Manuscript Details

Manuscript number	PMC_2019_122
Title	Predicting future sedentary behaviour using wearable and mobile devices
Article type	Research Paper

Abstract

Sedentarism is a common problem that can affect human health and wellbeing. Predicting sedentary behaviour is an emerging area that can benefit from data collected from sensors available in ubiquitous devices, such as wearables and smartphones. In this paper, we present an approach aiming at predicting the sedentary behaviour of a user from data collected from sensors installed in wearable/mobile devices. We compare personal and impersonal models using a real-life dataset consisting of sensing data of 48 users during 10 weeks. We found that impersonal models using Deep Neural Networks were able to accurately predict the subject's future sedentary behaviour.

Keywords	Sedentary behaviour prediction; machine learning; user modelling; wearable and mobile devices

Corresponding Author	Marcelo Armentano
Corresponding Author's Institution	ISISTAN Research Institute (CONICET/UNICEN)
Order of Authors	Martín Santillán Cooper, Marcelo Armentano

Submission Files Included in this PDF

File Name [File Type]

cover_letter.docx [Cover Letter]

highlights.docx [Highlights]

manuscript.docx [Manuscript File]

To view all the submission files, including those not included in the PDF, click on the manuscript title on your EVISE Homepage, then click 'Download zip file'.

Research Data Related to this Submission

Data set http://studentlife.cs.dartmouth.edu/dataset.html

StudentLife Dataset

StudentLife is the first study that uses passive and automatic sensing data from the phones of a class of 48 Dartmouth students over a 10 week term to assess their mental health (e.g., depression, loneliness, stress), academic performance (grades across all their classes, term GPA and cumulative GPA) and behavioral trends (e.g., how stress, sleep, visits to the gym, etc. change in response to college workload -- i.e., assignments, midterms, finals -- as the term progresses).

March 25, 2018

Dear Editors,

We are pleased to submit an original research entitled "Predicting future sedentary behaviour using wearable and mobile devices", authored by Martín Santillán Cooper and Marcelo G. Armentano, to be considered for publication in the journal *Pervasive and Mobile Computing*.

We believe that this manuscript is appropriate for publication in this journal since we addressed the question of it is possible to predict the future sedentary behaviour of a subject, based on the observation of values obtained from multiple sensors of wearable and mobile devices. To validate our hypothesis, we used a publicly available dataset consisting of sensing data of 48 users during 10 weeks. We evaluated both personal and impersonal models for predicting future sedentary behaviour according to the Metabolic Equivalent of Tasks (MET), which is a standard way of measuring the energy expenditure of an activity. At the best of our knowledge, the use of MET for predicting future sedentary behaviour remained unexplored. We believe that our findings would appeal to the readership of this special issue in particular and of the journal in general.

This manuscript has not been published and is not under consideration for publication elsewhere. All authors have approved the manuscript and agree with its submission to the *Pervasive and Mobile Computing* journal. We have no conflicts of interest to disclose.

Please let us know of your decision at your earliest convenience.
With my best regards,
Sincerely yours,

Marcelo G. Armentano
ISISTAN (CONICET/UNICEN)
Campus Universitario
Tandil, Argentina
email: marcelo.armentano@isistan.unicen.edu.ar

Highlights

- We present an approach aiming at predicting future sedentary behaviour of a user
- We used data collected from sensors installed in wearable/mobile devices
- We propose the use of MET for considering an activity to be sedentary
- We evaluated both personal and impersonal models using a real-word dataset
- Deep Neural Networks obtained better performance than Logistic Regression

Predicting future sedentary behaviour using wearable and mobile devices

Martín Santillán Cooper and Marcelo G. Armentano

ISISTAN Research Institute (CONICET/ UNICEN)

Campus Universitario. Paraje Arroyo Seco. Tandil (7000). Argentina.

mscooper@exa.unicen.edu.ar; marcelo.armentano@isistan.unicen.edu.ar

Tel: +54 0249 4385650 / Fax: +54 0249 4385681

Abstract

Sedentarism is a common problem that can affect human health and wellbeing. Predicting sedentary behaviour is an emerging area that can benefit from data collected from sensors available in ubiquitous devices, such as wearables and smartphones. In this paper, we present an approach aiming at predicting the sedentary behaviour of a user from data collected from sensors installed in wearable/mobile devices. We compare personal and impersonal models using a real-life dataset consisting of sensing data of 48 users during 10 weeks. We found that impersonal models using Deep Neural Networks were able to accurately predict the subject's future sedentary behaviour.

Keywords: Sedentary behaviour prediction; machine learning; user modelling; wearable and mobile devices

1 - Introduction

The definition of sedentary behaviour has been evolving over the years, at the same time so did the way of measuring it. The Sedentary Behaviour Research Network[1] defines *sedentary behaviour* as any waking behaviour characterized by an energy expenditure ≤1.5 METs (Metabolic Equivalent of Tasks) while in a sitting or reclining posture (Tremblay 2012). MET measures the intensity of an activity in multiples of resting energy expenditure. Examples of sedentary activities are watching television (1.0 MET), eating while sitting (1.5 MET) or playing video games (1.0 MET) and driving (1.3 MET).

Research done in the area demonstrates strong and consistent associations between sedentary time and diabetes, cardiovascular disease and all-cause mortality (Wilmot et al. 2012) (Carter et al. 2017). However, the reported associations were largely independent of physical activity. Thus, it is important to note that sedentary behaviour does not represent the opposite of physical activity and that it is possible for an individual to have high levels of both moderate to vigorous physical activity (MVPA) and sedentary behaviour. In general, the sedentary time has shown to be detrimentally associated with health outcomes and with markers of metabolic risk across diverse population groups. In addition, it has been highlighted the importance of not only stimulating MVPA but also reducing sedentary time, since sedentary behaviour is a risk factor for mortality independent of MVPA (Koster et al. 2012).

Other research line focused on the associations between short breaks in sedentary time with metabolic outcomes (Paing et al. 2018) and with optimization of cognitive operations (Felez-Nobrega et al. 2018)(Magnon, Vallet, and Auxiette 2018)(Falck, Davis, and Liu-Ambrose 2017). Benatti and Ried-Larsen (Benatti and Ried-Larsen 2015) claimed that there is enough evidence to show the positive effects of breaking up prolonged time spent sitting on metabolic outcomes.

[1] https://www.sedentarybehaviour.org/

Methods for measuring sedentary behaviour can be classified as subjective (self- and proxy-report questionnaires, diaries) and objective (common sensors of ubiquitous devices). Subjective methods are being surpassed by new technologies that can provide, for all population groups, second-by-second information on posture, movement (or lack of movement) and patterns within and between days. (Atkin et. al. 2012).

Although mobile devices can be considered as one of the causes of sedentary behaviour (He and Agu 2016b), they also offer new opportunities to prevent it. Nowadays, portable mobile devices, such as smartphones, smartwatches and fitness trackers are equipped with a wide variety of sensors that can be used for human activity and behaviour analysis. The use of objective methods to assess sedentary behaviour is growing in popularity as the costs portable mobile devices decrease and are easier to use. In 2018, 91% of people between 18 and 49 years old in the US own a smartphone[2]. In this context, these devices can be seen as an opportunity for developing complex objective methods for measuring sedentary behaviour.

