

# Personalized Effect of Health Behavior on Blood Pressure: Machine Learning Based Prediction and Recommendation

Po-Han Chiang and Sujit Dey

Mobile Systems Design Lab, Dept. of Electrical and Computer Engineering, University of California, San Diego

Email: {pochiang, sdey}@ucsd.edu

**Abstract** — Blood pressure (BP) is one of the most important indicator of human health. In this paper, we investigate the relationship between BP and health behavior (e.g. sleep and exercise). Using the data collected from off-the-shelf wearable devices and wireless home BP monitors, we propose a data driven personalized model to predict daily BP level and provide actionable insight into health behavior and daily BP. In the proposed machine learning model using Random Forest (RF), trend and periodicity features of BP time-series are extracted to improve prediction. To further enhance the performance of the prediction model, we propose RF with Feature Selection (RFFS), which performs RF-based feature selection to filter out unnecessary features. Our experimental results demonstrate that the proposed approach is robust to different individuals and has smaller prediction error than existing methods. We also validate the effectiveness of personalized recommendation of health behavior generated by RFFS model.

**Index Terms**— Machine Learning; Blood Pressure; Random Forest, Wearables

## I. INTRODUCTION

Health behavior (e.g. exercise, nutrition and sleep) is widely acknowledged as having key impact on human health condition. Traditionally, such relationship between health behavior and health condition indicator (e.g. blood pressure (BP) and glucose level) is studied through clinical trials in ambulatory settings. Popular wearables, such as Fitbit and Apple Watch, collect a great amount of continuous sensor data such as heart rate and steps count, which can provide detailed health behavior such as duration and quality of exercise and sleep. However, the potential of wearables for medical studies has not been utilized due to the lack of connection to health condition either from self-measurement or electronic medical records.

BP is the pressure of circulating blood on the walls of blood vessels and a direct indicator of hypertension, one of the most prevalent chronic disease in the world. BP is usually expressed in terms of the systolic pressure (SBP, maximum BP during one heart beat) and diastolic pressure (DBP, minimum BP in between two heart beats) and is measured in millimeters of mercury (mmHg). For accurate diagnosis and treatment of hypertension, regular BP measurement is necessary. The traditional cuff-based BP measurement method using mercury sphygmomanometers is inconvenient for constant measurement [1]. Although there has been a huge attention to continuous BP monitoring [2], the accuracy and cost limit the viability of such methods. Alternatively, with the

information of historical BP and health behavior, the prediction of BP will provide users a quick and convenient way to understand their future health condition. Sleep and exercise are proved to be statistically correlated with BP with randomized controlled trials [3, 4]. However, the personalized effect of health behavior, that is which of the health behavior factors has most important effect on an individual's BP level, has not yet been studied.

In this paper, we propose to use machine learning (ML) techniques to construct a personalized model to predict daily BP using an individual's historical BP and health behavior, and estimate the effect of the individual's health behavior on his/her BP. To the best of our knowledge, this is the first work investigating daily BP prediction and its relationship between health behavior data collected by wearables. Besides numerical prediction of daily BP, we propose to investigate the personal effects of health behavior on daily BP with the relative importance among different predictor variables (features) generated by the proposed ML model. With this information, we aim to provide personalized actionable insight to individuals and their healthcare providers to improve BP through sleep and exercise.

### A. Related Work

Prediction of BP level is proposed in [5-8]. Artificial neural network is used in [5] to predict SBP using contextual data; however, the prediction is static and cannot be applied to continuous BP prediction. In [6,7], pulse transit time (PTT) is used to predict short-term BP level with tree-based models [6] and neural network models [7]. However, PTT is only applicable for very short time horizon prediction (less than 10 minutes) while our technique is proposed to predict BP one day ahead, which allows actionable behavior change. The authors in [8] attempt to solve the temporal dependency between BP level and contextual data using neural network models. The data in [8] is measured daily and then averaged on monthly basis. The potential of temporal dependency of data is therefore not fully utilized due to insufficient temporal resolution and information loss in averaging or accumulation process of input data. Most importantly, none of the above research jointly use the BP data and wearable devices data to make the prediction, which provides timely information (BP prediction and health behavior effect) for users to plan their health behavior accordingly to improve their BP.

