A Machine Learning Approach for Prediction of Sedentary Behavior Based on Daily Step Counts

Evangelos Papathomas, Andreas Triantafyllidis, Rafail-Evangelos Mastoras, Dimitrios Giakoumis, Konstantinos Votis and Dimitrios Tzovaras

Information Technologies Institute, Centre for Research and Technology Hellas
6th Klm. Charilaou - Thermi Road, P.O. BOX 60361 GR - 570 01, Thessaloniki, Greece
{epapathom, atriand, remastoras, dgiakoum, kvotis, dimitrios.tzovaras}@iti.gr

Abstract— **Sedentary behavior is considered as a major public health challenge, linked with many chronic diseases and premature mortality. In this paper, we propose a steps counting -based machine learning approach for the prediction of sedentary behavior. Our work focuses on analyzing historical data from multiple users of wearable physical activity trackers and exploring the performance of four machine learning algorithms, i.e., Logistic Regression, Random Forest, XGBoost, Convolutional Neural Networks, as well as a Majority Vote Ensemble of the algorithms. To train and test our models we employed a crowd sourced dataset containing a month's data of 33 users. For further evaluation, we employed a dataset containing 6 months of data of an additional user. The results revealed that while all models succeed in predicting next-day sedentary behavior, the ensemble model outperforms all baselines, as it manages to predict sedentary behavior and reduce false positives more effectively. On the multi-subjects test dataset, our ensemble model achieved an accuracy of 82.12% with a sensitivity of 74.53% and a specificity of 85.71%. On the additional unseen dataset, we achieved 76.88% in accuracy, 63.27% in sensitivity and 81.75% in specificity. These outcomes provide the ground towards the development of real-life artificially intelligent systems for sedentary behavior prediction.**

Keywords— **healthcare, sedentary behavior, wearable devices, machine learning**

I. Introduction

Sedentary behavior among children and adults has become a matter of serious concern over the last decades. Recent studies have associated physical inactivity with the development of various physical and mental health issues such as obesity [1], cardiovascular disease [2], diabetes [3] and depression [4]. As a result, organizations and researchers around the world focus on providing useful recommendations and guidelines that encourage healthier lifestyles and promote physical activity. Piercy et al. [5] propose duration-based physical activity guidelines for children and adults, while also considering their health history and status. On the other hand, Tudor-Locke et al. [6] took advantage of objective monitoring methods using widely available pedometers and accelerometers to propose step-based guidelines for healthy children and adults, and individuals with disabilities and/or chronic diseases.

In [7], Tudor-Locke et al. explore the harmful effects of prolonged sedentary behavior on health and examine its association with low daily step counts. Even though the lack of evidence prevents them from proposing a minimum daily step count for children and adolescents, Tudor-Locke et al. define a sedentary lifestyle for adults using a threshold of 5,000 steps/day. Studies conducted by [8], [9], [10] and [11] have supported this threshold by demonstrating the effects of an abrupt transition from high to low daily step-based physical activity on healthy individuals. Krogh-Madsen et al. [9] have shown that a transition from >10,000 steps/day to <2000 steps/day in young men for a 2-week period reduced skeletal muscle insulin sensitivity and induced a 7% decline in maximal oxygen consumption. In addition, Mikus et al. [10] determined that even a three-day reduction from 10,000 steps/day to 5,000 steps/day in healthy, active individuals leads to substantial increases in insulin and C-peptide responses to the oral glucose tolerance test (OGTT).

As it is evident, monitoring the daily step counts of individuals and encouraging healthy habits can potentially prevent the development of several life-threatening health issues. The wide availability and popularity of wearable devices for physical activity tracking reinforces physical activity promotion. However, to effectively intervene and avert cases of sedentary behavior, it is important to regularly look for patterns that reveal sedentary inclinations. By predicting days where the person is prone to demonstrate sedentary behavior, a physical activity monitoring system could be better informed beforehand and take more proactive decisions on when and how to support the user in remaining physically active.