Many smartphone applications have been implemented with the aim of alerting the user when sedentary behaviour is recognized (He and Agu 2014; Grundgeiger et al. 2017; Fahim et al. 2017). Predicting future sedentary behaviour can help to enable preventive interventions such as reminders and suggestion for different activities based on Theory of Planned Behaviour (TPB) (Ajzen, 1991). TPB postulate that a subject is more likely to participate in recommended interventions to reduce sedentary behaviours if such activities are included in their plans. Following this idea, our hypothesis is that if we were able to predict at time *t* that a subject will be sedentary in time *t+1,* we could recommend activities aiming at changing his/her sedentary routine in a long term. This kind of interventions can lead to better opportunities for changing the subjects' behaviour to healthier outcomes.

We define future sedentary behaviour prediction (FSBP) as predicting whether the physical activity of a user will surpass, on average, 1,5 METs in the near future or not. The

---

[2]http://www.pewresearch.org/fact-tank/2018/09/28/internet-social-media-use-and-device-ownership-in-u-s-have-plateaued-after-years-of-growth/

problem of predicting the future sedentary behaviour has been previously addressed by analyzing only the inactive/stationary time of a subject and the performance of several models has been measured and compared. However, although MET is a standard metric in the area of health for measuring the intensity of an activity in terms of energy expenditure, the use of this metric for predicting sedentary behaviour by using wearables and mobile devices remains unexplored. In this paper, we present a novel approach for predicting the future sedentary behaviour of a subject in terms of its MET level (Tremblay 2012). In addition, we present a comparison between different models for FSBP.

To sum up, the research question that we address in this article is whether it is possible to predict the future sedentary behaviour of a subject, based on the observation of values obtained from multiple sensors of wearable and mobile devices.

The remaining of this article is organized as follows. In Section 2 we present some related work in the use of wearable and mobile devices to predict human activity, focusing particularly on sedentary behaviour prediction. Section 3 defines Personal and Impersonal models and Section 4 describes the two predictive models used in this research. Section 5 describes the dataset used for our experiments, detailing the data cleaning procedure and the definition of features from the raw data. Furthermore, we analyze some sedentary behaviour that can be observed from the dataset. In Section 6, we present the experiments performed to evaluate the proposed models. Finally, in Section 7, we present our conclusions and future work.

2 - Related Work

As mobile and wearable devices, such as smartphones, became popular, many studies have been carried out to find usage patterns that allow correlating, inferring and predicting different types of human behaviours in the context of health and wellbeing. Harari et al (Harari et al. 2017) demonstrated the viability of using smartphone sensing methods to track health-related behaviours in the context of students' daily lives. They identified a number of significant correlations between objective sensing data from smartphones and mental wellbeing and academic performance outcomes. Gong et al (Gong et al. 2019), in a collaboration between engineers and psychologists, demonstrated a statistically significant relationship between smartphone sensing data and social anxiety levels of college students. From the experiments performed, the authors concluded that there is a significant difference in movement behaviours tied to social anxiety level and that these behaviours vary across different semantic locations. These results open new possibilities for passively monitoring behavioural markers of social anxiety through the integration of the accelerometer and the GPS data. Kanjo et al (2019) adopted a deep learning approach for emotion classification trained and tested using a dataset collected from smartphones and wearable devices in a real-world study. They achieved a classification accuracy of 95% using hybrid models composed by Convolutional Neural Network and Long Short-term Memory Recurrent Neural Network. Wu et al (Wu et al. 2018) used Bluetooth encounter networks to predict their cognitive stress levels. The results indicated a potential value of incorporating Bluetooth encounter data into mental health monitoring practice via mobile sensing technology. Zia et al (Zia et al. 2016) studied the viability of neural networks in predicting Freezing of Gait trained and tested with data collected from wearable devices. The aim of this research was to analyze if wearable devices could be used to help Parkinson's patients live safer and more independent lives. Particularly, they used a class of neural networks known as Layered Recurrent Networks and reached 42%, 89%, and 57% precision in predicting Freezing of Gait for the three participants in the study.

Activity Prediction is an area of research that consists of hypothesizing information about the activities that a subject will be involved in the near future. In this context, our research focused on predicting future sedentary behaviour. Several authors have addressed a similar problem. Q. He and E. Agu proposed several models for FSBP. First, they proposed a frequency domain algorithm for detecting recurrent sedentary patterns from activity time-series data. In their experiment, subjects who exhibited recurrent sedentary behaviours yielded periodic functions with a Mean Square Error as low as 0.003817 for predicting recurrent sedentary behaviours (He and Agu 2016a). In a second study, the same authors explore whether the contexts that can be sensed by users' smartphones can be used to predict their future sedentary behaviours reliably, in order to enable more effective interventions based on the theory of planned behaviour. By using logistic regression, they were able to classify user context variables to predict if the subject will be "very sedentary" in the next hour with a precision of 73.1% and a recall of 87.7% (He and Agu 2016b). It is important to notice that the reported figures correspond to the prediction of only the "very sedentary" class. Parallely, they proposed an approach to automatically discover patients' temporal patterns of sedentary behaviour from raw activity logs by using an autoregressive model with maximum entropy method. They have been able to model the three important individual determinants of sedentary behaviours: time, daily rhythm, and past sedentary habits (He and Agu 2016c). Finally, they focused on detecting the prevailing rhythms of sedentary behaviours and modelling the cyclical rhythm and linear rhythm in Lefebvre's philosophy using periodic functions (history-free) and linear functions (history-dependent) respectively (He and Agu 2017). In spite of the fact that these authors built several interesting models in their research, it is very important to note that they measured sedentary behaviour in terms of the average of sedentary records per hour. Instead, we follow the definition of the Sedentary Behaviour Research Network and we approximated the MET level thanks to the Compendium of physical activities (Ainsworth et al. 2011). In addition, a comparison between personal and impersonal models (in terms of the data used to train the models) has been never done FSBP.

3 - Personal and Impersonal models

In this paper, we aimed at comparing personal and impersonal models for FSBP, taking into account the advantages and disadvantages of each of them. Lockhart and Weiss (Lockhart and Weiss 2014), define impersonal models as those that "*use training data from a panel of users that will not subsequently use the model*". This definition implies that the training and test sets have no users in common. On the other hand, personal models are trained only with information taken from the same user for whom the model is intended. Therefore, personal models require a training phase to collect labeled data from each user. The training and test data come from the same user but contain different and disjoint examples.

Both, personal and impersonal models have its own advantages and disadvantages. On the one hand, impersonal models have the advantage that they can include data from many to train the model. In consequence, it is possible to train impersonal models with large amounts of data collected from other users and complex non-linear non-parametric functions can be searched for FSBP. Additionally, impersonal models can be built once for all users, as they can be considered universal. Finally, impersonal models have the advantage of not suffering the problem of cold-start (Schein et al. 2002). The cold-start problem refers to the situation in which a new user is added to a system and there is almost no information about him. Impersonal models can be used to generate predictions to a new user without any additional labelled training data or model regeneration. Among the disadvantages of impersonal models, we can highlight that building a model with data from many users and using it to classify activities of a target user may be prone to introduce noise due to the diversity among users. In consequence, users with particular behaviour patterns should have a better performance with personal models than with impersonal models.

