancing the darkness of the sleep environment at night, and improving sleep through healthy routines. In the second example (Figures 5 and 6), a different individual slept for more than 10.50 h, walked for less than 15 min, and had no conversation from 18 to 20. In this case, the module provides different suggestions, i.e., sleeping for less time, increasing physical activity, and engaging in meaningful social interactions in the evening. We believe that such context-aware suggestions are likely to be more effective than generic ones that do not take into account the user's actual behavior.

### 3.6. LLM-Based Context-Aware Explainer

For applications related to health and well-being, it is fundamental to provide the user with explanations about the suggestions received by the ML algorithm. Hence, our system includes a dedicated module that dynamically produces a prompt asking the GPT to explain the reasons underlying the generated behavior change tips.

Of course, the explanations of the behavioral change tips must be linked closely to the individual's specific past behavior. This tight coupling leverages the individual's historical actions, making the explanations directly relevant to his/her habits. In contrast, generic explanations would lack this contextual connection, diminishing their impact and applicability to the user's unique lifestyle and routines. For this reason, we considered different prompts to enable the GPT to produce explanations that are both easily understandable and tightly coupled to the individual's unique context.

As a result of our tests, our choice was to ask the GPT to produce explanations considering the decision path of the decision tree. Indeed, the decision path includes the specific conditions that determined the sleep quality forecast (e.g., too long or too short sleep time, social interactions, exposure to noise, etc.). The chosen prompt, together with the generated response, is shown in Figure 7 considering the use case of Figures 5 and 6. Note that, since the prompt of the "LLM-based context-aware explainer" module is issued in the same context as the "LLM-based behavior change tip generation" one, we do not need to explicitly report the decision path in the former prompt.

```

Prompt: Explain the reasons why you provided those suggestions considering the
        decision path of the decision tree.

Response:
Certainly! Let's break down the personalized suggestions based on the decision path of
        the decision tree:

Optimize Sleep Duration:
    Decision Path: Previous day sleep duration > 10.50 hours.
    Reasoning: Sleeping more than 10.50 hours may lead to feelings of grogginess and
        can negatively impact sleep quality. The suggestion is to aim for a more
        balanced sleep duration, typically between 7 to 9 hours, which is considered
        optimal for most adults.

Increase Physical Activity:
    Decision Path: Duration of walking activity < 15 minutes.
    Reasoning: The decision tree suggests that the individual has a low duration of
        walking activity. Regular physical activity is associated with improved sleep
        quality. Therefore, increasing the duration of daily walking can contribute to
        better overall sleep.

Extend Conversation Time in the Evening:
    Decision Path: Duration of Conversation in the time range 18-20 < 1 minute.
    Reasoning: The limited duration of evening social interactions might contribute to
        a lack of relaxation and social connection before bedtime. Increasing the
        time spent in meaningful conversations during the early evening hours can
        positively impact mood and potentially improve sleep quality.
    
```

Figure 7. LLM-based explanations of the use case illustrated in Figures 5 and 6.

### 3.7. Smartphone Application

On a daily basis, a custom smartphone application receives behavior change tips and corresponding explanations, parses them, and displays them in a user-friendly manner to the user. Samples of the app tips and explanations are shown in Figures 8 and 9, respectively.

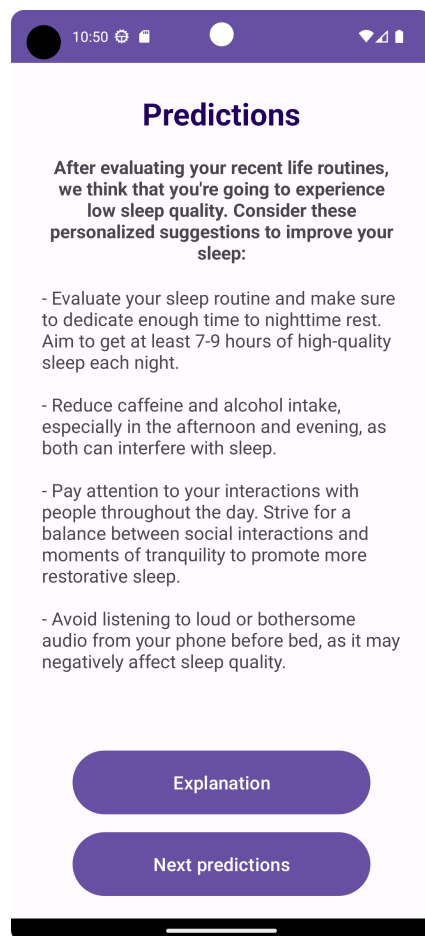


Figure 8. Smartphone app prediction.

