

Original Paper

Sleep Quality Prediction From Wearable Data Using Deep Learning

Aarti Sathyaranayana¹, MSc; Shafiq Joty¹, PhD; Luis Fernandez-Luque¹, PhD; Ferda Ofli¹, PhD; Jaideep Srivastava¹, PhD; Ahmed Elmagarmid¹, PhD; Teresa Arora², PhD; Shahrad Taheri², MBBS, PhD

¹Qatar Computing Research Institute, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar

²Department of Medicine, Weill Cornell Medical College in Qatar, Qatar Foundation, Doha, Qatar

Corresponding Author:

Luis Fernandez-Luque, PhD

Qatar Computing Research Institute

Hamad Bin Khalifa University, Qatar Foundation

HBKU Research Complex

Doha, 5825

Qatar

Phone: 974 50173040

Fax: 974 44540630

Email: lluque@qf.org.qa

Related Article:

This is a corrected version. See correction statement: <http://mhealth.jmir.org/2016/4/e130/>

Abstract

Background: The importance of sleep is paramount to health. Insufficient sleep can reduce physical, emotional, and mental well-being and can lead to a multitude of health complications among people with chronic conditions. Physical activity and sleep are highly interrelated health behaviors. Our physical activity during the day (ie, awake time) influences our quality of sleep, and vice versa. The current popularity of wearables for tracking physical activity and sleep, including actigraphy devices, can foster the development of new advanced data analytics. This can help to develop new electronic health (eHealth) applications and provide more insights into sleep science.

Objective: The objective of this study was to evaluate the feasibility of predicting sleep quality (ie, poor or adequate sleep efficiency) given the physical activity wearable data during awake time. In this study, we focused on predicting good or poor sleep efficiency as an indicator of sleep quality.

Methods: Actigraphy sensors are wearable medical devices used to study sleep and physical activity patterns. The dataset used in our experiments contained the complete actigraphy data from a subset of 92 adolescents over 1 full week. Physical activity data during awake time was used to create predictive models for sleep quality, in particular, poor or good sleep efficiency. The physical activity data from sleep time was used for the evaluation. We compared the predictive performance of traditional logistic regression with more advanced deep learning methods: multilayer perceptron (MLP), convolutional neural network (CNN), simple Elman-type recurrent neural network (RNN), long short-term memory (LSTM-RNN), and a time-batched version of LSTM-RNN (TB-LSTM).

Results: Deep learning models were able to predict the quality of sleep (ie, poor or good sleep efficiency) based on wearable data from awake periods. More specifically, the deep learning methods performed better than traditional logistic regression. “CNN had the highest specificity and sensitivity, and an overall area under the receiver operating characteristic (ROC) curve (AUC) of 0.9449, which was 46% better as compared with traditional logistic regression (0.6463).

Conclusions: Deep learning methods can predict the quality of sleep based on actigraphy data from awake periods. These predictive models can be an important tool for sleep research and to improve eHealth solutions for sleep.

(*JMIR Mhealth Uhealth* 2016;4(4):e125) doi:[10.2196/mhealth.6562](https://doi.org/10.2196/mhealth.6562)

KEYWORDS

wearables; sleep quality; sleep efficiency; actigraphy; body sensor networks; mobile health; connected health; accelerometer; physical activity; pervasive health; consumer health informatics; deep learning

Introduction

Background

The importance of sleep is paramount to health and performance. Insufficient sleep can impede physical, emotional, and mental well-being [1,2] and can lead to a multitude of health complications such as insulin resistance [3-5], high blood pressure [6], cardiovascular disease [7,8], a compromised immune or metabolic system [9,10], mood disorders (such as depression or anxiety) [11,12], and decreased cognitive function for memory and judgment [13-15].

There are many indicators of sleep quality, an important one being sleep efficiency. Sleep efficiency is a metric that takes into consideration the time spent asleep, the time it takes to fall asleep, and the time asleep but with disturbance (ie, awakenings). Poor sleep efficiency can lead to sleep deprivation, which is found to be a major health risk with links to diseases such as diabetes and obesity [16]. Also, sleep behavior has been found to impact adolescent health [17-19]. Recent systematic reviews have shown the relevance of physical activity to sleep, including sleep efficiency [20-22]. Although the relationship between physical activity and sleep is not yet fully understood, it is thought to be a strong and complex correlation contributing to multiple lifestyle diseases such as type 2 diabetes mellitus and obesity [20-22]. However, the underlying mechanism between physical activity and sleep is not yet fully understood.

Role of Wearables in Sleep Health and eHealth

There are many research tools to study the link between physical activity and sleep, including standardized questionnaires, actigraphy sensors, and polysomnography (PSG). All of these methods have different clinical indications (eg, diagnosis of different sleep disorders) [23]. For example, PSG is considered the “gold standard” for sleep medicine, as it involves the use of multiple sensors during sleep such as electroencephalogram, motion sensors, breathing sensors, SpO₂(oxygen saturation), and so forth. These sensors monitor and observe a patient overnight [24] and can be used for diagnosing different sleep disorders. The elaborate nature of PSG means that it is generally limited to one overnight observation. Even the portable solutions, which permit PSG assessment in the patient’s home, are complex to perform and not without limitations [25,26].

To better understand the impact of daily physical activity on sleep behavior, new tools are needed. Sleep researchers in the early 1990s developed a technique called *actigraphy* to study sleep interactions using wearable devices [27]. Although actigraphy traditionally uses wearable devices to evaluate the sleep period of a patient, it can also be used to observe physical activity. Actigraphy has become a widely used tool, as it has been found to be much more reliable than subjective or self-reported sleep diaries and behavior logs [28]. Patients wear the device for a period of time as they continue their daily routines. The technique has been especially influential for large

cohort studies where PSG is not feasible [29]. Moreover, actigraphy allows for the continuous longitudinal monitoring of a patient. This is particularly impactful for the study of diseases such as chronic obstructive pulmonary disease where sleep disturbances can be a predictor of exacerbation of the disease [30]. Current approaches to the analysis of actigraphy data involve sleep experts performing a number of steps manually. This is a bottleneck, and hence there may be troves of actigraphy data left unanalyzed.

Furthermore, there are hundreds of consumer-grade physical activity and sleep tracking devices (eg, Fitbit) collecting motion data similarly to actigraphy devices. These consumer devices are being used by millions of people collecting huge amounts of data. Some devices even allow the collection of data for very long periods (eg, several months), as they require only an occasional need for battery recharge (ie, Garmin VivoFit). Recent studies have found that these consumer-grade devices can sometimes have similar precision to clinical-grade actigraphy sensors [31]. There are successful examples of the integration of physical activity wearable data into eHealth tailored applications [32], including smart-watch health applications [33] that can collect physical activity and sleep data directly from the watch.

Objective

As previously noted, physical activity and sleep are interrelated. In this study, we tested the feasibility of predicting poor or good sleep efficiency based on physical activity data from the awake periods from a wearable device (ie, actigraphy). We took a detour from classical methods and proposed deep learning approaches to modeling the relationship between sleep and physical activity. The importance of this research is two-fold.

First, since our approach can be used in cases where sleep sensor data is not available, our models can be used in the early detection of potential low sleep efficiency. This is a common problem with consumer-grade wearable devices, as users might not wear them during the night (battery recharging, sensors embedded in smart jewelry, and so forth).