On the other hand, personal models have the advantage of matching the idiosyncrasies of each user, at the cost of requiring each user to provide training (labelled) data. However, personal models suffer from the cold-start problem since the amount of data available for new users is scarce.

4 - Predictive Models

As we defined in Section 1, we define FSBP as the problem of predicting whether the physical activity of a user will surpass, on average, 1.5 METs in the near future. Then, if we consider that MET levels lower to 1.5 correspond to sedentary behaviour and MET levels higher than 1.5 correspond to non-sedentary behaviour, we are in presence of a binary classification problem.

Classification models can be divided into *discriminative* models and *generative* models (Bishop 2006). Given a set of data points$\{x\,(1), \ldots, x(m)\}$ associated to a set of outcomes $\{y\,(1), \ldots, y(m)\}$, discriminative models estimates $P(Y|\,X)$while generative models estimate $P(X|\,Y)$and $P(Y)$ to compute the posterior probability $P(Y|\,X)$. In this paper, we analyze the performance of discriminative models. Particularly, we use Logistic Regression and Deep Neural Networks for analyzing whether the user will exhibit sedentary behaviour in the next hour, given a set of features extracted from mobile devices' sensors. We chose these two machine learning techniques in order to compare the performance between a linear and a non-linear approach.

Both Logistic Regression and Deep Neural Networks classifiers share similar characteristics. The main problem they approach is to fit a function f that maps an input vector x to a category y. Formally, this mapping can be seen as a function $f(x;\theta) = y$, where θ is the set of parameters of the model, x is the vector of features extracted from the mobile devices' sensors and $y \in \{0;1\}$, where 0 corresponds to non-sedentary behaviour and 1 corresponds to sedentary behaviour. The parameters are learned by minimizing a loss function. The method that minimizes the lost function is called optimizer. In general, and albeit having many variations, the optimizer uses gradient descent to search the optimal value of the parameters that minimize the loss function.

Logistic Regression (LR) is a linear classifier that uses the sigmoid function as its hypothesis. The logistic function is defined as $g(z) = \frac{1}{1 + e^{-z}}$ and has the property that its co-domain goes from 0 to 1, having a horizontal asymptote in both values. Thus, the output of the sigmoid function can be seen as the confidence that the classifier has on that the example corresponds to a particular class. For example, the closer to 1 the output, the greater the confidence of the classifier in the fact that the input corresponds to the 1 class. Logistic regression is a linear classifier since the input of the predicted log-odds is a linear function of x. The advantage of linear models is that they can be fit e□ciently and reliably, but the capacity of the model is limited to linear functions, so the model cannot understand the interaction between any two input variables.

Deep Neural Network (DNN) is an algorithm that belongs to the Deep Learning field. DNNs can learn multiple levels of representation in order to model complex relationships among data. In them, higher-level features and concepts are defined in terms of lower-level ones. DNN is a type of model composed of a sequence of layers (Goodfellow, Bengio, and Courville 2016). At the same time, each layer is composed of a group of neurons, also called units. The first layer in a DNN is the *input layer* and corresponds to the x feature vector. The input layer is followed by one or more hidden layers. The last layer in a DNN is the *output layer*. Each hidden unit computes the dot product between the output of the previous layers and its corresponding weights -i.e the parameters of the DNN-. Then, an activation function is applied to each unit to introduce non-linear complexities to the model. In comparison to linear models, DNNs in general and especially those used in this research, have many parameters that have to be learned. To learn those parameters, DNNs use backpropagation with different optimization techniques. Another difference between DNNs and linear models is that DNNs have some hyperparameters that have to be tuned in order to maximize the performance of the model. Some of the decisions that have to be made are: how many layers the network should contain, how these layers should be connected to each other, how many units should exist in each layer, which activation function should use each unit, which optimizer should be used, how many epochs should be used for training, which regularization technique should be used, etc.

After finishing the process of tuning the DNN used in this work, we obtained a network with 4 hidden layers architecture composed by 256x128x64x32 units, respectively. As the activation function, we used Rectified Linear Units (ReLu), which uses the activation function $g(z) = max\{0, z\}$. To prevent overfitting, each hidden layer used Batch

Normalization (Ioffe and Szegedy 2015) and $L2$ regularization. $L2$, also known as weight decay, drives the weights closer to the origin by adding the sum of the square of each parameter of the model to the loss function. The output layer has only one neuron with a sigmoid activation function. We used binary cross-entropy as the loss function and Adaptive Moment Estimation (Adam) (Kingma and Ba 2015) as the optimization method. Each DNN model was fitted with 30 epochs and a batch size of 512, in the case of impersonal models, and a batch size of 32, in the case of personal models.

All DNN architectures tested were designed and implemented using the deep learning library Keras[3] for Python. The Logistic Regression implementation used in our experiments was the one provided by Scikit-Learn[4], a machine learning library implemented in Python.

5 - Dataset analysis and processing

In this section, we describe the dataset used to validate our models, the pre-processing applied to the raw data, the features extracted and some considerations about the hours of the sleep of the subjects

5.1 - Dataset description

We analyzed and preprocessed the StudentLife Dataset (Wang et al. 2014) in order to prove the validity of our models. The dataset was collected from 30 undergrads and 18 graduate students over a 10-week term in spring 2013. The students consisted of 38 males and 10 females. Two of them were first-year, 14 second-year, 6 third-year, and eight fourth-year Bachelor's students. There were also 13 first-year and 1 second-year Master's student, and

[3] https://keras.io/
[4] https://scikit-learn.org/stable/

3 PhD students. Participants were racially diverse, with 23 Caucasians, 23 Asians, and 2 African-Americans. Available data included:

1. **Activity data**, including activity duration (total time that the user moves per day), indoor mobility and the total travelled distance (i.e., outdoor mobility) per day;
2. **Conversation data**, including conversation duration and frequency per day;
3. **Sleep data**, including sleep duration, sleep onset and waking time; and finally
4. **Location data**, including GPS, inferred buildings when the participant is indoors, and the number of co-located Bluetooth devices.

The StudentLife Dataset was collected using the StudentLife app, which is a smartphone application and sensing system that automatically infers human behaviour in an energy-efficient manner. StudentLife app uses a continuous sensing engine, which balances the performance needs of the application and the resource demands of continuous sensing on the phone.

Different types of sensing information were logged with different frequencies, depending on the sensor type and the global workload of the phone. For example, activity logs are sampled every 2–3 seconds in 1 of every 4 minutes, smartphone app usage was logged every 20 minutes, and location logs were sampled every 10 minutes.