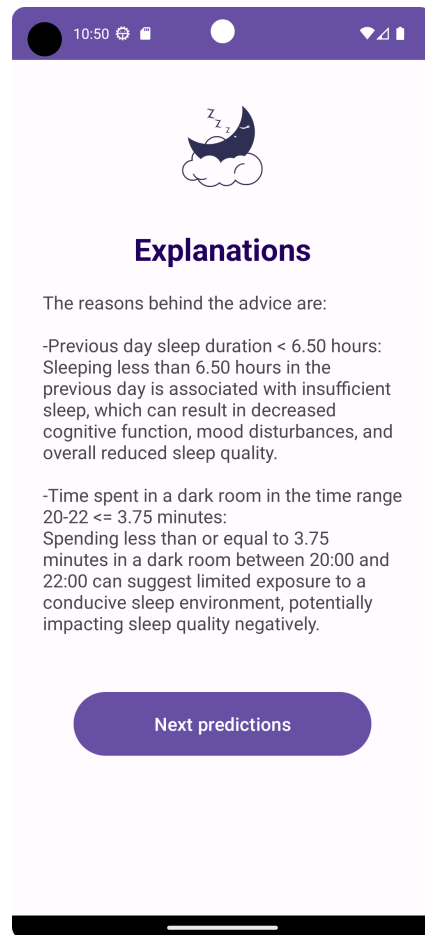


Figure 9. Smartphone app explanation.

#### 4. Experimental Evaluation

We conducted an extensive experimental evaluation of our system to assess both its predictive capabilities and the user experience.

##### 4.1. Dataset

In order to evaluate the prediction performance of our system, we used the StudentLife [32] dataset. The StudentLife dataset was collected by a research team from Dartmouth College within the activities of the StudentLife project, a research initiative aimed at understanding the daily activities, behaviors, and well-being of college students. This dataset incorporates information obtained from smartphone sensors and wearable devices, encompassing data on physical activity, location, sleep patterns, modalities, and communication. Additionally, it includes survey responses addressing different health parameters, including sleep quality.

Approval of all ethical and experimental procedures and protocols of the StudentLife dataset was granted by the Institutional Review Board at Dartmouth College [32]. The data collection lasted 10 weeks and was carried out with 48 individuals (30 undergraduates and 18 graduate students; 10 females and 38 males) using Android smartphones. Smartphone data were automatically gathered without user input and sent to the StudentLife project cloud infrastructure. The app automatically inferred participants' physical activity, sleep duration, and sociability, including the number and duration of conversations. The sleep recognition algorithm achieved 95% accuracy in correctly recognizing bedtime, with an error  $\pm 25$  min. The app also gathered data from different sensors, including inertial sensors, location/co-location/proximity, and audio and light sensor readings. Moreover, participants were prompted to respond to various Ecological Momentary Assessment

(EMA) questions related to different health dimensions, including perceived sleep quality. The EMA reports, supplied by a medical doctor and psychologists from the research team, were administered multiple times a day, with an average of 8 EMAs per day. Different incentives were provided to participants to promote compliance and ensure data quality. The acquired data were fully anonymized to enforce data privacy and were stored on secure servers. Overall, the total size of the dataset is 52.6 GB. The dataset is represented in tabular format, where each row corresponds to a sensor event. The dataset encompasses a comprehensive array of information, comprising approximately 23 million accelerometer events, 99 million audio-related events, 79 thousand conversation events, 7 thousand events associated with light levels, and 9 thousand events related to phone usage. The dataset is publicly available at <https://studentlife.cs.dartmouth.edu/> (accessed on 29 January 2024).