Machine learning has been widely used in various healthcare applications, like risk detection from medical

images [9], disease diagnosis [10], and health condition estimation from electronic medical records [11]. There are also multiple studies in diagnosing hypertension using homographic or contextual data (e.g. gender, body-mass index and alcohol habit) [12] or data from wearables [13]. In [10], many deep learning (DL) techniques have been applied on medical data, such as multilayer perceptron (MLP), recurrent neural networks (RNNs) and Long Short-Term Memory (LSTM), etc. There are three main challenges using DL: 1) The black-box nature of DL makes it hard to understand the correlation and effects between features and BP while they are as important as predicted outcome in medical study, 2) supervised DL generally requires a significant amount of labeled examples to achieve satisfying performance, but building such databases is costly and labor-intensive in most medical applications, 3) the performance largely depends on the setting of parameters of the DL models, which makes the optimization process opaque and intractable.

Random forest (RF) algorithm, introduced by [14], is one of the most popular methods based on aggregating a large collection of tree-based estimators. RF is known to work well for high-dimensional and small-sample problems. Moreover, RF provides measures of the relative importance of the features with respect to the prediction of the target variable. It has been shown that the importance measure is an efficient tool for selecting relevant variables [15].

Considering the above advantages, we choose RF as our candidate machine learning model. To enhance the state-independent RF model by utilizing the temporal correlation of BP, we extract historical values and moving average of BP as features to enhance the performance. We also examine the feature importance obtained from RF and propose a stable feature selection method to remove redundant and irrelevant features. We show that our proposed technique outperforms other popular ML methods. Moreover, we show that the proposed personalized BP model achieves better result than the aggregated model based on a larger but non-personalized dataset. Finally, we conduct experiments to validate the personal effect of health behavior on BP suggested by our proposed model.

The rest of the paper is organized as follows. In section II, data collection, feature extraction, and theoretical background of RF are described and the proposed RF with Feature Selection (RFFS) is presented. In section III, the performance of the proposed algorithm is evaluated. We also validate the personalized health behavior recommendation suggested by the proposed RFFS model. Finally, we conclude the paper in section IV.

## II. PROPOSED METHOD

### A. Data Acquisition and Representation

The data set that we used for evaluating the proposed method was collected from self-tracking experiment on 8 participants for 90 days using Fitbit Charge HR and Omron Evolv wireless BP monitor. The subjects include 4 males and 4 females and their age ranges from 25 to 79. The subjects were randomly selected and we did not intend to draw any gender or age specific conclusions. Since sleep and exercise data is sampled at 1 Hz and BP is measured twice (morning and evening) daily, the dataset consists time series with mixed sampling rates. To model the target variables (SBP and DBP) in terms of the features over time, we aggregated the sleep and exercise data on daily basis.

The target and predictor variables (features) are summarized in Table I. Our objective is to predict daily BP level using historical measurement of BP as well as exercise and sleep on the previous day. In addition to steps count and distance, exercise data includes calories burned and different levels of active time is obtained from Fitbit API based on heart rate and metabolic equivalent, a physiological measure expressing the energy cost of physical activities. Sleep data includes minutes asleep, minutes awake, awakening count, bed time and wake time. Bed time and wake time refer to the time participants go to bed and wake up respectively and are converted to numerical values. Each sample consists of one measurement of BP in the morning or the evening. The samples with any missing data were omitted. After filtering, there were at least 75 days for each user in the dataset and totally 1059 samples are used.

### B. Random Forest

There may be potentially redundant and irrelevant features in the proposed dataset either from raw data (e.g. steps count and distance) or derived features (e.g. historical values of BP). Moreover, the relevant features set may differ from person to person, so it is not advisable to select useful feature with pre-defined rules. RF has been proved to be robust with redundant and irrelevant features [Brei]. Furthermore, RF only has few main parameters to be tuned and it is not sensitive to these parameters [liaw], which facilitates the application of our proposed technique to construct personalized prediction model of BP.

#### *Preliminaries: regression tree*

Decision tree is a non-parametric and tree-like model used in machine learning problems. Decision tree is built by learning simple decision rules inferred from the input data, and it consists of three kinds of nodes: 1) root nodes, which represent all samples that are ready to be split later, 2) decision nodes, which split the samples into multiple sub-trees or leaf nodes based on decision rules, 3) leaf nodes, where no more split is performed. If the target variable is continuous, we call the

TABLE I. Target and Predictor Variables

| Target Variables | Original Predictor Variables (Features)                                                                                                                                                                                              | Derived Predictor Variables (Features)                                                                          |
|------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|
| SBP, DBP         | Heart rate, calories burned, steps, distance, floors, sedentary minutes, lightly active minutes, fairly active minutes, very active minutes, exercise calories, minutes asleep, minutes awake, awakening count, bed time, wake time, | Historical BP ( $BP_{t-1}, BP_{t-2} \dots, BP_{t-k}$ ), $EWMA_{BP}$ , days (categorical), morning (categorical) |

decision tree a regression tree. At each level of the regression tree, the best feature and its split threshold are chosen to separate the samples into distinct classes. The choice of the feature and the split point leads to the greatest reduction in variance for the prediction of the target variable. The estimated value of the target variable is the mean value of all observed samples in that leaf node. In other words, if a new data observation falls into that node, its prediction of the target variable is the mean value of all training samples in that class.