The StudentLife dataset has been used in several fields, both in human activity learning and prediction (Chen et al. 2014) and in mental wellness assessment (Saeb et al. 2016) (Wang et al. 2017)(Harari et al. 2017). This dataset is, to the best of our knowledge, the best suited for studying and analyzing models for FSBP, due to its 10-weeks duration (which let us analyze how sedentary behaviour change along time) and its wide variety of sensing data. Furthermore, this dataset was used in several works for FSBP, fact that enables to compare the results and the performance of the proposed models.

5.2 - Data Cleaning

All of the sensing data in the StudentLife dataset was tagged with a timestamp. We decided to discretize the time-series into one-hour buckets since this is the granularity used in most of the related work for sedentary prediction. Then, all the features that were generated from the raw dataset correspond to a particular user/hour combination. For example, a specific 1-hour-bucket might correspond to subject 10 and hour 2013-04-24 19:00–20:00.

In the dataset, students' physical activity was labelled as *stationary*, *walking*, *running* and *unknown*. We discarded the physical activity logs labelled with *unknown,* since they do not provide enough information regarding the activity that was being performed by the subjects. Then, we only kept the 1-hour-buckets in which there were available logs of the physical information of the subject since the sedentary level is computed from it. Since our goal is to predict future sedentary behaviour we also removed the buckets for which we had no physical activity information of the next hour. Finally, we removed subject 52 of the dataset due to inconsistencies found in its data. For example, Fig 3. shows a sample of the cumulative activity time (%) for each type of activity, for each hour of the day for subject 52. We can observe that records for this subject indicate an improbable running/walking activity in three consecutive days.

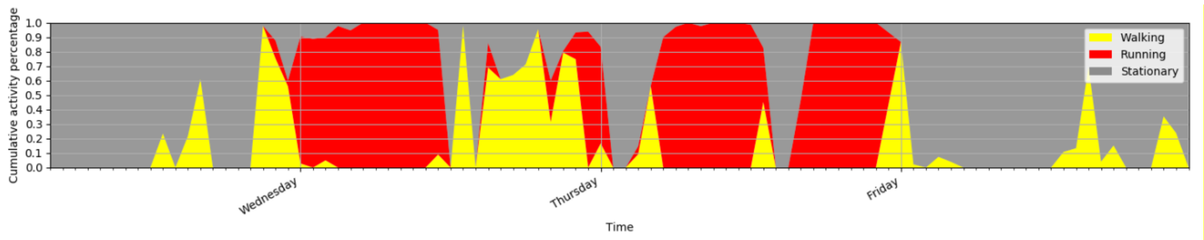

Fig.3 Cumulative activity time (percentual) for each type of activity, for each hour of the day for subject 52.

After discarding of invalid buckets, we obtained 60,819 one-hour-bucket for all the 48 remaining subjects.

For the remaining buckets, we computed the sedentary level, in terms of MET, from the physical activities performed by each subject. We established the MET level of each physical activity according to the Compendium of physical activities[5] (Ainsworth et al. 2011). For those activities with "stationary" labels in the dataset, we set a MET value 1.3 (corresponding to activities such as "sitting quietly", "lying quietly", "doing nothing", "lying in bed awake", "listening to music (not talking or reading)" in the Compendium). For those activities labelled as "walking" in the dataset, we set a MET value of 5.0 (corresponding to activities such as "carrying 15-pound load -e. g. suitcase- level ground or downstairs" in the Compendium). Finally, for activities labelled as "running" in the dataset, we set a MET value of 8.3, which correspond to activities such as "running, 5 mph (12 min/mile)" in the Compendium. From the resulting MET values, we classified each 1-hour-bucket as sedentary or non-sedentary. Each 1-hour bucket is classified as sedentary if the average of the MET values of the physical logs of that particular bucket is equal or less than 1.5 (Tremblay 2012). For example, if a 1-hour-bucket contains 100 stationary logs, 10 walking logs and 10 running logs, the MET level of that bucket is calculated as (100*1.3 + 10*5 + 10*8.5)/120 = 2.2. Therefore, the sedentary class for that particular 1-hour bucket will be categorized as non-sedentary.

5.3 - Features definition

Since we discretized all the data in 1-hour buckets, we had to hypothesize which information was going to be useful to predict the next hour sedentary behaviour. The variables used were chosen based on the data available in the StudentLife dataset and on previous research (He and Agu 2016b; Cook and Krishnan 2015).

---

[5] https://sites.google.com/site/compendiumofphysicalactivities/

Although subjects in the dataset were students, we avoided to include data that might only be present in this kind of subjects, for example, the time remaining for a scheduled exam, if he/she is attending to a given class, etc. Therefore, in order to make our approach more general, we only defined features that might be computed from data collected from other population.

We describe next the 20 smartphones sensed context variables selected and computed by processing the StudentLife dataset.

GPS features

- distanceTraveled: the total haversine distance travelled calculated from the GPS logs in a 1-hour time bucket;
- locationVariance: used to measure the variability in a participant's GPS location, computed as described in (Saeb et al. 2015).

Time features

- hourSine: sine transformation of the hour;
- hourCosine: cosine transformation of the hour;
- dayOfWeek: the day of the week;
- pastMinutes: the number of minutes that passed since the beginning of the day;

  remainingMinutes: the number of minutes left to finish the day;

Physical activity features

- Stationary level: the percentage of stationary instances of physical activity in a 1-hour time bucket;
- Walking level: the percentage of walking instances of physical activity in a 1-hour time bucket;
- Running level: the percentage of running instances of physical in a 1-hour time bucket;
- activityMajor: the type of audio with the most instances in a 1-hour time bucket;

Audio features

- SilenceLevel: the percentage of silence instances of audio activity in a 1-hour time bucket;
- voiceLevel: the percentage of voice instances of audio activity in a 1-hour time bucket;
- noiseLevel: the percentage of noise instances of audio activity in a 1-hour time bucket;
- numberOfConversations: the number of conversations that the student had in a 1-hour time bucket;

Other features

- isCharging: whether the smartphone was charged;
- isLocked: whether the smartphone was locked;
- isInDark: whether the smartphone was in the darkness;
- hasCalendarEvent: whether an event is scheduled in the calendar;
- wifiChanges: the number of wifi changes in a 1-hour time bucket.
- actualClass: if the user's current class is sedentary or not.

We hypothesize that time-related features may have an important role in FSBP. Therefore, we pay special attention to the more convenient time-related features that we should extract to characterize each 1-hour bucket. We took the decision of treating the hour of the day as a numerical variable and not as a categorical variable, but since it is a cyclical numerical variable, we performed a sine and cosine transformation to the hour resulting in two features for expressing the hour of the day.

In contrast, we took the day of the week as a categorical variable, since when we analyzed the activities of the students along the whole week we saw that each particular day had its own particular patterns. Thus, we can assume that this categorical variable has no particular ordering.

Finally, since the features *dayOfWeek* and *activityMajor* are categorical features, they were dummy-encoded. Dummy encoding is a method for converting a categorical input features into a binary features, such that they can be correctly interpreted by the predictive model. With this transformation, we obtained 7 dummy features representing the categories of the features *dayOfWeek and 3* for the features *activityMajor.* Finally, we end up with 28 smartphone variables for predicting the sedentary behavior of the subjects.