#### 4.2. Sleep Quality Forecasting

To evaluate the effectiveness of our system in forecasting the individual's sleep quality in the following night based on past smartphone data, we used a leave-one-person-out cross-validation approach, i.e., the data of one person were used for testing, while the data of the remaining people were used for training, and testing was iteratively executed with all subjects. The code to run the experiments was implemented in Python, using its machine learning libraries.

As explained before, we treat sleep quality forecasting as a regression problem, where the target value (i.e., the quality of sleep of the next night) is represented by a real number ranging from 1 (very bad sleep quality) to 4 (very good sleep quality). We performed the experiments with two well-known algorithms: the random forest regressor and the random tree regressor. We chose these algorithms because they achieved good performance in similar tasks even when the dimension of the available training set was relatively small [33].

We measured the effectiveness of prediction using the Pearson correlation coefficient  $r$  [34]. The results, shown in Table 1, indicate that the random tree regressor and the random forest regressor have similar performance on this task. Indeed, the former achieves an  $r$  value of 0.33, while the latter achieves  $r = 0.22$ . However, the logic behind the predictions is more understandable when using the random tree because it consists of a series of binary decisions based on input features. Instead, the logic behind the predictions in a random forest is more complex to interpret because it consists of the aggregation of predictions from several trees. Hence, in order to provide the LLM with clear and simple information about the model prediction to generate suggestions and explanations, for the rest of our experiments, we used the random tree regressor to implement the sleep quality forecasting module.

**Table 1.** Sleep quality forecasting results.

| Regressor     | $r$  |
|---------------|------|
| Random Forest | 0.22 |
| Random Tree   | 0.33 |

The achieved results indicate that there is a statistical correlation between the forecast of our system and the actual sleep quality perceived by the individual. In particular, a correlation value  $r$  between 0.2 and 0.5 indicates a moderate correlation between the predicted score and the ground truth. We point out that, in our experiments, the models were trained with data acquired from a relatively small number of individuals. We expect to obtain higher correlation scores using a larger training set.

#### 4.3. User Study

In order to evaluate the utility of the behavior change tips of our system, we conducted a usability test with a group of students from the University of Cagliari, Italy. The study was conducted in June 2023. Participants were recruited through personal contacts and

networking. The participants of our study consisted of 10 students (7 males and 3 females) aged 21 through 26. The average age was 23. All of them had experience in the usage of smartphones and electronic interfaces.

In our study, we employed the assessment framework introduced by Mohseni et al. [35]. The evaluation process involved administering questionnaires and conducting interviews with participants after presenting them with an overview of the system and the related application. To measure agreement with the statements in the questionnaire, we utilized the *Likert scale* method [36].

The questionnaire is presented in Table 2. Each participant had an individual meeting during which we provided a high-level explanation of the system’s functionality. Subsequently, we demonstrated a customized version of our smartphone application without explanations, allowing participants to familiarize themselves with the interface. Afterward, we introduced the original application with explanations and observed the participants’ interactions for a period. We then instructed the participants to independently use both versions of the application until they felt confident about the system’s functionality and potential utility. Finally, the participants completed the questionnaire online through a form.

**Table 2.** Questionnaire for the user study.

| #   | Statement   |
|-----|---|
| Q1  | I have the feeling of trust in the system without explanations regarding its prediction of the sleep quality. |
| Q2  | I have the feeling of trust in the system with explanations regarding its prediction of the sleep quality.    |
| Q3  | The system without explanations can assist me in improving my sleep quality.                                  |
| Q4  | The system with explanations can assist me in improving my sleep quality.                                     |
| Q5  | The explanations would help me decide whether the forecast of the system is correct.                          |
| Q6  | Explanations are easily understandable.   |
| Q7  | Explanations are detailed enough.   |
| Q8  | Explanations are not lengthy.   |
| Q9  | Explanations are useful to understand the reason for the system’s forecast.                                   |
| Q10 | Explanations help reduce the learning time on the system.   |
| Q11 | In general, I am satisfied with the explanations provided by the system.                                      |

The questions primarily focused on the application’s utility, both with and without explanations for the obtained forecast, and whether the explanations were easy to understand. The participation of potential users, particularly students, was crucial in obtaining diverse perspectives and evaluating the application’s suitability within its user base. Their responses contributed to a comprehensive assessment of the application’s utility, highlighting its strengths and areas that required further attention.