Second, our study was focused on advanced deep learning methods. Traditional prediction models applied to activity raw accelerometer data (eg, logistic regression) suffer from at least 2 key limitations: (1) They are not robust enough to learn useful patterns from noisy raw accelerometer output. As a result, existing methods for classification and analysis of physical activity rely on extracting higher-level features that can be fed into prediction models [34]. This process often requires domain expertise and can be time consuming. (2) Traditional methods do not exploit task labels for feature construction, and thus can be limited in their ability to learn task-specific features. Deep learning has the advantage that it is robust to raw noisy data, and can learn, automatically, higher level abstract features by passing raw input signals through nonlinear hidden layers while also optimizing on the target prediction tasks. We leveraged this characteristic by building models using a range of deep

learning methods on raw accelerometer data. This reduced the need for data preprocessing and feature space construction and simplified the overall workflow for clinical practice and sleep researchers.

Methods

Overview

The methodology of this study involved several steps: (1) the collection of wearable data with regards to physical activity and sleep patterns, (2) data processing and representation, (3) data modeling, and (4) performance evaluation. The following subsections explain each of these steps.

Data Collection

The deidentified data used in this study were collected by Weill Cornell Medical College-Qatar for a research study called Qatar's Ultimate Education for Sleep in Teenagers. The aim of the study was to determine different sleep patterns in adolescents residing in Qatar and how these related to body weight status. The institutional review board (IRB) approval was initially obtained by the joint IRB for Hamad Medical Corporation and Weill Cornell Medical College. The selected cohort was chosen from adolescents attending 1 of the 2 high schools and registered

in grades 7-11. The dataset used contained complete actigraphy data from a subset of 92 adolescents over 1 week. There were 322 total sleep instances: 102 from boys and 220 from girls.

After agreeing to participate in the study, the adolescent participants were provided with an ActiGraph GT3X+ device, placed on their nondominant wrist, and were instructed to wear it at all times for 7 consecutive days and nights. The ActiGraph GT3X+, shown in [Figure 1](#), is a clinical-grade wearable device that samples a user's activity at 30-100 Hz (we used 30 Hz in this study). The effectiveness of this device has been successfully validated against clinical polysomnography [35]. The participants were instructed not to remove the water-resistant device at any time. We used the manufacturer's software (ActiLife version 6; available from ActiGraph, LLC) to export the data. Although the device is triaxial, our methods used the vertical axis only.

The subjective self-reported sleep diaries were collected, but they were not considered for the study, since actigraphy data provides a more objective measurement of physical activity and sleep patterns. Sadeh's and American Academy of Sleep Medicine sleep definitions were used in this interpretation [27,36,37]. For calculating the sleep efficiency score, we considered each individual sleeping period as sleep.

Figure 1. ActiGraph Gt3X+.



Data Processing and Representation

Sleep Quality Definitions

In a person's activity time series, that is, the continuous data collected from a wearable device, there are moments when an individual is awake and when they are asleep. The latter is referred to as the sleep period. The boundary from awake time to the sleep period is called the sleep onset time, and the boundary from sleep period to awake time is referred to as the sleep awakening time. The period of time between the self-reported time to bed and the sleep onset time is called the latency.

To measure sleep quality we determined sleep efficiency (see [Figure 2](#)), which is the ratio of total minutes asleep to total minutes in bed. Those achieving a sleep efficiency score of $\geq 85\%$ are thought to be good-quality sleepers, and those with a score of $< 85\%$ are thought to have poor-quality sleep [38].

Total minutes in bed represents the amount of time that an individual spends asleep as well as the amount of time the individual takes to fall asleep, that is, latency. Total sleep time represents the amount of time that an individual spends asleep, less the amount of time the person awakens. This is calculated by subtracting the wake after sleep onset (WASO) from the

duration of the sleep period. WASO is the sum of all moments of wakefulness lasting longer than 5 minutes (see [Figure 3](#)).

Figure 2. Sleep efficiency equation as defined by sleep specialists.

$$\begin{aligned} \text{Sleep Efficiency} &= \frac{\text{TotalSleepTime}}{\text{TotalMinutesinBed}} \\ &= \frac{||\text{SleepPeriod}|| - \text{WASO}}{||\text{SleepPeriod}|| + \text{Latency}} \end{aligned}$$

Figure 3. The adapted wake after sleep onset calculation.

$$\text{WASO} = \sum_{\substack{\text{Sleep} \\ \text{Awakening} \\ \text{Time}}}^{\text{Sleep} \\ \text{Onset} \\ \text{Time}} \left\{ \begin{array}{ll} ||\text{WakefulPeriod}||, & ||\text{WakefulPeriod}|| > 5 \\ 0, & \text{otherwise} \end{array} \right.$$

Data Processing

The data collected from the actigraphy device contained triaxial accelerometer movement. Previous research has identified that for devices worn on the wrist, the vertical axis from accelerometer data is the most indicative of physical activity [39]. Consequently, we switched to 1-dimensional input for a simpler and less noisy modeling (ie, each row in our dataset contained a 1-dimensional time series of vertical axis accelerometer data).

The raw accelerometer data was aggregated into minute-by-minute epochs using a script written in R version 3.3.1, a statistical Open Source computing software developed as part of the R Project. In this dataset, the time series contains activity data during both the *awake time* and the *sleep time* (as shown in [Figure 4](#)). We used the awake time to form the prediction (ie, as the model input) and the sleep time to determine the ground truth of sleep quality. Sleep quality (good or poor) was determined from the sleep time actigraphy data, and each time series was labeled as such.

As mentioned earlier, sleep onset time and sleep awakening time are metrics that form the boundary of the sleep period [36]. We interpreted and expanded these values for accelerometer data according to Sadeh's actigraphy definitions [37].

Sleep onset time is traditionally defined as the first minute of 15 continuous minutes of sleep after a self-reported bedtime,

and the *sleep awakening time* is the last minute of 15 continuous minutes of sleep that is followed by 30 minutes of movement [37]. To automate this interpretation directly from the accelerometer output, we developed the concept of candidate rows. *Candidate rows* denote moments (or designated epochs) in time with a lack of triaxial movement and require a subsequent pass to determine whether the individual is asleep or awake. Each row is iterated upon and run through a state machine, as illustrated in [Figure 5](#). Since there is no self-reported bedtime, we inferred it as the beginning of sedentary behavior immediately preceding and adjacent to the start of the sleep period. The duration of this sedentary time is the *latency*.

The handling of nonwearing time is very important and should be minimized wherever possible. The device was water-resistant and did not require recharging during the time period the participants were instructed to wear it. If a participant removed the device, the triaxial accelerometer would record zero values for the time it was not worn. Natural human behavior makes micro-movements that are sensed by the accelerometer even during sleep, and so periods with a continuous lack of movement indicate device removal. In other words, during the time the device was not worn, our algorithm would identify candidate rows for the entire period, denote them as sleep time, and score the sleep as having a perfect sleep efficiency of 1. Alternatively, the ActiLife software includes an algorithm to remove nonwear periods from physical activity calculations [40].

Figure 4. Example of sleep definitions on accelerometer data of an actigraphy device.