In this vein, we introduce a machine learning approach towards the prediction of sedentary behavior. By employing a series of machine learning algorithms, we perform classification on daily step count time series segments of fixed length that are annotated using the step count of the following day. To the authors' knowledge, this is one of the first works on the prediction of sedentary behavior utilizing a machine learning approach based on step count data. In the following sections we present a brief theoretical background, we detail our implementation, we evaluate the results and, finally, we provide insight for future improvements.

II. Background and Related Work

In the last few years, the dramatic growth of temporal data has led to the development of several time series classification (TSC) algorithms [12]-[14]. Modern approaches to TSC explore a variety of machine learning architectures on both univariate and multivariate time series datasets.

A. Related Work

So far, there have been only a handful of studies for the prediction of sedentary behavior. In 2016, He et al. [15] introduced an autoregressive model with maximum entropy

method (MEM) [16] to predict sedentary behavior using raw activity logs from individuals. In search for the optimal parameters for their model, the research group utilized the physical activity logs of the StudentLife dataset [17]. The group determined that a person's sedentary behavior the next hour is correlated with their previous behaviors in the past six hours, and that sedentary patterns are often repeated in a daily and weekly basis.

He et al. proposed a rhythm analysis-based model [18] for the prediction and the prevention of sedentary behavior. The model detects the rhythms of sedentary behavior and models cyclical and linear rhythms through periodic and linear functions, respectively. Again, through validating their model on the physical activity logs of the StudentLife dataset, the group was able to detect half-day rhythms, daily rhythms, weekly rhythms and biweekly rhythms among the students.

In 2019, Ozogur et al. [19] proposed a deep learning approach to the prediction of sedentary behavior. Using a 6-hour window of historical physical activity data, the team trained a recurrent neural network (RNN) model that predicts the sedentary levels of the next hour for each student in the StudentLife dataset. After evaluating the RNN model through comparing it directly to a single neuron model, they concluded that it performs better for most students.

The aforementioned approaches focus on predicting sedentary behavior rhythms on an hourly rather than a daily basis, and they are based on activity states. In contrast with the above works, our proposed implementation focuses on daily predictions of step-defined sedentary behavior using objective monitoring methods, while following widely accepted physical activity recommendations. As such, we attempt to prevent sedentary behaviors a day before they occur towards encouraging individuals to become more active.

B. Background

In this work, we explore the Logistic Regression, Random Forest [20], XGBoost [21] and CNN [13] algorithms for the prediction of daily sedentary behavior. Logistic Regression utilizes a logistic function to linearly combine independent variables and model their relationship with one binary dependent variable. Random Forest is a tree ensemble algorithm that creates a forest of random decision trees, fitted using a bagging technique. XGBoost is a recent, and rather popular, ensemble learning algorithm that combines the predictions of multiple regression trees, creating a scalable tree boosting system.

Ultimately, to obtain improved performance over the performance achieved by each algorithm individually, we explore a voting ensemble [22] of all the aforementioned algorithms. Ensembles typically yield better results in cases where there is substantial diversity between models. As a result, the combination of different algorithms has the potential to promote diversity and enhance predictions.

III. IMPLEMENTATION

In this section we present an overview of the dataset that we utilized, the data preprocessing techniques we applied on it, as well as the machine learning algorithms we examined and evaluated.

A. Dataset Overview

We selected a crowd-sourced dataset [23] comprising physical activity data from several Fitbit users. The dataset contains daily logs of physical activity data, separated into activity intensity, walked steps and burnt calories. For the prediction of sedentary behavior, in our implementation we utilize the daily step count data of the dataset, which consist of 31 days of logged step counts for the majority of the 33 Fitbit users included. We utilized only the daily step count data of the Fitbit dataset, considering that the rest of the data is prone to errors [24] and also highly correlated to the step count.