5.4 - Sedentary Behaviour Analysis

We generated heat maps with the average and standard deviation of the MET levels of all users according to the day of the week. To do this, we grouped 1-hour buckets per day of the week (y-axis) and time of day (x-axis). Heat maps plotting the average level of sedentary behaviour had already been used in other works but without taking into account the MET level. In these related works (He and Agu 2016a); (He and Agu 2016b); (He and Agu 2016c), heat maps were generated with the percentage of activities classified as stationary during each 1-hour bucket. Heat maps for the standard deviation of each 1-hour bucket have not been analyzed in previous works. We think that both types of heat maps complement each other because together they allow us to generate hypotheses about the routines of the users in relation to their sedentary behaviour and to understand the importance of time features when generating predictive models.

Fig. 1 and Fig. 2 show as an example the heat maps of the average and the standard deviation of subject 46 in the dataset. We can observe that the hours of the day with greatest sedentary behaviour (that is, MET levels lower than 1.5) also correspond to a low standard deviation. The same observation can be made for most of the users in the dataset. A low standard deviation suggests that the observed sedentary behaviour correspond to a routine behaviour, while non-sedentary behaviour are not part of the user's routine due to the high standard deviation.

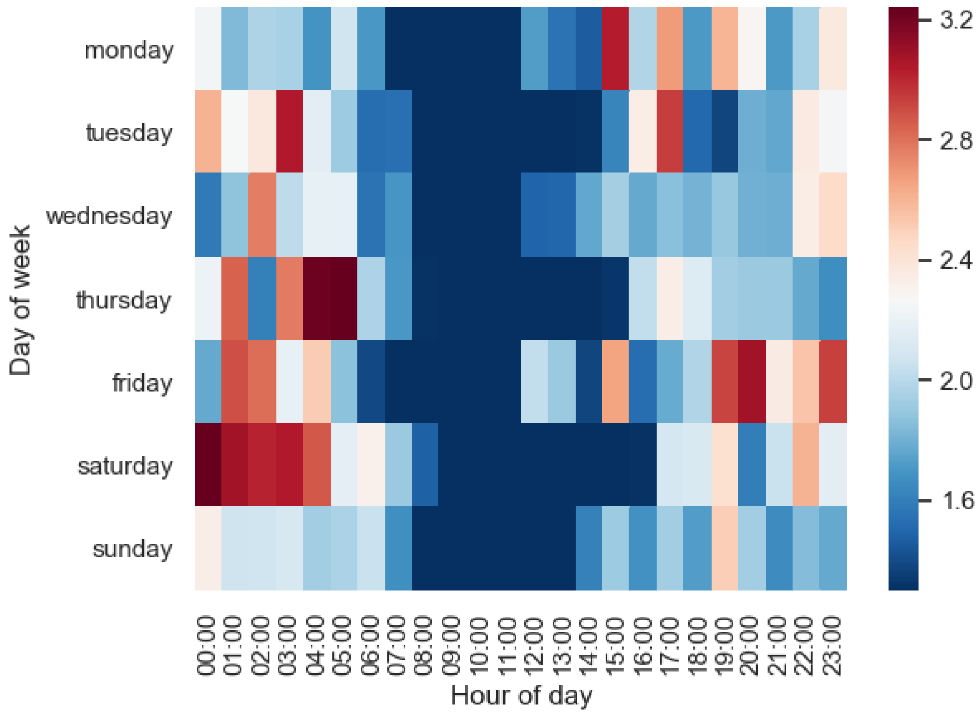

Fig. 1: Subject 46's average of the sedentary level in 1-hour buckets grouped by the day of week and hour of the day

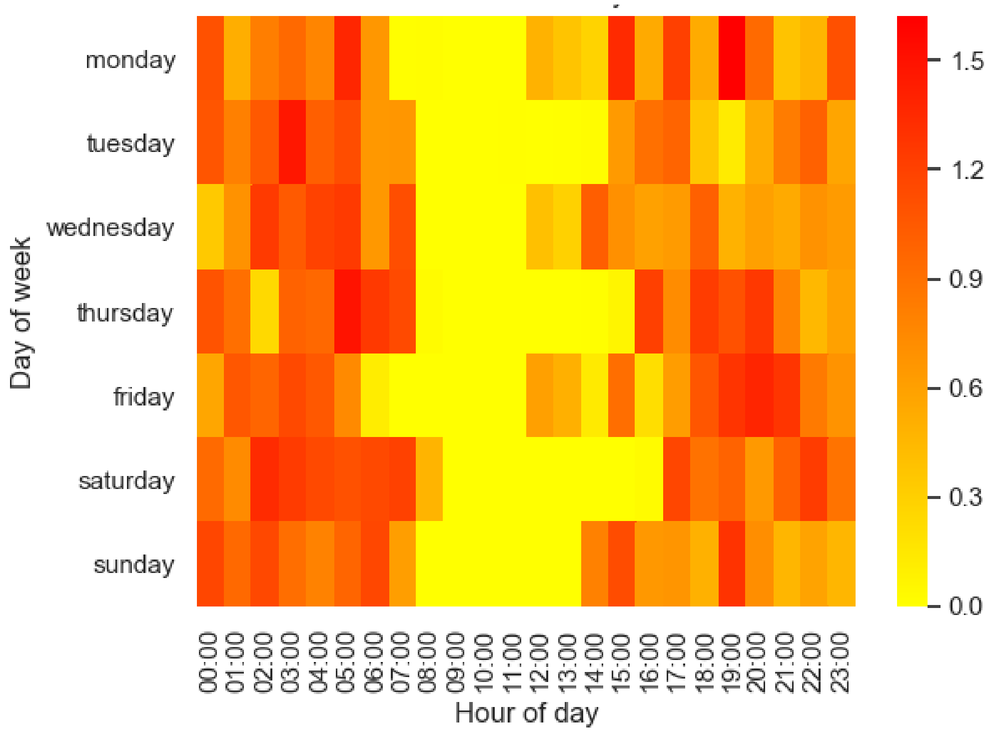

Fig. 2: Subject 46's standard deviation of the sedentary level in 1-hour buckets grouped by the day of week and hour of the day

We computed the Pearson correlation between the mean and the standard deviation of all users for all possible hour/day of week combination. Then, we computed the averaged correlation and resulted in 0.87, which supports our observation that sedentary behaviour is correlated with routine behaviour. Since the subjects that participated in the experiments correspond to students, this correlation can be explained by the fact that two possible routine behaviours that have a low level of energy expenditure are the hours of sleep and the schedule of courses. The information about the classes to which each subject assisted was part of the dataset, so we were able to verify that the class schedules of the students correspond to the hours of low energy expenditure. In, summary, users tended to be more sedentary in routine activities than in the non-routine activities. This observation leads to the hypothesis that sedentary behaviour is more predictable than non-sedentary behaviour in terms of time features.