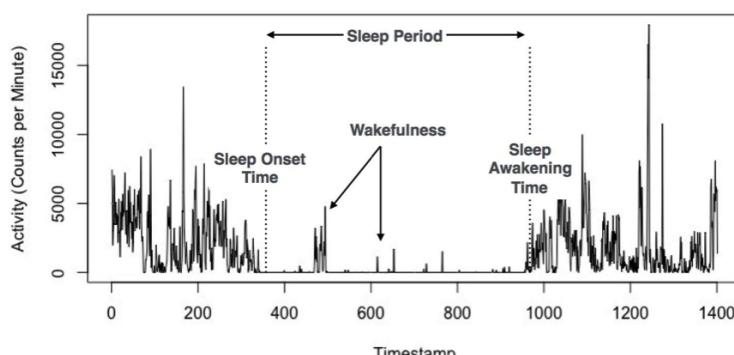
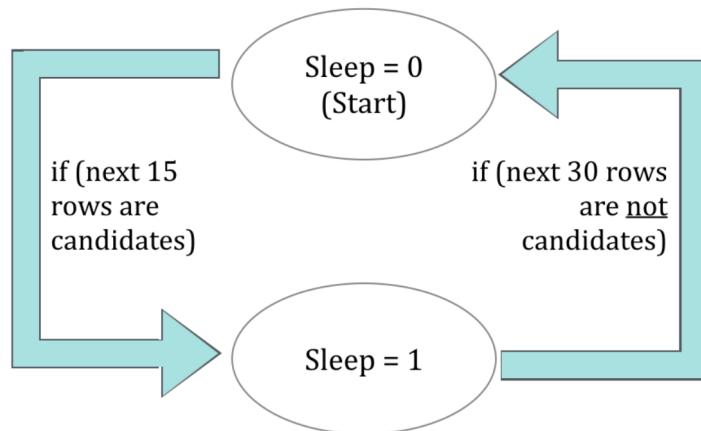


Figure 5. State machine diagram explaining the designation of sleep or awake time periods.

Models for Sleep Quality

As explained in more detail below, in this study we explored the use of deep learning methods to predict sleep quality based on actigraphy data. We compared the results of our deep learning models to those of logistic regression, a standard statistical method. In this section, we explain how each model was built. The models used in our study were as follows:

- Logistic regression, a nondeep learning model
- Multi-layer perceptrons (MLPs), a deep learning model
- Convolutional neural network (CNN), a deep learning model
- Recurrent neural networks (RNN), a deep learning model
- Long short-term memory (LSTM) RNN, a deep learning model
- Time-batched long short-term memory (TB-LSTM) RNN, a deep learning model

For logistic regression, we fed the input signals directly into the output layer for prediction. In contrast, for the deep learning models, we passed the inputs through one or more hidden layers before they were fed to the output layer for prediction. Each hidden unit in the hidden layers used a nonlinear activation function. In our study, we experimented with different hidden layer units using rectified linear unit as our activation function.

Data Partitioning

To train our models without over-fitting and test their performance afterward, we created a random partitioning of the dataset. Each time series was assigned to a partition randomly while maintaining an even class distribution of the target variable, sleep quality. The data were split with a 70%-15%-15% ratio for training, testing, and validation sets, respectively.

Input and Output of the Models

The input of the models were time series vectors, $X=(x_1, \dots, x_T)$, representing the physical activity of a person's awake time. Each vector corresponded to a continuous period of awake time, and so for each individual, there might be multiple such vectors over the 7 days. Each x_T represented the value of the vertical axis at time t .

The output of the model was a binary classification decision between *good* and *poor* sleep quality based on sleep efficiency

(%). These classifications corresponded to the sleep efficiency definitions as described earlier. In addition to the binary decision, the model also gave its confidence (a score between 0.0 and 1.0) in that decision.

Training the Models

To be able to predict, we first trained the models on the training dataset. We used an online training algorithm RMSprop [41], which relied on a number of preset parameters:

- Mini-batch size: how many training instances to consider at one time.
- Learning rate: the rate at which parameters are updated.
- Max epoch: maximum number of iterations over the training set.
- Dropout ratio: ratio of hidden units to turn off in each mini-batch training.

The training algorithm minimizes the cross-entropy between the predicted distribution and the actual (gold) target labels. To avoid over-fitting, we used early stopping based on the model's performance on the validation set. In particular, we evaluated the model after every epoch on the validation set and stopped when its accuracy went down. To reduce the cross-entropy between the predicted distributions and the target distributions, RMSprop was used setting the maximum number of epochs to 50 as suggested by the authors [41].

Logistic Regression (a Nondeep Learning Model)

As a baseline, we used logistic regression (LR) to predict the sleep quality. LR is a generalized linear classification model that does not have any hidden layers. For the LR, the raw input signals X are directly fed to the output layer for prediction without any nonlinear hidden layer transformations. The optimal setting for logistic regression (LR) was with a mini-batch size of 5 and a dropout ratio of 0.5.

Multilayer Perceptrons (a Deep Learning Model)

MLPs, also known as feed-forward neural networks, are the simplest models in the deep learning family. They have one or more hidden layers. In fact, MLP without any hidden layers is equivalent to logistic regression. In MLP, all the units of a hidden layer are fully connected to the units in the previous layer. The best parameter configuration for MLP was with a

mini-batch size of 20, a dropout ratio of 0.1, and a hidden layer size of 15.

Convolutional Neural Network (a Deep Learning Model)

CNNs are a more complex type of deep learning method that includes repetitive filters or kernels applied to local time slots, thereby composing a high level of abstract features. This operation is called *convolution*. After convolution, a *max-pooling* operation is performed to select the most significant abstract features. This design of CNNs yields fewer parameters than its fully connected counterpart (MLP), and therefore generalizes well for target prediction tasks. For its best configuration, we used 25 hidden nodes, filter length of 5 and pooling length of 4, 5 mini-batch size, and 0.0 dropout ratio.

Recurrent Neural Networks (a Deep Learning Models)

RNNs compose abstract features by processing activity measures in an awake time sequentially, at each time step combining the current input with the previous hidden state. RNNs create internal states by remembering the previous hidden layer, which allows them to exhibit dynamic temporal behavior. These features make RNNs a good deep learning method for temporal series. RNNs performed best with a mini-batch size of 5, a dropout ratio of 0.1 and a hidden layer size of 75. To avoid over-fitting, we used a technique based on dropout of hidden units and early stopping based on the loss on the development set [42].

Long Short-Term Memory (a Deep Learning Model)

A subtype of RNN, *LSTM* uses specifically designed memory blocks as units in the recurrent layer to capture longer-range dependencies. The optimal configuration values for LSTM were a mini batch size of 5, dropout ratio of 0.5, and hidden layer size of 100.

Time-Batched Long Short-Term Memory (a Deep Learning Model)

To further improve our implementation of LSTM, we constructed batches of time steps by merging accelerometer measures over time steps. We referred to this version of the model as *TB-LSTM*. The configuration values for TB-LSTM were mini-batch size of 5, dropout ratio of 0.5, and hidden layer size of 100.

Performance Evaluation

For the evaluation of the performance of the different models, we reported on several well-known metrics such as accuracy, precision, recall, F1-score, and area under the receiver operating characteristic (ROC) curve (AUC). These metrics are commonly used in data mining and clinical decision support systems.

Accuracy

It is computed as the proportion of correct predictions, both positive and negative (sum of true positives and true negatives divided by the number of all instances in the dataset).

Precision

It is the fraction of the number of true positive predictions to the number of all positive predictions (ie, true positives divided by the sum of true positives and false positives). In our case, precision described what percentage of the time the model predicted “good-quality sleep” correctly. Note that precision is also known as positive predictive value.

Specificity

It is the fraction of the number of true negative predictions to the actual number of negative instances in the dataset (ie, true negatives divided by the sum of true negatives and false positives). In our case, specificity referred to the percentage of the correctly predicted “poor-quality sleep” to the total number of “poor-quality sleep” instances in the dataset. Note that specificity is also known as true negative rate.

Recall or Sensitivity

It is the fraction of the number of true positive predictions to the actual number of positive instances in the dataset (ie, true positives divided by the sum of true positives and false negatives). In our case, recall referred to the percentage of the correctly predicted “good-quality sleep” to the total number of “good-quality sleep” instances in the dataset. Note that recall is also known as true positive rate or sensitivity.