B. Data Preprocessing

To ensure the quality of the dataset before training and validating our neural network, we thoroughly inspected the daily step counts of each user. After arranging all user logs in the dataset by date, we observed that several missed a small number of dates that other user logs contained. In order to maintain uniformity across the dataset, we injected the missing dates of each user log as dates that contain NaN values. Next, we removed all daily step counts under 500, as we assumed that they were not appropriately counted [25]. In addition, to balance the distribution of the data, we capped all daily step counts to a maximum of 10,000 steps, i.e. the suggested target for healthy adults [6].

To handle the missing data, i.e. 17.69% of the data, we calculated the moving average of each time series in the dataset using a 7-day sliding window. We set the minimum required number of observations with non-missing values in the window to 1, so that, otherwise, the moving average returns missing values. After the calculation, we replaced the missing data using the corresponding values of the moving average. Any missing data that the moving average was unable to calculate were replaced using the mean value of each time series.

To approach the prediction of sedentary behavior through exploiting the historical data of users, we converted each time series to sequences of 7 consecutive days, using a sliding window. We selected a window size of 7 to include potential weekly seasonality. Since each user's time series consists of 31 days, the resulting number of windows for each user was 24. As target data, we selected the next day for each sequence and labeled it as sedentary, if it did not surpass the 5,000 steps/day threshold, or as non-sedentary if it did.

Finally, we created a training and a test set by splitting each time series. Specifically, in the training set we included the first 14 windows and target days of each time series and in the test set we included the remaining 10 windows and target days. This way we involve all users in the training and the evaluation of our models.

C. Logistic Regression, Random Forest and XGBoost

We approached the prediction of sedentary behavior as a binary time series classification problem. Thus, to classify the input sequences we examined several machine learning algorithms, one of them being Logistic Regression. We fitted a Logistic Regression model using the training set discussed in the previous subsection. In addition, due to class imbalance

that resulted from the shortage of sedentary target days in the training set, we calculated the class weights using (1), where W_{class} is the weight of the class, s_{total} is the total number of samples, s_{class} is the number of samples of the class and n is the number of classes. Finally, we applied the class weights to the model before fitting the training data.

$$W_{class} = \frac{s_{total}}{s_{class*n}} \qquad (1)$$

Using the aforementioned training data, we additionally fitted a Random Forest model of 1,000 decision trees and an XGBoost model of 100 boosting rounds. We applied the class weights calculated using (1) to the Random Forest model before the training process. Similarly, before training the XGBoost model, we applied to it the ratio of the negative class samples to the positive class samples.

D. Convolutional Neural Network

To examine the results of a more advanced approach, we designed a neural network consisting of two 1D convolutional layers, a flattening layer and an output dense layer of a single neuron, as shown in Fig. 1.
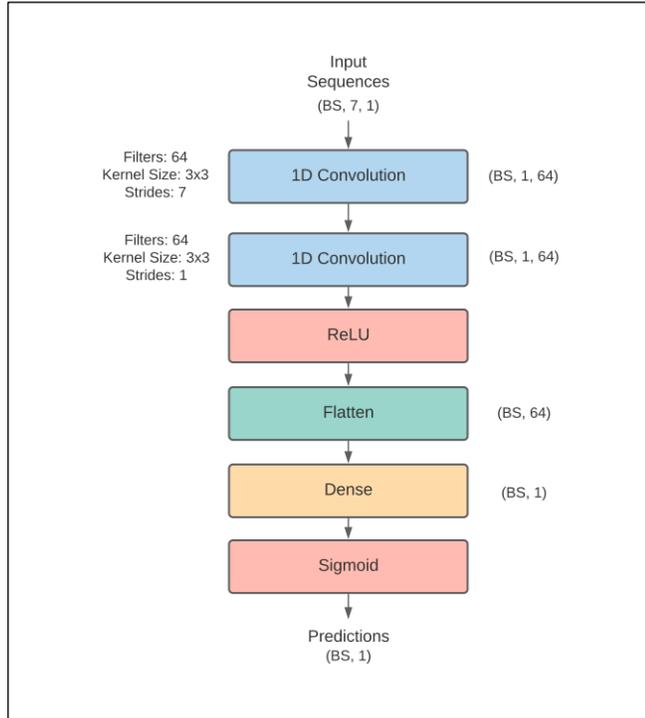

Figure 1.   The architecture of the Convolutional Neural Network.