The results, illustrated in Figure 10, are visualized using a horizontal bar chart. The chart displays the percentage of responses for each question. This type of chart provides a clear and immediate representation of the emerging trends in the questionnaire, identifying the predominant trends and differences in user responses to the various questions.

The average results for the questions regarding the system without explanations (Q1 and Q3) reached a mean value of 2.7 and 2.8, respectively. In contrast, the average results for the questions regarding the system with explanations (Q2 and Q4) were close to the value of 4 (mean value of 4.1 and 3.8, respectively). These average values highlight an improvement in the trust, usefulness, and understanding of the system when explanations are provided regarding the prediction of sleep quality. The responses suggest that users perceive practical value in the system with explanations and consider them helpful in assessing the system’s reliability and making informed decisions. According to the results, the explanations

provided by the system are considered easily understandable and sufficiently detailed. Moreover, users appreciate the clarity and accessibility of the explanations.

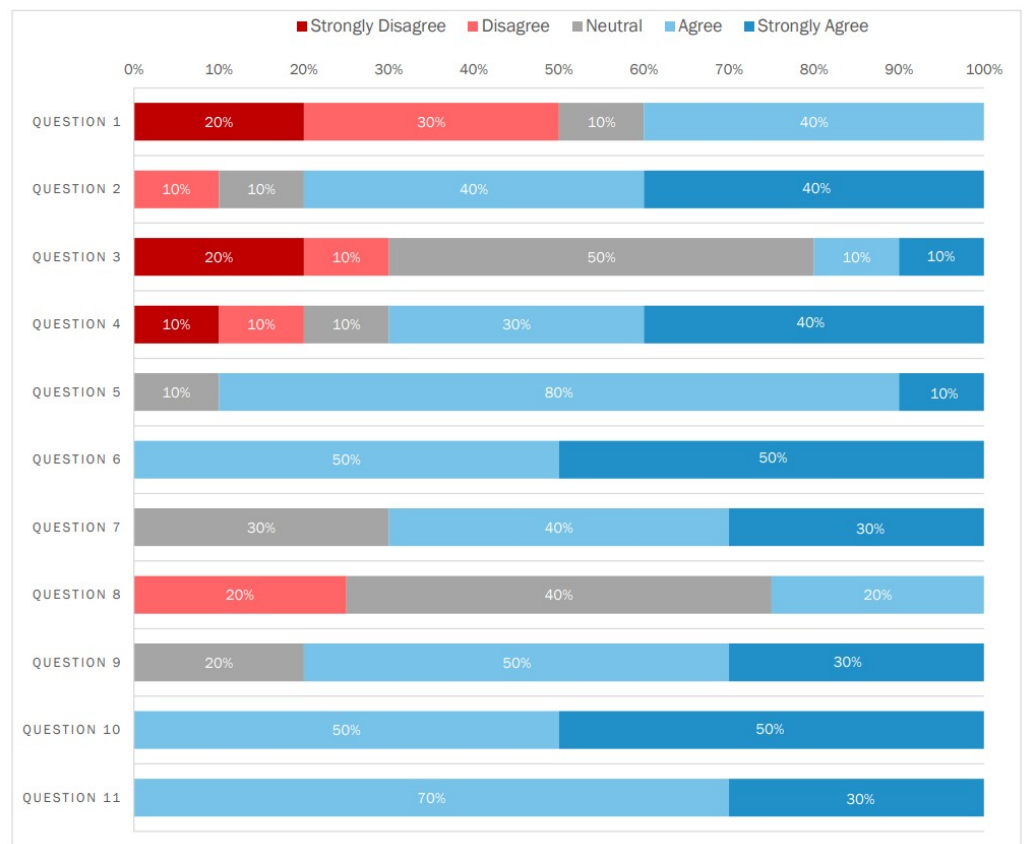


Figure 10. Results of the questionnaire.