F1-Score

There is usually an inverse relationship between precision and recall. That is, it is possible to increase the precision at the cost of decreasing the recall, or vice versa. Therefore, it is more useful to combine them into a single measure such as F1 score, which computes the harmonic mean of precision and recall.

Area Under the ROC Curve

It represents the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance. Hence, AUC defines an effective and combined measure of sensitivity and specificity (which are often inversely related, just like precision and recall) for assessing inherent validity of a classifier.

Results

Comparison Between Deep Learning and Logistic Regression

As shown in Table 1 and Figure 6, the performance of the logistic regression in the metrics previously explained performed worse than the models based on deep learning. “Only the simple RNN performed worse than logistic regression in both F1-score (harmonic mean of precision and recall) and accuracy.

Table 1. Results on raw accelerometer data.

Data models	AUC ^a	F1-Score	Precision	Recall	Accuracy
LR ^b	0.6463	0.8193	0.7083	0.9714	0.7321
MLP ^c	0.9449	0.9118	0.9394	0.8857	0.8929
CNN ^d	0.9456	0.9444	0.9189	0.9714	0.9286
RNN ^e	0.7143	0.7711	0.6667	0.9143	0.6607
LSTM ^f -RNN	0.8531	0.8500	0.7556	0.9714	0.7857
TB-LSTM ^g	0.9714	0.9211	0.8537	1.000	0.8929

^aAUC: Area under receiver operating characteristic (ROC) curve.

^bLR: logistic regression.

^cMLP: Multilayer perceptrons.

^dCNN: convolutional neural networks.

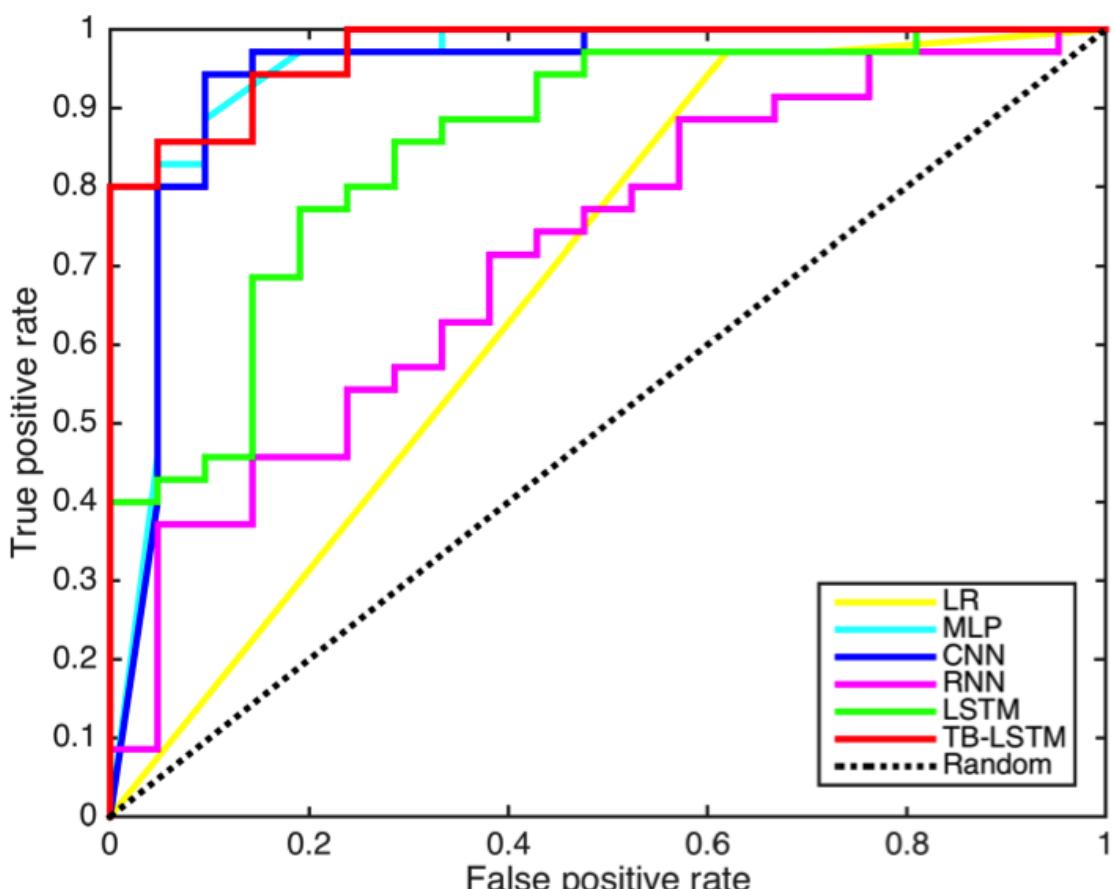
^eRNN: recurrent neural networks.

^fLSTM: long short-term memory.

^gTB-LSTM: time-batched LSTM.

As shown in **Table 1**, the AUC of the logistic regression model was low. The AUC value for LR was 0.6463, which was close to 0.5 (equivalent to a random prediction). This showed the limitation of classical models in analyzing raw accelerometry.

In contrast, all the AUC values for the deep learning models were better with a range from 0.7143 to 0.9714, TB-LSTM being the best performer and RNN, the worst. Time-batched LSTM, CNN, and MLP performed the best with AUC scores showing an improvement over LR by 50%, 46%, and 46%, respectively.

Figure 6. Receiver operating characteristic (ROC) curves for each model's prediction of sleep efficiency.

Comparison Between Deep Learning Models

Upon comparing the deep learning neural network models, we noticed that CNN yielded slight improvement over MLP in AUC (0.07% absolute), but more in F1 (3.57%) and accuracy (4.00%). These improvements could be attributed to the time-invariant convolution-pooling operations of the CNN model to pick key local patterns, which generalized well for small training data. The F1-score improved by 12% for time-batched LSTM, by 15% for CNN, and by 11% for MLP. The accuracy improved by 22% for time-batched LSTM, by 27% for CNN, and by 22% for MLP.

A comparison of the RNN models revealed that LSTM outperformed simple RNN by a wide margin; 19%, 10%, and 19% in AUC, F1, and accuracy, respectively. These gains over simple RNNs could be attributed to the specially designed gates of LSTMs that could capture long-range dependencies between physical activities in the sequences.

However, this is not surprising. Both simple and LSTM RNNs operate on sequences, where each time step comprises only one activity value. This often results in very long sequences. As mentioned earlier, in this setting RNNs cannot compose higher-level features effectively because of low-dimensional input at each time step, and also suffer from vanishing gradient problems due to lengthy sequences.

Our solution to surmount this problem was to use a time-batched input to LSTM. When we compared the results of our time-batched LSTM (TB-LSTM) with those of MLP and CNN, we found that TB-LSTM outperformed both MLP and CNN in AUC by 3%; in fact, it had the highest AUC score. It achieved better F1 score than MLP (1%), but worse than CNN (-2%). When we observed their precision and recall values, we found

that TB-LSTM had a very high recall but lower precision, which meant that it tended to predict more goods than gold standard. For the same reason, its accuracy was also lower than that of CNN.

Discussion

Principal Findings

In our study, we focused on the prediction of poor versus good sleep efficiency. That is a simple, but important, problem, as sleep efficiency has been found to be a crucial sleep parameter with important health consequences [38,43,44]. Furthermore, we did not quantify in our prediction the overall sleep efficiency but simply the differentiation between two classes (poor versus good sleep efficiency). This classification is consequently not an indicator of sleep patterns, but the prediction of a sleep quality parameter that might indicate a potential sleep problem.