The input convolutional layer contains 64 filters with a kernel size of 3x3 and acts as a downsampling layer by sliding the convolution window 7 elements at a time. The following convolutional layer also contains 64 filters with a 3x3 sized kernel but slides the convolution window 1 element at a time. To improve our network's performance, we applied the ReLU activation function only to the inner convolutional layer, following the design proposed by Zhao et al. [26]. Next, we flattened the window of the inner convolutional layer and we passed the result to the dense layer, where the sigmoid activation function was applied, to transform the output to a probability.

We trained the neural network using the aforementioned training set. During training, we selected a batch size of 16 samples/step and a 10% validation split. We used the Adam optimizer with a learning rate of 10^{-4}, the binary cross-entropy loss function and an early stopping mechanism. In addition, due to the imbalance between the two classes in the training data, we calculated weights for each class using (1) and applied them on the loss function.

E. Majority Vote Ensemble

To evaluate a combination of all models, we developed a Majority Vote Ensemble. The ensemble determines the output result by taking as input the labeled results of each model and counting the votes for each class. In the case of a draw, the input is labeled as sedentary, as obtaining some false positive results is less crucial than obtaining false negative results. The architecture of the ensemble is shown in Fig. 2.
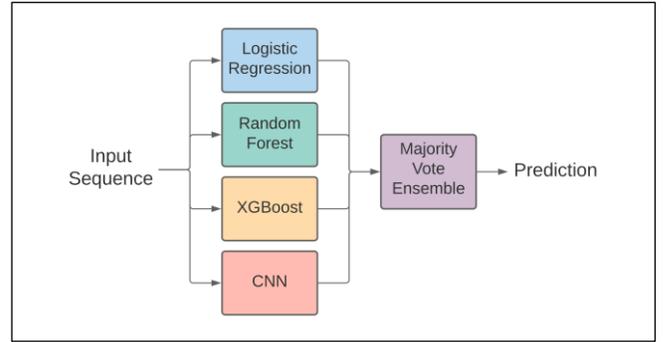

Figure 2.   The architecture of the Majority Vote Ensemble.

IV.   RESULTS

In this section we present and discuss the results of evaluating our models on the remaining unseen data of the dataset, i.e. the remaining 10 sequences of each user's time series. Furthermore, we evaluate our models on an unseen dataset of 200 days from an additional source.

A. Dataset Users

To properly evaluate the performance of our models, we tested their predictions on the remaining sequences from each user's time series. In order to optimize each model's results, under the assumption of equal misclassification cost, we selected a probability threshold using Youden's index. Next, we applied the thresholds to each model's output probabilities to generate the labeled results. In tables I and II, we respectively present the selected probability thresholds and the sensitivity, specificity and accuracy scores that each model achieved.

TABLE I.   THE MODELS' PROBABILITY THRESHOLDS.

Model	Threshold
Logistic Regression	0.8687
Random Forest	0.2870
XGBoost	0.4292
CNN	0.5872

TABLE II. THE MODELS' PERFORMANCE ON MULTIPLE USERS.

Model	Sensitivity	Specificity	Accuracy
Logistic Regression	0.6698	0.8929	0.8212
Random Forest	0.7547	0.8348	0.8091
XGBoost	0.6887	0.8661	0.8091
CNN	0.6981	0.8705	0.8152
Majority Vote Ensemble	0.7453	0.8571	0.8212

We observed that the Random Forest model achieved the highest sensitivity score but the lowest score in specificity. In contrast, the Logistic Regression model achieved the highest specificity score but the lowest sensitivity score. The XGBoost and CNN models attempted to balance their sensitivity and specificity scores, however with a noticeable incline in specificity. Finally, even though its scores did not exceed the highest measured scores, the Majority Vote Ensemble balanced the results in the most efficient manner.