Furthermore, users believe that the explanations are useful in understanding the system’s reasoning (Q9, average score 4.1) and reducing the time required to learn how the system functions. This suggests that the explanations provide valuable insights into the system’s decision-making process, helping users to grasp its functioning more efficiently. However, it is worth noting that the average score for the question about the length of explanations (Q8) is 2.4. This suggests that users generally would like the explanations to be more concise, aiming for greater efficiency.

Overall, the system with explanations appears to have a positive impact on the user experience compared to the system without explanations. Users find the explanations easily understandable, appreciate their level of detail, and perceive them as helpful in understanding the system’s reasoning and reducing the learning time.

### 5. Discussion and Limitations

The difficulty of acquiring large training sets of accurately labeled data is a challenging issue in computing and behavior recognition, due to problems of labeling costs and privacy issues [37]. Hence, the experimental evaluation of our work is relatively limited by the dimensions of the available dataset. Our results are, however, in line with those found in the literature. In [38], the authors presented a system to predict sleep quality through the analysis of EEG signals and health and lifestyle information such as ethnicity, gender, education, marital status, age, height, hip and waist circumference, body mass index, weight, and caffeine or alcohol intake. The model is composed of a recurrent neural network that analyzes and extracts information from EEG signals and a regression algorithm that, by analyzing the inferred information from EEG and the most traditional features, predicts sleep quality. The results were evaluated using the coefficient of determination  $r^2$ , which is



equal to the square of the Pearson correlation coefficient  $r$  [39]. The algorithms chosen by the authors were the random forest, AdaBoost, gradient boosting regression, support vector machine, and multilayer perceptron. The best results were obtained in the prediction of the self-reported sleep duration, while, for the prediction of the depth of sleep, the prediction of sleep restfulness, and the prediction of sleep quality compared to the usual, the  $r^2$  values obtained were all less than 0.2, corresponding to  $r \sim 0.44$ . Furthermore, since massive data about users' behavior and sleep quality are not currently available, we could not apply deep learning algorithms, which are known to perform well in this kind of task [24,40] but require very large training sets [41]. As a consequence, we adopted classical ML algorithms that may provide good recognition rates even with few examples [42]. In particular, the random tree algorithm has proven to be effective in sleep monitoring systems [27]. Indeed, based on the achieved results, we can observe a statistical correlation between the forecast of our system and the actual sleep quality perceived by the individual. However, the achieved Pearson correlation value indicates only a moderate correlation. We believe that training the model with more instances would lead to a relevant increase in the recognition performance; however, this should be confirmed by experiments.

As anticipated, with a larger dataset size, predictions of sleep quality could be improved using deep learning methods. The use of deep learning techniques, on the other hand, would make it more difficult to provide easily understandable explanations, due to the extreme complexity of deep learning models [43]. Another limitation of our system is the type of sensors used, which is limited to smartphone sensors. Indeed, integrating other sensors in our system, such as wearables and environmental sensors, would provide additional information to improve sleep forecasting.

The most innovative contribution of our work is the use of LLM to explain the factors influencing the sleep quality forecasts of the ML reasoner, and to provide transparent, context-aware suggestions aimed at enhancing sleep quality. The objective is to foster trust and encourage users' involvement, empowering them to take an active part in improving their sleep. Concerning this aspect of the system, an initial user study indicated that providing explanations on sleep quality forecasts and suggestions to improve it increases the trust, usefulness, and understanding of the system. However, these results should be confirmed through a larger study involving a vast number of volunteers. A longer user study would also allow us to assess the users' adherence to the system's suggestions and to quantitatively assess improvements in sleep quality.

An intrinsic problem of our approach is that, due to its non-deterministic nature, the GPT would likely generate different outputs given the same prompt. This fact introduces uncertainty in the tip generation process. Another limitation of the current work is that the tips are not personalized based on an extensive user profile, since the LLM-based modules only consider the contextual features represented in the feature vectors. Our system should be extended to consider additional profile information that may impact sleep, such as age and occupation. Furthermore, practitioners should be enrolled to validate the correctness of the suggestions provided by our system in order to fine-tune the GPT model and further improve the behavior change tips. Finally, our system could be improved by including domain-specific theoretical knowledge of sleep behavior to refine the ML models for the addressed tasks.