As in prediction or diagnostic problems, our results need to be discussed in terms of sensitivity and specificity (see [Table 2](#)). The deep learning methods of CNN and TB-LSTM were the best performers overall. Their sensitivity (0.97 and 1, respectively) showed that these models were able to detect nearly all the cases of “good-quality sleep,” meaning that in a tool for screening potential sleep problems these models will be able to detect easily people with normal sleep quality. Often high sensitivity comes at the price of low specificity (ie, failing to identify negative cases, or true negative error). This was the case of logistic regression, which had a high sensitivity but a specificity of 0.3, meaning that in such models many “poor sleeps” would have been wrongly classified as good sleep. This is very important, since misidentifying poor sleep cases can lead to underdiagnosis of problematic sleep.

Table 2. Sensitivity and specificity results.

Data models	Sensitivity or recall	Specificity
LR ^a	0.9714	0.333
MLP ^b	0.8857	0.9048
CNN ^c	0.9714	0.8571
RNN ^d	0.9143	0.2381
LSTM ^e	0.9714	0.4762
TB-LSTM ^f	1.000	0.7143

^aLR: logistic regression.

^bMLP: multilayer perceptrons.

^cCNN: convolutional neural networks.

^dRNN: recurrent neural networks.

^eLSTM: long short-term memory.

^fTB-LSTM: time-batched LSTM.

The sensitivity (also known as recall) and specificity of each of the models are reported in [Table 2](#). The high sensitivity values of each of the models indicate that deep learning has a strong capability of correctly identifying individuals with good sleep patterns from their preceding awake activity. The specificity is high for TB-LSTM, MLP, and CNN, indicating that these

models were also able to successfully distinguish those with poor sleep patterns.

Relevance of Findings

Previous algorithms, such as by Sadeh et al [27,37], do not focus on the prediction of sleep quality based on physical activity

during awake time, but rather on prediction of sleep quality based on accelerometer data during the sleep periods. The objective of those studies was the validation of the actigraphy data compared with PSG. Consequently, the objectives and findings of those studies cannot be compared with our study.

There are several data analytics areas relevant to our work. To our knowledge, our research was the first one looking into the use of deep learning for the study of actigraphy data related to physical activity and sleep. Previous research on the use of deep learning for sleep science has been focused on PSG data [45,46]. In other application areas, deep learning has been used for human activity recognition [47,48] which is a similar technical problem. In a previous study, we combined human recognition of actigraphy data with other machine learning algorithms, but not deep learning [49].

Impact in Sleep Science

Sleep insufficiency is highly prevalent in contemporary society, and has been shown to influence energy balance by altering metabolic hormone regulation. Consequently, health researchers are exploring the impact of sleep and physical activity on many health conditions. A major bottleneck for this research is that current approaches for studying actigraphy data require intensive manual work by human experts. Furthermore, huge datasets of actigraphy data are emerging from health research, including the study of sleep disorder patients, healthy populations, and epidemiological studies. Furthermore, millions of consumers are buying wearables that incorporate activity sensors. This burst of human activity data is a great opportunity for health research, but to achieve this paradigm shift, it is necessary to develop new algorithms and tools to analyze this type of data.

As explained in the results, our findings supported the feasibility of using physical activity data to predict the quality of sleep in terms of sleep efficiency. These findings were by no means aiming to substitute well-studied algorithms and methods for studying sleep and physical activity data, such as the methods described by Sadeh [37]. Improved algorithms, such as the ones we presented in this study, for actigraphy analysis can lead to a paradigm shift in the study of lifestyle behaviors such as sleep and physical activity, just as electrocardiography became crucial for cardiology and clinical research.

Our research showed that deep learning performed better than classical methods in terms of learning useful patterns from raw accelerometer data for the task of sleep quality prediction. Since deep learning models compute abstract features from raw input signals while optimizing on the actual sleep quality, this process yields a more robust solution. Furthermore, the good results of deep learning showed that raw accelerometer data had more “signal” regarding sleep quality, which traditional models such as logistic regression are not able to capture right now. More research needs to be done to understand why deep learning performs better, which eventually can help in identifying new factors influencing the quality of sleep.

Impact in eHealth

Our study provided an early example of how advanced deep learning methods could be used to infer new insights from raw actigraphy data. Our focus on predicting and forecasting can

help design new eHealth applications where predictions are made to personalize coaching for patients or to facilitate decision making of professionals.

There is an increased interest in sleep in the health domain. This is consequently being reflected in an increase use of eHealth for sleep [50-54] and also in the use of social media for sharing sleep logs [53]. The expansion of eHealth into sleep is not limited to sleep disorders, but also to improve sleep for people living with chronic conditions such as cancer [53]. These developments are closely related to the concept of Quantified Self for health [55,56]. Furthermore, we can assume that predicting sleep quality based on physical activity data acquired by accelerometer data (both from actigraphy or activity trackers) can be used to provide personalized feedback, such as momentary ecological interventions based on mobile technology [57].

In our study, we attempted to predict a parameter regarding the quality of sleep solely relying on the physical activity during the awake time. To our knowledge, this was not done earlier. The advantage of this approach is that eventually the same approach can be used to predict sleep quality with data from smart watches and other wearable devices that are not necessarily used during sleep. Therefore, our models can eventually be used within eHealth applications that do not require wearing a sensor during sleep. This is of special interest for the development of smart watch health apps [33], as they might require frequent battery charging.

Our research has many limitations as explained in the next subsection. However, this is the first study, to our knowledge, that focused on the prediction of sleep quality from physical activity accelerometer data during awake periods. Our methodology and results can be used as the baseline for further studies looking into predicting sleep quality from mobile and wearable devices. This is a source of major concern, since many sleep apps in the making predict with unclear methodology and performance [50-52]. Although more studies are highlighting the increasing reliability of consumer sleep wearables [31], we do not know how they calculate or predict sleep quality parameters. To maximize the potential of the wearable mass adoption for sleep health, we need research on not only the reliability of consumer-grade devices but also their data processing and modeling techniques.

Limitations

There were some limitations in our study regarding the generalization of our results. Sleep behavior can be affected by cultural aspects and also change with age. Our study sample drew sleep data from adolescents, aged 10-17 years, living in Qatar. Future research will need to evaluate whether applications of deep learning for sleep research using actigraphy will yield similar results in different populations (eg, adults and people with chronic conditions).

In our study, the prediction was simplified to “good” and “poor” sleep quality with regards to sleep efficiency. This may be an oversimplification of complex sleep problems. To provide more precise predictions (eg, quantitative value of sleep quality), these techniques will need to be validated. A prediction, such

as the one presented in this study, might be useful for the detection of people with unhealthy sleep patterns, but not to identify the causes of poor sleep efficiency.

There is also a limitation in the interpretation of deep learning. Deep learning models are “black boxes” and do not provide explanation of their sleep efficiency prediction. Other techniques such as logistic regression can provide insights on which features contribute to the prediction. However, this study showed that the performance of such models was much lower than that of deep learning. New techniques in deep learning are being researched to facilitate the interpretation of such models.

In our study, the models used the data from the individual’s awake time to predict of sleep quality. The prediction was made at the last moment before sleep using the full awake time activity. If these models were to be used to provide personalized feedback to individuals with sleep problems, they will need to

be tested with fragments of the awake time, giving an individual time to alter their behavior. Since our data was fragmented into sleep periods and awake times specific to an individual, the models would be able to handle varying durations of awake time.

Conclusions

Our study showed the feasibility of deep learning in predicting sleep efficiency using wearable data from awake periods. This is of paramount importance because deep learning eliminated the need for data preprocessing and simplified the overall workflow in sleep data research. The feasibility of our approach can lead to new applications in sleep science and also to the development of more complex eHealth sleep applications for both professionals and patients. These models can also be integrated in the broader context of quantified self [55,56].