Our models are capable of identifying patterns and correctly classifying sedentary behavior in most cases. However, several users engage in sedentary behavior in a random manner and without any prior indication. Specifically, we noticed a number of sequences in the test data that, even though the included step counts were non-sedentary, the target step count was sedentary. To demonstrate this, in Fig. 3 we present examples of time series with correctly and incorrectly classified sedentary days by the CNN model. The blue line signifies the time series of a user, while the dashed vertical purple line separates the target days in the training set, left of the separator, from the target days in the test set, right of the separator. The green dots mark the correctly classified sedentary days, i.e. the true positives, and the red dots mark the incorrectly classified sedentary days, i.e. the false negatives.
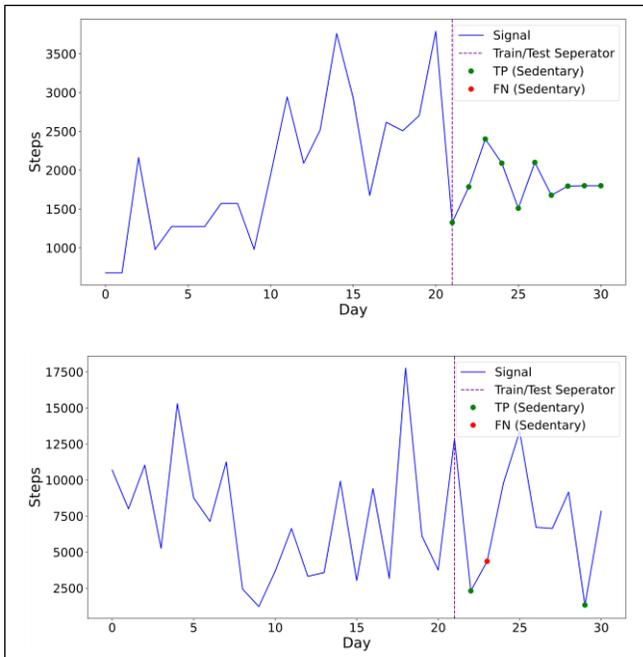

Figure 3. Examples of correct and incorrect classifications.

B. Additional User

To further explore the performances of all models, we evaluated their predictions on an additional dataset. The dataset contained Fitbit Charge logged daily step counts from an additional user in course of 200 days. To prepare the dataset for our models, we applied the preprocessing pipeline we described in the previous section. Furthermore, to obtain the labeled results, we applied the probability thresholds we calculated using the original test dataset. In table III, we present the sensitivity, specificity and accuracy scores that each model achieved on these additional data.

TABLE III. THE MODELS' PERFORMANCE ON THE ADDITIONAL USER.

Model	Sensitivity	Specificity	Accuracy
Logistic Regression	0.5714	0.9051	0.8172
Random Forest	0.6122	0.8248	0.7688
XGBoost	0.6122	0.7956	0.7473
CNN	0.5918	0.8102	0.7527
Majority Vote Ensemble	0.6327	0.8175	0.7688

The results showed that our implementations succeeded in detecting sedentary behavior a day earlier whenever sedentary patterns are available. In detail, the Logistic Regression model achieved the highest specificity score but the lowest score in sensitivity. The Random Forest, XGBoost and CNN models achieved quite similar results, attempting to balance both scores. And finally, the Majority Vote Ensemble achieved the highest sensitivity score, surpassing all baseline models, and a fairly high specificity score.

V. DISCUSSION

In this paper we proposed a machine learning approach to predict sedentary behavior using historic daily steps count data originating from wearable devices. Considering that physical inactivity is a major public health challenge linked with higher morbidity and mortality, the implementation of new ICT technologies is required to facilitate its prevention. Consequently, this paper focuses on predicting physical inactivity using ambient intelligent systems, as a resolution to this rising issue.