5.4 - Consideration regarding hours of sleep

The standard definition of sedentary behaviour emphasizes the fact that for an activity to be considered sedentary, the subject who is carrying it out must be awake. Nevertheless, previous studies used all the 1-hour buckets to train the models, including those in which the subjects were sleeping. Then, in order to be able to compare the results obtained, all the experiments have been performed by including and excluding the hours in which the subjects were sleeping.

We consider a 1-hour bucket in which the user might be sleeping as a bucket between 22 and 5 o'clock classified as sedentary (MET level ≤ 1.5). One of the benefits of removing the hours of sleep is that the class imbalance is reduced, significantly improving the performance of the predictive models. In spite of having reduced the imbalanced of the dataset, the data of some particular subjects continued to be very unbalanced due to their extremely sedentary behaviour.

After discarding the hours of sleep, we obtained 41,852 one-hour-bucket for all the 48 remaining subjects.

6 - Experiments

In this section, we describe the experiments that were carried out to test our predictive models. We first describe the methodology (Section 6.1) and the evaluation metrics (Section 6.2) used. Finally, in Section 6.3, we describe the results obtained.

6.1 - Methodology

The evaluation of personal and impersonal models is different, so the procedure to measure the performance of each one was different. To have a reliable measurement of the performance of personal models, we applied a 10-fold cross validation over each user personal records. For example, if a user had 1000 1-hour buckets (i.e. 1000 records in the dataset), for each fold, we trained the model with 900 1-hour buckets and tested it with the remaining 100 1-hour buckets. Once all the models were trained and tested, the weighted F1-score of each fold are averaged.

Unlike personal models, impersonal models are trained with the records corresponding to all the users but the one that will be used for testing. This technique is called Leave-One-Subject-Out validation. In our case, we evaluated 48 different models, one for each subject, by training the model with the information available from the remaining 47 subjects.

6.2 Evaluation metric

Usually, precision and recall are used to evaluate classification models and are calculated using the concepts of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). Precision for the positive class (i.e. sedentary behaviour) is defined as the fraction of sedentary behaviours reported by the model that was correctly predicted and is calculated as $\frac{TP}{TP + FP}$. Precision for the negative class (i.e. non-sedentary behaviour) is defined as the fraction of non-sedentary behaviours reported by the model that was correctly predicted and is calculated as $\frac{TN}{TN + FN}$.

On the other hand, recall for the positive class is defined as the fraction of true sedentary behaviours that were correctly detected and is computed as $\frac{TP}{TP + FN}$. Recall for the negative class (sedentary behaviour) is defined as the fraction of true non-sedentary behaviours that were correctly detected and is computed as $\frac{TN}{TN + FP}$.

Finally, F1-Score is the harmonic average of the precision and recall, where an F1 score reaches its best value at 1 (perfect precision and recall) and worst at 0. F1-score is computed, for each class, as $\frac{2 * Precision * Recall}{Precision + Recall}$.

We took the decision of using the weighted F1-Score, because of the problem of class imbalance. The weight for each class is calculated as the total number of samples that are being used to test the model. In the dataset, there are about 4 times more 1-hour buckets classified as sedentary than 1-hour buckets classified as non-sedentary.

In summary, the measuring procedure for each model was as follow. First, for each class, precision and recall are obtained. Then, we compute the F1-Score for each class. Finally, we take the weighted average between the F1-Score of the two classes, in place of the macro average, to measure the performance of the classifier for each class.

6.3 - Results

We evaluated eight model types, categorized in three axes: personal and impersonal, Logistic Regression and DNN and taking into account the hours of sleep or not. In total, more than 2 thousand instances of these model types have been trained and tested. Table 1 shows the mean and the standard deviation of the F1 score for all the model types and Fig. 3 shows their corresponding boxplots.

Table 1. The mean and standard deviation of the F1-score for each predictive model

			Mean	Standard Deviation
Personal model	With hours of sleep	LR	0.739	0.042
		DNN	0.747	0.047
	Without hours of sleep	LR	0.754	0.046
		DNN	0.761	0.049
Impersonal model	With hours of sleep	LR	0.745	0.042
		DNN	0.767	0.078
	Without hours of sleep	LR	0.757	0.052
		DNN	**0.8**	0.056

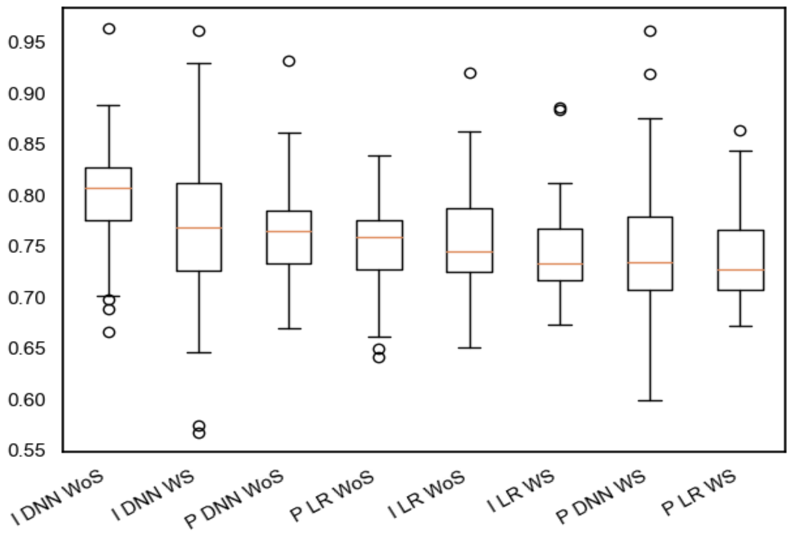

Fig. 3: Boxplot comparing the eight model types evaluated (I = Impersonal, P = Personal, DNN = Deep Neural Network, LR = Logistic Regression, WS = With sleeping hours, WoS = Without sleeping hours)

We can highlight three observations from the analysis of these results:

1. the performance of DNN always surpassed LR;
2. Impersonal models had a better performance than personal models;
3. As discussed in the previous section, removing the hours of sleep reduced the imbalance of the dataset. As expected, models trained without hours of sleep achieved a better performance than models that had been trained using this information.

Next, we compare the differences between personal and impersonal models for the predictive model with better performance, that is, DNN without considering hours of sleep. Fig 4 shows the F1-Score achieved by personal and impersonal models for all of the subjects in the dataset. We can observe that, for most of the subjects, the impersonal model obtained a better performance than the corresponding personal model. It is interesting to notice that only for 2 subjects out of a total of 48 (subjects 9 and 10) the F1-Score achieved by the personal model type surpassed by more than 5% the F1-Score achieved by the impersonal model type. We hypothesize that this may happen because subjects 9 and 10 exhibited a very particular behaviour, which is different from that exhibited for the rest of the users in the dataset. Since personal models can learn the idiosyncrasies of each particular user, these type of model were able to capture these particular behaviours. From the discussion presented in Section 5.4, we also deduce that impersonal models can have difficulties with subjects having a non-routine activity rate greater than the average non-routine activity rate. Since we focused mainly on time-related features, routine-behaviour has an important influence on the predictability of a subject future sedentary behaviour.