Acknowledgments

ST and TA were funded by the Biomedical Research Program at Weill Cornell Medicine in Qatar, supported by Qatar Foundation. The authors wish to thank the clinical research core at Weill Cornell Medicine in Qatar for their support in conducting the study, and the schools, teaching support staff, students, and parents for their participation.

Conflicts of Interest

None declared.

Authors' Contributions

AS was involved in all the research and manuscript preparation. SJ and FO advised AS on the technical aspects of the study and participated in manuscript preparation. LFL, AE, and JS advised AS on the study design and participated in manuscript preparation. LFL is the Principal Investigator of the eHealth project 360QS [58] at QCRI. ST and TA participated as health researchers in the study design, data collection, and the institutional review board application, and reviewed and approved the manuscript. ST is the Principal Investigator of the clinical study.

References

1. Strine TW, Chapman DP. Associations of frequent sleep insufficiency with health-related quality of life and health behaviors. *Sleep Med* 2005 Jan;6(1):23-27. [doi: [10.1016/j.sleep.2004.06.003](https://doi.org/10.1016/j.sleep.2004.06.003)] [Medline: [15680291](https://pubmed.ncbi.nlm.nih.gov/15680291/)]
2. Colten HR, Altevogt BR, 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; 2006.
3. Knutson KL, Ryden AM, Mander BA, Van CE. Role of sleep duration and quality in the risk and severity of type 2 diabetes mellitus. *Arch Intern Med* 2006 Sep 18;166(16):1768-1774. [doi: [10.1001/archinte.166.16.1768](https://doi.org/10.1001/archinte.166.16.1768)] [Medline: [16983057](https://pubmed.ncbi.nlm.nih.gov/17000571/)]
4. Gottlieb DJ, Punjabi NM, Newman AB, Resnick HE, Redline S, Baldwin CM, et al. Association of sleep time with diabetes mellitus and impaired glucose tolerance. *Arch Intern Med* 2005 Apr 25;165(8):863-867. [doi: [10.1001/archinte.165.8.863](https://doi.org/10.1001/archinte.165.8.863)] [Medline: [15851636](https://pubmed.ncbi.nlm.nih.gov/15851636/)]
5. Nilsson PM, Rööst M, Engström G, Hedblad B, Berglund G. Incidence of diabetes in middle-aged men is related to sleep disturbances. *Diabetes Care* 2004 Oct;27(10):2464-2469. [Medline: [15451917](https://pubmed.ncbi.nlm.nih.gov/15451917/)]
6. Palagini L, Bruno RM, Gemignani A, Baglioni C, Ghiaodoni L, Riemann D. Sleep loss and hypertension: a systematic review. *Curr Pharm Des* 2013;19(13):2409-2419. [Medline: [23173590](https://pubmed.ncbi.nlm.nih.gov/23173590/)]
7. Kasabeh E, Chi DS, Krishnaswamy G. Inflammatory aspects of sleep apnea and their cardiovascular consequences. *South Med J* 2006 Jan;99(1):58-67; quiz 68-9. [doi: [10.1097/01.smj.0000197705.99639.50](https://doi.org/10.1097/01.smj.0000197705.99639.50)] [Medline: [16466124](https://pubmed.ncbi.nlm.nih.gov/16466124/)]
8. Meier-Ewert HK, Ridker PM, Rifai N, Regan MM, Price NJ, Dinges DF, et al. Effect of sleep loss on C-reactive protein, an inflammatory marker of cardiovascular risk. *J Am Coll Cardiol* 2004 Feb 18;43(4):678-683 [FREE Full text] [doi: [10.1016/j.jacc.2003.07.050](https://doi.org/10.1016/j.jacc.2003.07.050)] [Medline: [14975482](https://pubmed.ncbi.nlm.nih.gov/14975482/)]
9. Cohen S, Doyle WJ, Alper CM, Janicki-Deverts D, Turner RB. Sleep habits and susceptibility to the common cold. *Arch Intern Med* 2009 Jan 12;169(1):62-67 [FREE Full text] [doi: [10.1001/archinternmed.2008.505](https://doi.org/10.1001/archinternmed.2008.505)] [Medline: [19139325](https://pubmed.ncbi.nlm.nih.gov/19139325/)]
10. Opp MR, Toth LA. Neural-immune interactions in the regulation of sleep. *Front Biosci* 2003 May 01;8:d768-d779. [Medline: [12700057](https://pubmed.ncbi.nlm.nih.gov/12700057/)]