Surprisingly, even though sedentary behavior is an important public health challenge, the predictive capabilities of relevant ICT systems have not been thoroughly investigated, nor exploited, thereby motivating the current work. Our outcomes, following the experimentation with different machine learning algorithms, provide an important ground towards the development of real-life artificially intelligent systems for sedentary behavior prediction on a large-scale.

In search for the most optimal prediction technique, we experimented with 4 machine learning algorithms and their ensemble. The results on two different test datasets showed that the ensemble outperforms the baseline models by balancing the metric scores more efficiently and achieving

higher scores in some cases. This finding is in line with outcomes of previous studies in other areas of healthcare, which showed that ensemble models outperform baseline models.

The ensemble's performance improvement is a product of the diverse ways the models fit to the data. In our case, this can be observed in the results of the baseline models, presented in tables II and III. As expected, models with higher sensitivity scores perform worse in specificity, while models with higher specificity scores obtain lower sensitivity, compared to the other models.

The limitations we encountered throughout our implementation were a consequence of the quality of the training data and the similarity between the two classes. Due to the presence of missing values in the data, we included a data augmentation process in the preprocessing pipeline. However, even though augmentation can be considerably beneficial, it cannot be a replacement for the lost information. Furthermore, through manually inspecting the data, we concluded that the extent of the class similarity can introduce confusion to the models and hinder their performance. Thus, we experimented with various machine learning algorithms since their diversity is a potential solution to the problem.

For future work, we intend to test our approach on additional datasets using also additional machine learning algorithms. Furthermore, we will examine model personalization techniques using these datasets, for performance improvements. Finally, we would like to explore additional features that could enhance the predictive capabilities of our approach.

In conclusion, we presented a machine learning modeling approach to predict sedentary behavior based on day-to-day step count data. Our work, based on evaluation outcomes from two different test datasets, provides evidence that such predictive capabilities are feasible.

REFERENCES

[1] M. Á. Martínez-González, J. A. Martínez, F. B. Hu, M. J. Gibney, and J. Kearney, "Physical inactivity, sedentary lifestyle and obesity in the European Union," Int. J. Obes., vol. 23, no. 11, pp. 1192–1201, 1999.

[2] G. Hu, P. Jousilahti, N. C. Barengo, Q. Qiao, T. A. Lakka, and J. Tuomilehto, "Physical activity, cardiovascular risk factors, and mortality among finnish adults with diabetes," Diabetes Care, vol. 28, no. 4, pp. 799–805, Apr. 2005.

[3] "Physical Activity/Exercise and Diabetes," Diabetes Care, vol. 27, no. SUPPL. 1. American Diabetes Association Inc., pp. s58–s62, Jan. 2004.

[4] T. C. Camacho, R. E. Roberts, N. B. Lazarus, G. A. Kaplan, and R. D. Cohen, "Physical Activity and Depression: Evidence from the Alameda County Study," Am. J. Epidemiol., vol. 134, no. 2, pp. 220–231, 1991.

[5] K. L. Piercy et al., "The physical activity guidelines for Americans," JAMA - J. Am. Med. Assoc., vol. 320, no. 19, pp. 2020–2028, Nov. 2018.

[6] C. Tudor-Locke et al., "How many steps/day are enough? for adults," Int. J. Behav. Nutr. Phys. Act., vol. 8, no. 1, p. 79, Jul. 2011.

[7] C. Tudor-Locke, C. L. Craig, J. P. Thyfault, and J. C. Spence, "A step-defined sedentary lifestyle index: <5000 steps/day," Applied Physiology, Nutrition and Metabolism, vol. 38, no. 2. NRC Research Press, pp. 100–114, Feb. 2013.