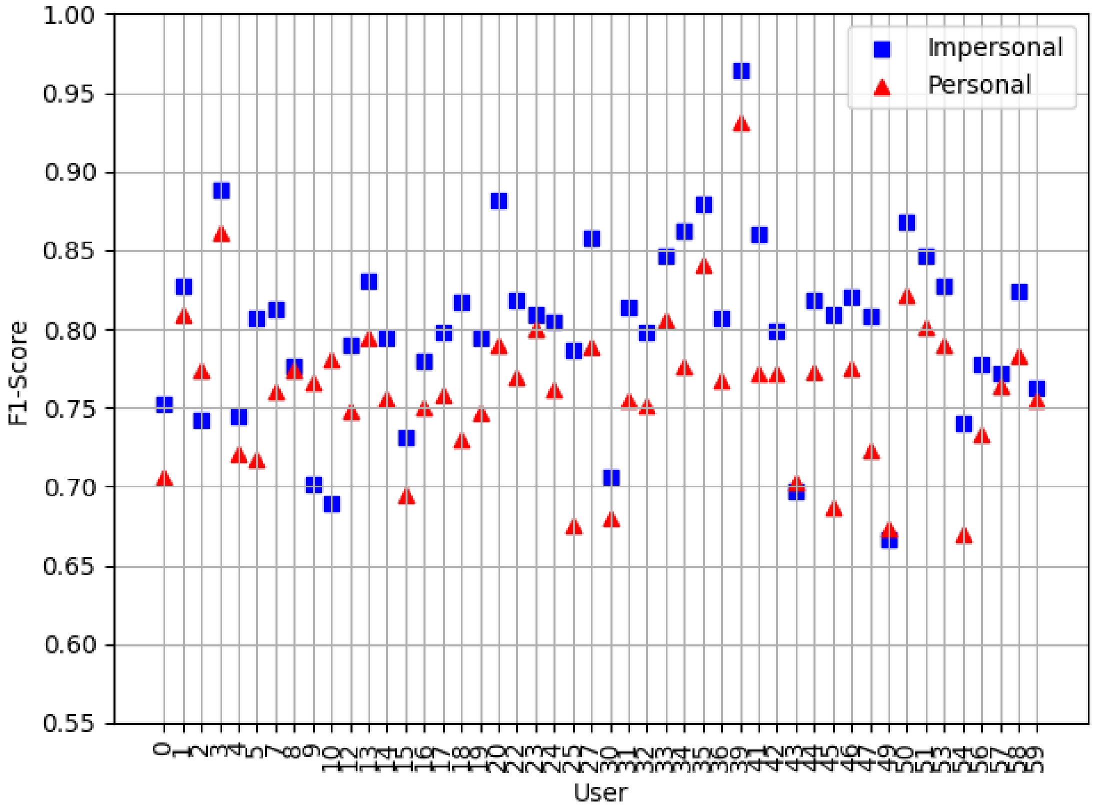

Fig. 4: Comparison of the F1 score achieved by personal and impersonal models for each user

6.7 - Comparison with related work

He and Agu (He and Agu 2016b) used Naıve Bayes classifier and Logistic Regression classifier to predict sedentary behaviour in three levels: Very Sedentary, Sedentary, and Less Sedentary. The results obtained in terms of correctly classified instances using 10-fold cross-validation, were 59.7% (with recall 64.2%) and 60.8% (with recall 62.6%) for Naıve Bayes classifier and Logistic Regression, respectively. In predicting Very Sedentary behaviour, which will likely require behaviour intervention, Naïve Bayes classifier achieved a precision of 70.2% (with recall 86.0%) and Logistic Regression classifier achieved a precision of 68.6% (with recall 89.3%). Finally, their results slightly improved by adding the user's current sedentary behaviour in addition to the current context variables.

7 - Conclusions and Future work

In this article, we presented an approach to the problem of using data collected from sensors installed in wearable/mobile devices to predict the future sedentary behaviour of a subject. We evaluated personal and impersonal models for two predictive techniques: Deep Neural Networks and Logistic Regression. We also compared the effect of excluding the hours of sleep from the analysis. Differently, from previous approaches, instead of using the percentage of sedentary/non-sedentary activities in each 1-hour bucket, we considered the Metabolic Equivalent of Tasks (MET). MET is a standard metric in the area of health for measuring the intensity of an activity in terms of energy expenditure. This metric, to the best of our knowledge, was not used for the problem of predicting sedentary behaviour by using wearables and mobile devices.

We evaluated the proposed approach with a real-life dataset consisting of sensing data of 49 users during 10 weeks. We found that impersonal models using Deep Neural Networks were able to predict the subject's future sedentary behaviour with an F1-score of 0.8.

One of the limitations of our work is that we assigned a fixed MET value for each of the three activities registered in the dataset. For example, we assigned a value of 5.0 MET for “walking” activities, while there are more than 50 different “walking” activities considered in the Compendium. Each of those activities has a different MET value assigned, varying from 2.0 to 12.0. We believe that better results could be obtained if we were able to capture more fine-grained activities, for example, by using the accelerometer sensing data.

Given the impact of time-related features in the predictive models, as a future work, we plan to explore the performance of sequential models such as Recurrent Neural Networks and Long Short-Term Memory Networks.

References

Cook, Diane J., and Krishnan, Narayanan C. 2015. *Activity Learning*, Willey Series on Parallel and Distributed Computing.

Ainsworth, Barbara E., William L. Haskell, Stephen D. Herrmann, Nathanael Meckes, David R. Bassett Jr, Catrine Tudor-Locke, Jennifer L. Greer, Jesse Vezina, Melicia C. Whitt-Glover, and Arthur S. Leon. 2011. “2011 Compendium of Physical Activities: A Second Update of Codes and MET Values.” *Medicine and Science in Sports and Exercise* 43 (8): 1575–81.

Ajzen, Icek. 1991. “The theory of planned behaviour” *Organizational Behaviour and Human Decision Processes* 50 (2): 179-211

Atkin AJ, Gorely T, Clemes SA, Yates T, Edwardson C, Brage S, Salmon J, Marshall SJ, Biddle SJ. 2012. “Methods of Measurement in epidemiology: sedentary Behaviour”. *International Journal of Epidemiology* 41 (5): 160-1471.

Benatti, Fabiana Braga, and Mathias Ried-Larsen. 2015. “The Effects of Breaking up Prolonged Sitting Time: A Review of Experimental Studies.” *Medicine and Science in Sports and Exercise* 47 (10): 2053–61.

Bishop, Christopher M. 2006. *Pattern Recognition and Machine Learning*. Springer.

Carter, Sophie, Yvonne Hartman, Sophie Holder, Dick H. Thijssen, and Nicola D. Hopkins. 2017. “Sedentary Behavior and Cardiovascular Disease Risk: Mediating Mechanisms.” *Exercise and Sport Sciences Reviews* 45 (2): 80–86.

Chen, Fanglin, Rui Wang, Xia Zhou, and Andrew T. Campbell. 2014. “My Smartphone Knows I Am Hungry.” In *Proceedings of the 2014 Workshop on Physical Analytics - WPA ’14*. https://doi.org/10.1145/2611264.2611270.