## 6. Conclusions and Future Work

In this paper, we addressed the challenging issue of providing context-aware and transparent behavior change tips to improve sleep quality considering the user's past behavior and the sleep forecast. Differently from previous works in this field, our system couples ML algorithms and LLM to fine-tune the tips and explanations to the past behavior and context of the user. Experiments with real-world data showed the effectiveness of the sleep forecast module, and a user study suggests that our system may increase users' trust and adherence. Several research directions should be investigated in the future. The performance of the system could be improved through the use of a larger dataset. This

would allow us to test additional ML algorithms that require very large training sets, such as deep learning algorithms. Furthermore, having a dataset acquired over a long period of time from a large number of subjects would allow us to evaluate additional characteristics of the system, including the user's adherence to the system's suggestions, and the effectiveness of the system in improving sleep quality in the long term. The performance of the system could also be improved by using different types of sensors, such as wearable sensors, capable of acquiring additional information about the user's body and everyday life. Our system could be further improved by fine-tuning the GPT model to obtain more effective tips. Fine-tuning will require the cooperation of medical personnel to validate and possibly correct the tips produced by the GPT based on the detected context and past behaviors.

**Author Contributions:** Conceptualization, E.C., S.M.M. and D.R.; Methodology, S.M.M. and D.R.; Software, E.C.; Validation, E.C.; Formal analysis, D.R.; Investigation, E.C., S.M.M. and D.R.; Data curation, E.C.; Writing—original draft, E.C., S.M.M. and D.R.; Supervision, D.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** We acknowledge the financial support under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.5—Call for Tender No. 3277, published on 30 December 2021 by the Italian Ministry of University and Research (MUR), funded by the European Union—NextGenerationEU. Project Code ECS0000038—Project Title eINS Ecosystem of Innovation for Next Generation Sardinia—CUP F53C22000430001—Grant Assignment Decree No. 1056, adopted on 23 June 2022 by the Italian Ministry of University and Research (MUR).

**Data Availability Statement:** The data presented in this study are available in this article.

**Acknowledgments:** The authors would like to thank the anonymous reviewers for their insightful comments and suggestions to improve the technical content and presentation of this paper.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Bixler, E. Sleep and society: An epidemiological perspective. *Sleep Med.* **2009**, *10*, S3–S6. [[CrossRef](#)]
2. Luyster, F.S.; Strollo, P.J., Jr.; Zee, P.C.; Walsh, J.K. Sleep: A health imperative. *Sleep* **2012**, *35*, 727–734. [[CrossRef](#)]
3. Grandner, M.A. Sleep, health, and society. *Sleep Med. Clin.* **2017**, *12*, 1–22. [[CrossRef](#)]
4. Diekelmann, S. Sleep for cognitive enhancement. *Front. Syst. Neurosci.* **2014**, *8*, 46. [[CrossRef](#)]
5. Nicholson, W.C.; Pfeiffer, K. Sleep disorders and mood, anxiety, and post-traumatic stress disorders: Overview of clinical treatments in the context of sleep disturbances. *Nurs. Clin.* **2021**, *56*, 229–247.
6. Besedovsky, L.; Lange, T.; Born, J. Sleep and immune function. *Pflügers-Arch.-Eur. J. Physiol.* **2012**, *463*, 121–137. [[CrossRef](#)] [[PubMed](#)]
7. Kim, T.W.; Jeong, J.H.; Hong, S.C. The impact of sleep and circadian disturbance on hormones and metabolism. *Int. J. Endocrinol.* **2015**, *2015*, 591729. [[CrossRef](#)]
8. Liew, S.C.; Aung, T. Sleep deprivation and its association with diseases—a review. *Sleep Med.* **2021**, *77*, 192–204. [[CrossRef](#)] [[PubMed](#)]
9. Kent, R.G.; Uchino, B.N.; Cribbet, M.R.; Bowen, K.; Smith, T.W. Social relationships and sleep quality. *Ann. Behav. Med.* **2015**, *49*, 912–917. [[CrossRef](#)]
10. Gordon, A.M.; Carrillo, B.; Barnes, C.M. Sleep and social relationships in healthy populations: A systematic review. *Sleep Med. Rev.* **2021**, *57*, 101428. [[CrossRef](#)]
11. Pan, Q.; Brulin, D.; Campo, E. Current status and future challenges of sleep monitoring systems: Systematic review. *JMIR Biomed. Eng.* **2020**, *5*, e20921. [[CrossRef](#)]
12. Shochat, T. Impact of lifestyle and technology developments on sleep. *Nat. Sci. Sleep* **2012**, *4*, 19–31. [[CrossRef](#)] [[PubMed](#)]
13. Troncoso-García, A.; Martínez-Ballesteros, M.; Martínez-Álvarez, F.; Troncoso, A. Explainable machine learning for sleep apnea prediction. *Procedia Comput. Sci.* **2022**, *207*, 2930–2939. [[CrossRef](#)]
14. Tsai, C.H.; Christian, M.; Kuo, Y.Y.; Lu, C.C.; Lai, F.; Huang, W.L. Sleep, physical activity and panic attacks: A two-year prospective cohort study using smartwatches, deep learning and an explainable artificial intelligence model. *Sleep Med.* **2023**, *114*, 55–63. [[CrossRef](#)] [[PubMed](#)]
15. Jany, R.; Ashmafee, M.H.; Hussain, I.; Hossain, M.A. SleepExplain: Explainable Non-Rapid Eye Movement and Rapid Eye Movement Sleep Stage Classification from EEG Signal. In Proceedings of the 2022 25th International Conference on Computer and Information Technology (ICCIT), Cox's Bazar, Bangladesh, 17–19 December 2022; pp. 248–253.