11. Peterman JS, Carper MM, Kendall PC. Anxiety disorders and comorbid sleep problems in school-aged youth: review and future research directions. *Child Psychiatry Hum Dev* 2015 Jun;46(3):376-392. [doi: [10.1007/s10578-014-0478-y](https://doi.org/10.1007/s10578-014-0478-y)] [Medline: [24962165](#)]
12. Murphy MJ, Peterson MJ. Sleep disturbances in depression. *Sleep Med Clin* 2015 Mar;10(1):17-23. [doi: [10.1016/j.jsmc.2014.11.009](https://doi.org/10.1016/j.jsmc.2014.11.009)] [Medline: [26055669](#)]
13. Williamson AM, Feyer AM. Moderate sleep deprivation produces impairments in cognitive and motor performance equivalent to legally prescribed levels of alcohol intoxication. *Occup Environ Med* 2000 Oct;57(10):649-655 [FREE Full text] [Medline: [10984335](#)]
14. Van Dongen HP, Maislin G, Mullington JM, Dinges DF. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep* 2003 Mar 15;26(2):117-126. [Medline: [12683469](#)]
15. Ellenbogen JM, Payne JD, Stickgold R. The role of sleep in declarative memory consolidation: passive, permissive, active or none? *Curr Opin Neurobiol* 2006 Dec;16(6):716-722. [doi: [10.1016/j.conb.2006.10.006](https://doi.org/10.1016/j.conb.2006.10.006)] [Medline: [17085038](#)]
16. Taheri S, Lin L, Austin D, Young T, Mignot E. Short sleep duration is associated with reduced leptin, elevated ghrelin, and increased body mass index. *PLoS Med* 2004 Dec;1(3):e62 [FREE Full text] [doi: [10.1371/journal.pmed.0010062](https://doi.org/10.1371/journal.pmed.0010062)] [Medline: [15602591](#)]
17. Arora T, Taheri S. Associations among late chronotype, body mass index and dietary behaviors in young adolescents. *Int J Obes (Lond)* 2015 Jan;39(1):39-44. [doi: [10.1038/ijo.2014.157](https://doi.org/10.1038/ijo.2014.157)] [Medline: [25135376](#)]
18. Arora T, Hosseini-Araghi M, Bishop J, Yao GL, Thomas GN, Taheri S. The complexity of obesity in U.K. adolescents: relationships with quantity and type of technology, sleep duration and quality, academic performance and aspiration. *Pediatr Obes* 2013 Oct;8(5):358-366. [doi: [10.1111/j.2047-6310.2012.00119.x](https://doi.org/10.1111/j.2047-6310.2012.00119.x)] [Medline: [23239604](#)]
19. Paruthi S, Brooks LJ, D'Ambrosio C, Hall WA, Kotagal S, Lloyd RM, et al. Recommended amount of sleep for pediatric populations: a consensus statement of the American Academy of Sleep Medicine. *J Clin Sleep Med* 2016 Jun 15;12(6):785-786. [doi: [10.5664/jcsm.5866](https://doi.org/10.5664/jcsm.5866)] [Medline: [27250809](#)]
20. Kredlow MA, Capozzoli MC, Hearon BA, Calkins AW, Otto MW. The effects of physical activity on sleep: a meta-analytic review. *J Behav Med* 2015 Jun;38(3):427-449. [doi: [10.1007/s10865-015-9617-6](https://doi.org/10.1007/s10865-015-9617-6)] [Medline: [25596964](#)]
21. Driver HS, Taylor SR. Exercise and sleep. *Sleep Med Rev* 2000 Aug;4(4):387-402. [doi: [10.1053/smrv.2000.0110](https://doi.org/10.1053/smrv.2000.0110)] [Medline: [12531177](#)]
22. Chennaoui M, Arnal PJ, Sauvet F, Léger D. Sleep and exercise: a reciprocal issue? *Sleep Med Rev* 2015 Apr;20:59-72. [doi: [10.1016/j.smrv.2014.06.008](https://doi.org/10.1016/j.smrv.2014.06.008)] [Medline: [25127157](#)]
23. Morgenthaler T, Alessi C, Friedman L, Owens J, Kapur V, Boehlecke B, Standards of Practice Committee, American Academy of Sleep Medicine. Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: an update for 2007. *Sleep* 2007 Apr;30(4):519-529. [Medline: [17520797](#)]
24. Hirshkowitz M. The history of polysomnography: tool of scientific discovery. In: Chokroverty S, Billiard M, editors. *Sleep Medicine: A Comprehensive Guide to Its Development, Clinical Milestones, and Advances in Treatment*. New York, USA: Springer; 2015:91-100.
25. Bruyneel M, Van den Broecke S, Libert W, Ninane V. Real-time attended home-polysomnography with telematic data transmission. *Int J Med Inform* 2013 Aug;82(8):696-701. [doi: [10.1016/j.ijimedinf.2013.02.008](https://doi.org/10.1016/j.ijimedinf.2013.02.008)] [Medline: [23529100](#)]
26. Masa JF, Corral J, Pereira R, Duran-Cantolla J, Cabello M, Hernández-Blasco L, et al. Therapeutic decision-making for sleep apnea and hypopnea syndrome using home respiratory polygraphy: a large multicentric study. *Am J Respir Crit Care Med* 2011 Oct 15;184(8):964-971. [doi: [10.1164/rccm.201103-0428OC](https://doi.org/10.1164/rccm.201103-0428OC)] [Medline: [21737584](#)]
27. Sadeh A. The role and validity of actigraphy in Sleep Medicine: an update. *Sleep Med Rev* 2011 Aug;15(4):259-267. [doi: [10.1016/j.smrv.2010.10.001](https://doi.org/10.1016/j.smrv.2010.10.001)] [Medline: [21237680](#)]
28. Arora T, Broglia E, Pushpakumar D, Lodhi T, Taheri S. An investigation into the strength of the association and agreement levels between subjective and objective sleep duration in adolescents. *PLoS One* 2013;8(8):e72406 [FREE Full text] [doi: [10.1371/journal.pone.0072406](https://doi.org/10.1371/journal.pone.0072406)] [Medline: [23951321](#)]
29. Lee I, Shiroma EJ. Using accelerometers to measure physical activity in large-scale epidemiological studies: issues and challenges. *Br J Sports Med* 2014 Feb;48(3):197-201 [FREE Full text] [doi: [10.1136/bjsports-2013-093154](https://doi.org/10.1136/bjsports-2013-093154)] [Medline: [24297837](#)]
30. Ehsan M, Khan R, Wakefield D, Qureshi A, Murray L, Zuwallack R, et al. A longitudinal study evaluating the effect of exacerbations on physical activity in patients with chronic obstructive pulmonary disease. *Ann Am Thorac Soc* 2013 Dec;10(6):559-564. [doi: [10.1513/AnnalsATS.201304-100OC](https://doi.org/10.1513/AnnalsATS.201304-100OC)] [Medline: [24053416](#)]
31. Montgomery-Downs HE, Insana SP, Bond JA. Movement toward a novel activity monitoring device. *Sleep Breath* 2012 Sep;16(3):913-917. [doi: [10.1007/s11325-011-0585-y](https://doi.org/10.1007/s11325-011-0585-y)] [Medline: [21971963](#)]
32. Poirier J, Bennett WL, Jerome GJ, Shah NG, Lazo M, Yeh H, et al. Effectiveness of an activity tracker- and Internet-based adaptive walking program for adults: a randomized controlled trial. *J Med Internet Res* 2016;18(2):e34 [FREE Full text] [doi: [10.2196/jmir.5295](https://doi.org/10.2196/jmir.5295)] [Medline: [26860434](#)]
33. Lu T, Fu C, Ma MH, Fang C, Turner AM. Healthcare applications of smart watches: a systematic review. *Appl Clin Inform* 2016 Sep 14;7(3):850-869. [doi: [10.4338/ACI-2016-03-R-0042](https://doi.org/10.4338/ACI-2016-03-R-0042)] [Medline: [27623763](#)]