Fahim, Muhammad, Thar Baker, Asad Masood Khattak, and Omar Alfandi. 2017. “Alert Me: Enhancing Active Lifestyle via Observing Sedentary Behavior Using Mobile Sensing Systems.” In *2017 IEEE 19th International Conference on E-Health Networking,* *Applications and Services (Healthcom)*. https://doi.org/10.1109/healthcom.2017.8210838.

Falck, Ryan S., Jennifer C. Davis, and Teresa Liu-Ambrose. 2017. "What Is the Association between Sedentary Behaviour and Cognitive Function? A Systematic Review." *British Journal of Sports Medicine* 51 (10): 800–811.

Felez-Nobrega, Mireia, Charles H. Hillman, Kieran P. Dowd, Eva Cirera, and Anna Puig-Ribera. 2018. "ActivPAL™ Determined Sedentary Behaviour, Physical Activity and Academic Achievement in College Students." *Journal of Sports Sciences* 36 (20): 2311–16.

Gong, Jiaqi, Yu Huang, Philip I. Chow, Karl Fua, Matthew S. Gerber, Bethany A. Teachman, and Laura E. Barnes. 2019. "Understanding Behavioral Dynamics of Social Anxiety among College Students through Smartphone Sensors." *An International Journal on Information Fusion* 49: 57–68.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press.

Grundgeiger, Tobias, Jürgen Pichen, Jennifer Häfner, Birgit Wallmann-Sperlich, Diana Löffler, and Stephan Huber. 2017. "Combating Sedentary Behavior." In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17*. https://doi.org/10.1145/3027063.3053094.

Harari, Gabriella M., Samuel D. Gosling, Rui Wang, Fanglin Chen, Zhenyu Chen, and Andrew T. Campbell. 2017. "Patterns of Behavior Change in Students over an Academic Term: A Preliminary Study of Activity and Sociability Behaviors Using Smartphone Sensing Methods." *Computers in Human Behavior* 67: 129–38.

He, Qian, and Emmanuel Agu. 2014. "On11." In *Proceedings of the 2014 Workshop on Physical Analytics - WPA '14*. https://doi.org/10.1145/2611264.2611268.

He, Qian, and Emmanuel O. Agu. 2016a. "A Frequency Domain Algorithm to Identify Recurrent Sedentary Behaviors from Activity Time-Series Data." In *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. https://doi.org/10.1109/bhi.2016.7455831.

———. 2016b. “Smartphone Usage Contexts and Sensable Patterns as Predictors of Future Sedentary Behaviors.” In *2016 IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT)*. https://doi.org/10.1109/hic.2016.7797695.

———. 2016c. “Towards Sedentary Lifestyle Prevention: An Autoregressive Model for Predicting Sedentary Behaviors.” In *2016 10th International Symposium on Medical Information and Communication Technology (ISMICT)*. https://doi.org/10.1109/ismict.2016.7498879.

———. 2017. “A Rhythm Analysis-Based Model to Predict Sedentary Behaviors.” In *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. https://doi.org/10.1109/chase.2017.122.

Ioffe Sergey, Szegedy Christian. 2015. “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift.” *Proceedings of the 32nd International Conference on International Conference on Machine Learning.* Volume 37: 448-456.

Kanjo, Eiman, Younis, Eman M.G., and Ang, Chee Siang. 2019. “Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection”. *Information Fusion* 49: 46-56

Kingma Diederik P, Ba Jimmy Lei. 2015.”Adam: A Method for Stochastic Optimization”. *ICLR 2015: International Conference on Learning Representations 2015*.

Koster, Annemarie, Paolo Caserotti, Kushang V. Patel, Charles E. Matthews, David Berrigan, Dane R. Van Domelen, Robert J. Brychta, Kong Y. Chen, and Tamara B. Harris. 2012. “Association of Sedentary Time with Mortality Independent of Moderate to Vigorous Physical Activity.” *PloS One* 7 (6): e37696.

Tremblay, Mark. 2012. “Letter to the Editor: Standardized Use of the Terms ‘sedentary’ and ‘sedentary Behaviours’” *Mental Health and Physical Activity* 6 (1): 55–56.

Lockhart, Jeffrey W., and Gary M. Weiss. 2014. “The Benefits of Personalized Smartphone-Based Activity Recognition Models.” In *Proceedings of the 2014 SIAM International Conference on Data Mining*, 614–22.

Magnon, Valentin, Guillaume T. Vallet, and Catherine Auxiette. 2018. “Sedentary Behavior at Work and Cognitive Functioning: A Systematic Review.” *Frontiers in Public Health* 6 (August): 239.

Paing, Aye C., Kathryn A. McMillan, Alison F. Kirk, Andrew Collier, Allan Hewitt, and Sebastien F. M. Chastin. 2018. “The Associations of Sedentary Time and Breaks in Sedentary Time with 24-Hour Glycaemic Control in Type 2 Diabetes.” *Preventive Medicine Reports* 12 (December): 94–100.

Saeb, Sohrab, Emily G. Lattie, Stephen M. Schueller, Konrad P. Kording, and David C. Mohr. 2016. “The Relationship between Mobile Phone Location Sensor Data and Depressive Symptom Severity.” *PeerJ* 4 (September): e2537.

Saeb, Sohrab, Mi Zhang, Christopher J. Karr, Stephen M. Schueller, Marya E. Corden, Konrad P. Kording, and David C. Mohr. 2015. “Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study.” *Journal of Medical Internet Research* 17 (7): e175.

Schein, Andrew I., Alexandrin Popescul, Lyle H. Ungar, and David M. Pennock. 2002. “Methods and Metrics for Cold-Start Recommendations.” In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR ’02*. https://doi.org/10.1145/564418.564421.

Wang, Rui, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T. Campbell. 2014. “StudentLife.” In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp ’14 Adjunct*. https://doi.org/10.1145/2632048.2632054.

———. 2017. “StudentLife: Using Smartphones to Assess Mental Health and Academic Performance of College Students.” In *Mobile Health*, 7–33.

Wilmot, E. G., C. L. Edwardson, F. A. Achana, M. J. Davies, T. Gorely, L. J. Gray, K. Khunti, T. Yates, and S. J. H. Biddle. 2012. “Sedentary Time in Adults and the Association with Diabetes, Cardiovascular Disease and Death: Systematic Review and Meta-Analysis.” *Diabetologia* 55 (11): 2895–2905.

Wu, Congyu, Mehdi Boukhechba, Lihua Cai, Laura E. Barnes, and Matthew S. Gerber.

2018. “Improving Momentary Stress Measurement and Prediction with Bluetooth Encounter Networks.” *Smart Health*. https://doi.org/10.1016/j.smhl.2018.07.017.

Zia, Jonathan, Arash Tadayon, Troy McDaniel, and Sethuraman Panchanathan. 2016. “Utilizing Neural Networks to Predict Freezing of Gait in Parkinson’s Patients.” In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS ’16*. https://doi.org/10.1145/2982142.2982194.