16. Mira, F.A.; Favier, V.; dos Santos Sobreira Nunes, H.; de Castro, J.V.; Carsuzaa, F.; Meccariello, G.; Vicini, C.; De Vito, A.; Lechien, J.R.; Estomba, C.C.; et al. Chat GPT for the management of obstructive sleep apnea: Do we have a polar star? In *European Archives of Oto-Rhino-Laryngology*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 1–7.
17. Kaul, V.; Enslin, S.; Gross, S.A. History of artificial intelligence in medicine. *Gastrointest. Endosc.* **2020**, *92*, 807–812. [[CrossRef](#)]
18. Zolfaghari, S.; Suravee, S.; Riboni, D.; Yordanova, K. Sensor-based Locomotion Data Mining for Supporting the Diagnosis of Neurodegenerative Disorders: A Survey. *ACM Comput. Surv.* **2023**, *56*, 1–36. [[CrossRef](#)]
19. Zolfaghari, S.; Khodabandehloo, E.; Riboni, D. TraMiner: Vision-based analysis of locomotion traces for cognitive assessment in smart-homes. *Cogn. Comput.* **2022**, *14*, 1549–1570. [[CrossRef](#)] [[PubMed](#)]
20. Secinaro, S.; Calandra, D.; Secinaro, A.; Muthurangu, V.; Biancone, P. The role of artificial intelligence in healthcare: A structured literature review. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 1–23. [[CrossRef](#)]
21. Thieme, A.; Belgrave, D.; Doherty, G. Machine learning in mental health: A systematic review of the HCI literature to support the development of effective and implementable ML systems. *ACM Trans. Comput. Hum. Interact. (TOCHI)* **2020**, *27*, 1–53. [[CrossRef](#)]
22. Scott, A.J.; Webb, T.L.; Martyn-St James, M.; Rowse, G.; Weich, S. Improving sleep quality leads to better mental health: A meta-analysis of randomised controlled trials. *Sleep Med. Rev.* **2021**, *60*, 101556. [[CrossRef](#)]
23. Mendonça, F.; Mostafa, S.S.; Morgado-Dias, F.; Ravelo-Garcia, A.G.; Penzel, T. A review of approaches for sleep quality analysis. *IEEE Access* **2019**, *7*, 24527–24546. [[CrossRef](#)]
24. Sathyanarayana, A.; Joty, S.; Fernandez-Luque, L.; Ofli, F.; Srivastava, J.; Elmagarmid, A.; Arora, T.; Taheri, S. Sleep quality prediction from wearable data using deep learning. *JMIR mHealth uHealth* **2016**, *4*, e6562.
25. Arora, A.; Chakraborty, P.; Bhatia, M. Analysis of data from wearable sensors for sleep quality estimation and prediction using deep learning. *Arab. J. Sci. Eng.* **2020**, *45*, 10793–10812. [[CrossRef](#)]
26. Sadeghi, R.; Banerjee, T.; Hughes, J.C.; Lawhorne, L.W. Sleep quality prediction in caregivers using physiological signals. *Comput. Biol. Med.* **2019**, *110*, 276–288. [[CrossRef](#)]
27. Chang, X.; Peng, C.; Xing, G.; Hao, T.; Zhou, G. Isleep: A smartphone system for unobtrusive sleep quality monitoring. *ACM Trans. Sens. Netw. (TOSN)* **2020**, *16*, 1–32. [[CrossRef](#)]
28. Alamoudi, D.; Breeze, E.; Crawley, E.; Nabney, I. The feasibility of using smartphone sensors to track insomnia, depression, and anxiety in adults and young adults: Narrative review. *JMIR mHealth uHealth* **2023**, *11*, e44123. [[CrossRef](#)] [[PubMed](#)]