RF is an enhanced approach of decision trees which usually over-fit the data and result in high variance. The main principle is using a group of decision trees to form a stronger prediction model. To illustrate RF, we will first introduce bootstrap aggregation.

Bootstrap samples are datasets randomly drawn with replacement from the training data, and each bootstrap sample is the same size as the original training set. Bootstrap aggregation, or namely bagging, averages the prediction learned from a set of bootstrap samples, thereby reducing its variance [16]. Moreover, since each tree generated in bagging is identically distributed, the expectation of an average of such trees is the same as the expectation of any one of them. Therefore, the bias of bagged trees remains unchanged as that of the individual trees [16]. This leads to a decrease in overall mean squared error, the sum of squared bias and variance. For regression, we train many independent regression trees based on bootstrap versions of the training data, and then average the prediction result.

In addition to bagging, RF further improves the variance by reducing the correlation between the trees without substantial increase of the bias [14]. This is achieved by feature bagging, which randomly selects a subset of input features when constructing each tree. In practice, RF generalizes well with both categorical and numerical input features, with minimal parameter tuning required.

### C. Feature Extraction for Time-series Prediction

In regression problems, a sample consists of observations of the target variable and the respective features. It is assumed that the samples are independently generated and used for training. However, we can utilize potential temporal dependency of BP time series by feature extraction specifically for time-series prediction. Time series prediction problems include a set of time-ordered observations of a variable,  $y_i \in Y, i = 1, 2 \dots t$  where  $y_i$  is the value of target variable  $Y$  measured at time  $i$ ,

and the task is defined as trying to predict the future values of  $y_u$  for time stamps  $u > t$ . In our problem we set  $u = t + 1$ . The features used for predicting the future values of  $Y$  are usually the most recent observations of  $Y$  (e.g.  $y_t, y_{t-1} \dots, y_{t-k}$ ), based on the assumption that there exists correlation between successive observations of the series. If  $k$  is appropriately chosen, it's possible to capture the dynamics of the time series. In addition to historical values of BP, we generate exponentially weighted moving average (EWMA) of BP series to capture the increasing or decreasing trend of BP series. Finally, we create two categorical features, which are named as Day and Morning, to capture the potential daily and weekly periodicity of BP time series. The above two categorical features are processed with one hot encoding to perform binarization of the category. These derived features are summarized in Table I.

### D. RF-based Feature Selection

Although RF is robust with regards to redundant and irrelevant features, high-dimensional data will degrade the performance especially when the dataset is not correspondingly large. This problem can be solved by selecting only a subset of the original feature set before training the primary RF model. Feature importance, which is computed by measuring how effective the feature is at reducing prediction variance when creating decision trees within RFs, is often used to rank and select features. However, the feature importance generated by a single run of the RF model varies due to the nature that samples and features are randomly selected for use in constructing each tree. To obtain accurate and stable ranking of features, we can train the RF model multiple times and average the feature importance [17]. Therefore, we propose a RF-based feature selection technique which uses an additional model consisting of multiple RF runs to sort and select important features. For the problem with feature set  $X$  and target variables  $Y$ , we define feature importance vector in  $j^{th}$  run of RF as  $I_{XY}(j)$  where  $j = 1, 2 \dots J$ . The average feature importance  $\bar{I}_{XY}$  can then be calculated by

$$\bar{I}_{XY} = \frac{\sum_{j=1}^J I_{XY}(j)}{J} . \quad (1)$$

Only the features with higher average feature importance will be used in the primary RF model. Here we use the median of average feature importance as the selection threshold. In our

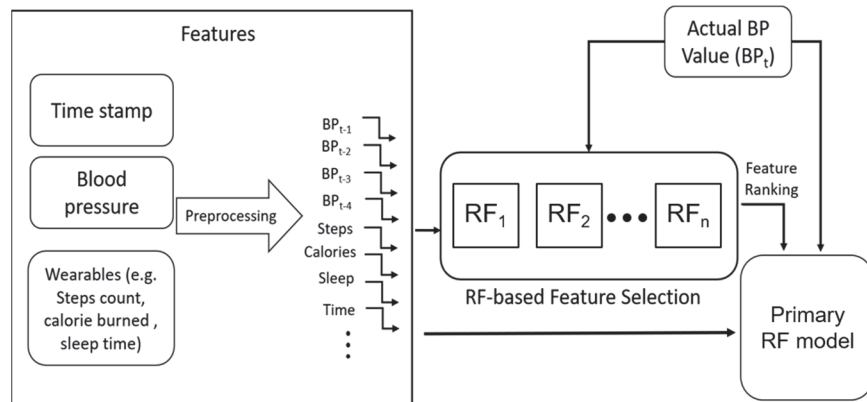


Figure 1. Block Diagram of proposed RF with Feature Selection (RFFS, training phase)