34. Wu W, Dasgupta S, Ramirez EE, Peterson C, Norman GJ. Classification accuracies of physical activities using smartphone motion sensors. *J Med Internet Res* 2012;14(5):e130 [FREE Full text] [doi: [10.2196/jmir.2208](https://doi.org/10.2196/jmir.2208)] [Medline: [23041431](https://pubmed.ncbi.nlm.nih.gov/23041431/)]
35. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc* 1998 May;30(5):777-781. [Medline: [9588623](https://pubmed.ncbi.nlm.nih.gov/9588623/)]
36. Grigg-Damberger MM. The AASM Scoring Manual four years later. *J Clin Sleep Med* 2012 Jun 15;8(3):323-332 [FREE Full text] [doi: [10.5664/jcsm.1928](https://doi.org/10.5664/jcsm.1928)] [Medline: [22701392](https://pubmed.ncbi.nlm.nih.gov/22701392/)]
37. Sadeh A, Raviv A, Gruber R. Sleep patterns and sleep disruptions in school-age children. *Dev Psychol* 2000 May;36(3):291-301. [Medline: [10830974](https://pubmed.ncbi.nlm.nih.gov/10830974/)]
38. Reed DL, Sacco WP. Measuring sleep efficiency: what should the denominator be? *J Clin Sleep Med* 2016 Feb;12(2):263-266 [FREE Full text] [doi: [10.5664/jcsm.5498](https://doi.org/10.5664/jcsm.5498)] [Medline: [26194727](https://pubmed.ncbi.nlm.nih.gov/26194727/)]
39. Kim Y, Lee J, Peters BP, Gaesser GA, Welk GJ. Examination of different accelerometer cut-points for assessing sedentary behaviors in children. *PLoS One* 2014;9(4):e90630 [FREE Full text] [doi: [10.1371/journal.pone.0090630](https://doi.org/10.1371/journal.pone.0090630)] [Medline: [24699259](https://pubmed.ncbi.nlm.nih.gov/24699259/)]
40. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. *Med Sci Sports Exerc* 2011 Feb;43(2):357-364 [FREE Full text] [doi: [10.1249/MSS.0b013e3181ed61a3](https://doi.org/10.1249/MSS.0b013e3181ed61a3)] [Medline: [20581716](https://pubmed.ncbi.nlm.nih.gov/20581716/)]
41. Tieleman T, Hinton G. COURSERA: Neural Networks for Machine Learning. YouTube. 2012. URL: <https://www.youtube.com/watch?v=LGA-gRkLEsI> [accessed 2016-10-27] [WebCite Cache ID [6lZHtYQZN](https://webcache.googleusercontent.com/search?q=cache:6lZHtYQZN)]
42. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014 Jan. URL: <https://www.cs.toronto.edu/~hinton/absps/JMLRdropout.pdf> [accessed 2016-10-28] [WebCite Cache ID [6lacTnQRx](https://webcache.googleusercontent.com/search?q=cache:6lacTnQRx)]
43. Roeser K, Eichholz R, Schwerdtle B, Schlarb AA, Kübler A. Relationship of sleep quality and health-related quality of life in adolescents according to self- and proxy ratings: a questionnaire survey. *Front Psychiatry* 2012;3:76 [FREE Full text] [doi: [10.3389/fpsyg.2012.00076](https://doi.org/10.3389/fpsyg.2012.00076)] [Medline: [22969731](https://pubmed.ncbi.nlm.nih.gov/22969731/)]
44. Gruber R, Somerville G, Enros P, Paquin S, Kestler M, Gillies-Poitras E. Sleep efficiency (but not sleep duration) of healthy school-age children is associated with grades in math and languages. *Sleep Med* 2014 Dec;15(12):1517-1525. [doi: [10.1016/j.sleep.2014.08.009](https://doi.org/10.1016/j.sleep.2014.08.009)] [Medline: [25441747](https://pubmed.ncbi.nlm.nih.gov/25441747/)]
45. Shmiel O, Shmiel T, Dagan Y, Teicher M. Data mining techniques for detection of sleep arousals. *J Neurosci Methods* 2009 May 15;179(2):331-337. [doi: [10.1016/j.jneumeth.2009.01.035](https://doi.org/10.1016/j.jneumeth.2009.01.035)] [Medline: [19428545](https://pubmed.ncbi.nlm.nih.gov/19428545/)]
46. Längkvist M, Karlsson L, Loutfi A. Sleep stage classification using unsupervised feature learning. *Adv Artif Neural Syst* 2012;1-9. [doi: [10.1155/2012/107046](https://doi.org/10.1155/2012/107046)]
47. Wang L, Qiao Y, Tang X. Action recognition with trajectory- pooled deep-convolutional descriptors. Presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June, 2015; Boston, MA, USA p. 4305-4314. [doi: [10.1109/CVPR.2015.7299059](https://doi.org/10.1109/CVPR.2015.7299059)]
48. Sun L, Jia K, Yeung Y, Shi H. Human action recognition using factorized spatio-temporal convolutional networks. Presented at: IEEE International Conference on Computer Vision (ICCV). IEEE; December, 2015; Santiago, Chile p. 4597-4605. [doi: [10.1109/ICCV.2015.522](https://doi.org/10.1109/ICCV.2015.522)]
49. Sathyaranayana, A, Ofli F, Fernandez-Luque L, Srivastava J, Elmagarmid A, Arora T, et al. Robust Automated Human Activity Recognition and its Application to Sleep Research. Presented at: IEEE International Conference on Data Mining (ICDM), Data Mining in Human Activity Analysis Workshop; 2016; Barcelona, Spain URL: <https://arxiv.org/abs/1607.04867>
50. Behar J, Roebuck A, Domingos JS, Gederi E, Clifford GD. A review of current sleep screening applications for smartphones. *Physiol Meas* 2013 Jul;34(7):R29-R46. [doi: [10.1088/0967-3334/34/7/R29](https://doi.org/10.1088/0967-3334/34/7/R29)] [Medline: [23771214](https://pubmed.ncbi.nlm.nih.gov/23771214/)]
51. de ZM, Baker FC, Colrain IM. Validation of sleep-tracking technology compared with polysomnography in adolescents. *Sleep* 2015 Sep 01;38(9):1461-1468 [FREE Full text] [doi: [10.5665/sleep.4990](https://doi.org/10.5665/sleep.4990)] [Medline: [26158896](https://pubmed.ncbi.nlm.nih.gov/26158896/)]
52. Bhat S, Ferraris A, Gupta D, Mozafarian M, DeBari VA, Gushway-Henry N, et al. Is there a clinical role for smartphone sleep apps? Comparison of sleep cycle detection by a smartphone application to polysomnography. *J Clin Sleep Med* 2015 Jul 15;11(7):709-715 [FREE Full text] [doi: [10.5664/jcsm.4840](https://doi.org/10.5664/jcsm.4840)] [Medline: [25766719](https://pubmed.ncbi.nlm.nih.gov/25766719/)]
53. Min YH, Lee JW, Shin Y, Jo M, Sohn G, Lee J, et al. Daily collection of self-reporting sleep disturbance data via a smartphone app in breast cancer patients receiving chemotherapy: a feasibility study. *J Med Internet Res* 2014;16(5):e135 [FREE Full text] [doi: [10.2196/jmir.3421](https://doi.org/10.2196/jmir.3421)] [Medline: [24860070](https://pubmed.ncbi.nlm.nih.gov/24860070/)]
54. McIver DJ, Hawkins JB, Chunara R, Chatterjee AK, Bhandari A, Fitzgerald TP, et al. Characterizing sleep issues using Twitter. *J Med Internet Res* 2015 Jun 08;17(6):e140 [FREE Full text] [doi: [10.2196/jmir.4476](https://doi.org/10.2196/jmir.4476)] [Medline: [26054530](https://pubmed.ncbi.nlm.nih.gov/26054530/)]
55. Almalki M, Gray K, Martin-Sanchez F. Activity theory as a theoretical framework for health self-quantification: a systematic review of empirical studies. *J Med Internet Res* 2016 May 27;18(5):e131 [FREE Full text] [doi: [10.2196/jmir.5000](https://doi.org/10.2196/jmir.5000)] [Medline: [27234343](https://pubmed.ncbi.nlm.nih.gov/27234343/)]
56. Paton C, Hansen M, Fernandez-Luque L, Lau AY. Self-tracking, social media and personal health records for patient empowered self-care: contribution of the IMIA Social Media Working Group. *Yearb Med Inform* 2012;7:16-24. [Medline: [22890336](https://pubmed.ncbi.nlm.nih.gov/22890336/)]

57. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [[FREE Full text](#)] [doi: [10.1348/135910709X466063](https://doi.org/10.1348/135910709X466063)] [Medline: [19646331](#)]
58. Haddadi H, Ofli F, Mejova Y, Weber I, Srivastava J. 360 Quantified Self. Presented at: IEEE International Conference on Healthcare Informatics (ICHI); 2015; Dalla, TX, USA p. 587-592. [doi: [10.1109/ICHI.2015.95](https://doi.org/10.1109/ICHI.2015.95)]

Abbreviations

AUC: Area under ROC curve
CNN: convolutional neural network
LR: logistic regression
LSTM-RNN: long short-term memory RNN
MLP: multilayer perceptron
PSG: polysomnography
ROC: receiver operating characteristic
RNN: recurrent neural network
SpO₂: oxygen saturation
TB-LSTM: time-batched version of LSTM-RNN

Edited by G Eysenbach; submitted 30.08.16; peer-reviewed by N Ridgers, YS Bin; comments to author 20.09.16; revised version received 04.10.16; accepted 22.10.16; published 04.11.16

Please cite as:

Sathyaranayana 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(4):e125

URL: <http://mhealth.jmir.org/2016/4/e125/>

doi: [10.2196/mhealth.6562](https://doi.org/10.2196/mhealth.6562)

PMID: [27815231](#)

©Aarti Sathyaranayana, Shafiq Joty, Luis Fernandez-Luque, Ferda Ofli, Jaideep Srivastava, Ahmed Elmagarmid, Teresa Arora, Shahrad Taheri. Originally published in JMIR Mhealth and Uhealth (<http://mhealth.jmir.org>), 04.11.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on <http://mhealth.jmir.org/>, as well as this copyright and license information must be included.