[8] R. H. Olsen, R. Krogh-Madsen, C. Thomsen, F. W. Booth, and B. K. Pedersen, "Metabolic responses to reduced daily steps in healthy nonexercising men," JAMA - Journal of the American Medical Association, vol. 299, no. 11. American Medical Association, pp. 1261–1263, Mar. 19, 2008.

[9] R. Krogh-Madsen et al., "A 2-wk reduction of ambulatory activity attenuates peripheral insulin sensitivity," J. Appl. Physiol., vol. 108, no. 5, pp. 1034–1040, May 2010.

[10] C. R. Mikus, D. J. Oberlin, J. L. Libla, A. M. Taylor, F. W. Booth, and J. P. Thyfault, "Lowering physical activity impairs glycemic control in healthy volunteers," Med. Sci. Sports Exerc., vol. 44, no. 2, pp. 225–231, Feb. 2012.

[11] Å. Ernersson, F. H. Nystrom, and T. Lindström, "Long-term increase of fat mass after a four week intervention with fast food based hyper-alimentation and limitation of physical activity," Nutr. Metab., vol. 7, no. 1, p. 68, Aug. 2010.

[12] A. Bagnall, J. Lines, A. Bostrom, J. Large, E. Keogh, and J. B. Fuernkranz Anthony Bagnall, "The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances," Data Min Knowl Disc, vol. 31, pp. 606–660, 2017.

[13] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P. A. Muller, "Deep learning for time series classification: a review," Data Min. Knowl. Discov., vol. 33, no. 4, pp. 917–963, Jul. 2019.

[14] J. C. B. Gamboa, "Deep Learning for Time-Series Analysis," arXiv, Jan. 2017.

[15] Q. He and E. O. Agu, "Towards sedentary lifestyle prevention: An autoregressive model for predicting sedentary behaviors," in International Symposium on Medical Information and Communication Technology, ISMICT, Jun. 2016, vol. 2016-June.

[16] T. J. Ulrych and R. W. Clayton, "Time series modelling and maximum entropy," Phys. Earth Planet. Inter., vol. 12, no. 2–3, pp. 188–200, Aug. 1976.

[17] R. Wang et al., "Studentlife: Assessing mental health, academic performance and behavioral trends of college students using smartphones," in UbiComp 2014 - Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Sep. 2014, pp. 3–14.

[18] Q. He and E. O. Agu, "A Rhythm Analysis-Based Model to Predict Sedentary Behaviors," in Proceedings - 2017 IEEE 2nd International Conference on Connected Health: Applications, Systems and Engineering Technologies, CHASE 2017, Aug. 2017, pp. 383–391.

[19] G. Ozogur, M. A. Erturk, and M. A. Aydin, "Prediction of physical activity times using deep learning method," in Lecture Notes in Electrical Engineering, 2019, vol. 504, pp. 299–307.

[20] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, Oct. 2001.

[21] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Aug. 2016, vol. 13-17-August-2016, pp. 785–794.

[22] R. Polikar, "Ensemble Learning," in Ensemble Machine Learning, Springer US, 2012, pp. 1–34.

[23] R. Furberg, J. Brinton, M. Keating, and A. Ortiz, "Crowd-sourced Fitbit datasets 03.12.2016-05.12.2016," May 2016.

[24] Xie, Junqing et al. "Evaluating the Validity of Current Mainstream Wearable Devices in Fitness Tracking Under Various Physical Activities: Comparative Study." JMIR mHealth and uHealth vol. 6,4 e94. 12 Apr.

[25] E. Wood et al., "Diabetes Mobile Care: Aggregating and Visualizing Data from Multiple Mobile Health Technologies.," AMIA Jt. Summits Transl. Sci. proceedings. AMIA Jt. Summits Transl. Sci., vol. 2019, pp. 202–211, 2019.

[26] G. Zhao, Z. Zhang, H. Guan, P. Tang, and J. Wang, "Rethinking ReLU to Train Better CNNs," in Proceedings - International Conference on Pattern Recognition, Nov. 2018, vol. 2018-August, pp. 603–608.