29. Deng, J.; Lin, Y. The benefits and challenges of ChatGPT: An overview. *Front. Comput. Intell. Syst.* **2022**, *2*, 81–83. [[CrossRef](#)]
30. Susnjak, T. Beyond Predictive Learning Analytics Modelling and onto Explainable Artificial Intelligence with Prescriptive Analytics and ChatGPT. *Int. J. Artif. Intell. Educ.* **2023**, 1–31. [[CrossRef](#)]
31. Bachechi, C.; Rollo, F.; Po, L. Real-time data cleaning in traffic sensor networks. In Proceedings of the 2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA), Antalya, Turkey, 2–5 November 2020; pp. 1–8.
32. Wang, R.; Chen, F.; Chen, Z.; Li, T.; Harari, G.; Tignor, S.; Zhou, X.; Ben-Zeev, D.; Campbell, A.T. StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Seattle, WA, USA, 13–17 September 2014.
33. Khodabandehloo, E.; Riboni, D.; Alimohammadi, A. HealthXAI: Collaborative and explainable AI for supporting early diagnosis of cognitive decline. *Future Gener. Comput. Syst.* **2021**, *116*, 168–189. [[CrossRef](#)]
34. Lee Rodgers, J.; Nicewander, W.A. Thirteen ways to look at the correlation coefficient. *Am. Stat.* **1988**, *42*, 59–66. [[CrossRef](#)]
35. Mohseni, S. Toward Design and Evaluation Framework for Interpretable Machine Learning Systems. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, Honolulu, HI, USA, 27–28 January 2019; Conitzer, V., Hadfield, G.K., Vallor, S., Eds.; ACM: New York, NY, USA, 2019; pp. 553–554.
36. Joshi, A.; Kale, S.; Chandel, S.; Pal, D.K. Likert scale: Explored and explained. *Curr. J. Appl. Sci. Technol.* **2015**, *7*, 396–403. [[CrossRef](#)]
37. Ye, J.; Dobson, S.; McKeever, S. Situation identification techniques in pervasive computing: A review. *Pervasive Mob. Comput.* **2012**, *8*, 36–66. [[CrossRef](#)]
38. Boussard, J.; Kochenderfer, M.J.; Zeitzer, J.M. Predicting Subjective Sleep Quality Using Recurrent Neural Networks. In Proceedings of the 2019 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), Philadelphia, PA, USA, 7 December 2019; pp. 1–6.
39. Schober, P.; Boer, C.; Schwarte, L.A. Correlation coefficients: Appropriate use and interpretation. *Anesth. Analg.* **2018**, *126*, 1763–1768. [[CrossRef](#)] [[PubMed](#)]
40. Park, S.; Li, C.T.; Han, S.; Hsu, C.; Lee, S.W.; Cha, M. Learning sleep quality from daily logs. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 2421–2429.
41. Sarker, I.H. Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Comput. Sci.* **2021**, *2*, 420. [[CrossRef](#)]

42. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
43. Arrieta, A.B.; Diaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; García, S.; Gil-López, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 82–115. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.