TABLE II. Comparison of MSE, MAE, and MAPE of different models

|      | SBP    |      |          | DBP   |      |          |
|------|--------|------|----------|-------|------|----------|
|      | MSE    | MAE  | MAPE (%) | MSE   | MAE  | MAPE (%) |
| SVM  | 64.06  | 6.30 | 5.26     | 44.72 | 5.30 | 6.74     |
| ANN  | 123.53 | 9.06 | 7.48     | 64.90 | 7.57 | 9.04     |
| LSTM | 115.47 | 8.63 | 7.16     | 62.00 | 6.23 | 8.12     |
| RF   | 57.82  | 5.75 | 4.80     | 45.98 | 5.08 | 6.69     |
| RFTS | 58.82  | 5.94 | 4.94     | 42.36 | 4.82 | 6.30     |
| RFFS | 47.33  | 5.18 | 4.31     | 37.45 | 4.30 | 5.64     |

proposed model, we use 5 RFs with 500 trees in our feature selection model since any larger number of RFs or trees achieves the same stable feature ranking. However, the selection of the number of RFs and threshold varies with different datasets and is out of scope of this paper. Integrated with RF-based feature selection, the block diagram of our proposed technique RF with Feature Selection (RFFS) is shown in Fig. 1.

### III. RESULTS AND DISCUSSION

In this section, we will discuss the experiment settings and results obtained by using the proposed RFFS algorithm and compare the results with existing methods. Finally, we will validate the effectiveness of personalized recommendation of health behavior generated by the RFFS model.

#### A. Experiment Setting

We implement and evaluate RFFS and other machine learning methods using the Scikit-learn library [18] and Keras [19] in python environment on an Intel i5 3.2GHz quad-cores and 16GB RAM computer. Mean square error (MSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) are calculated and used as our evaluation metrics. Their definitions are as follow:

$$MSE = \frac{\sum_{i=1}^n (P_{pred}^i - P_{actual}^i)^2}{n} \quad (2)$$

$$MAE = \frac{\sum_{i=1}^n |P_{pred}^i - P_{actual}^i|}{n} \quad (3)$$

$$MAPE = \frac{nMAE}{\sum_{i=1}^n |P_{actual}^i|} \times 100\% \quad (4)$$

We use 5-fold cross-validation to randomly split our data set into training (80%) and test (20%) sets 5 times and average the prediction results. We compare the performance of our

proposed approach RFFS with 1) Support Vector Machine (SVM), 2) Multilayer Perception (MLP), which is one of the most established ANN architectures, and 3) Long Short-Term Memory (LSTM), which is a popular DL model exploiting temporal dependency of data, as described in Section I. To show the effectiveness of the proposed feature extraction and selection, we also compare the result using RF with only original features (termed as RF) and RF with derived time-series related features but without feature selection (termed as RFTS). For all RF models, we set the number of trees as 500 and the minimum number of samples required to split an internal node as 2. Both MLP and LSTM models were trained using batches of size 20 and the Adam optimizer [20]. Due to limited training samples of our dataset, the maximum depth of both MLP and LSTM models was set as 4. We also use early stopping and insert dropout layers in both models with dropout rate equals 0.2 to avoid overfitting.

#### B. BP Prediction Results

The MSE, MAE and MAPE of BP prediction of our proposed method and other methods are summarized in Table II. Noted the value in Table II is the average MSE, MAE and MAPE of all participants and we will take a detailed look at individual participants later in this section. Table II shows that the proposed RFFS method produces the best prediction of BP in terms of MSE, MAE and MAPE. We observe that the error obtained from both DL models (ANN and LSTM) is among the highest of all methods. The higher forecasting error is possibly because the DL models overfits the data. Although LSTM performs slightly better than ANN, the result is still unsatisfactory possibly because LSTM cannot converge with the small training dataset (150~200 samples per participant) in our setup. Secondly, SVM and RF give similar performance and RF has the best result among existing methods. Note that although RFTS (using additional time-series but without feature selection) performs better than RF in terms of DBP, it

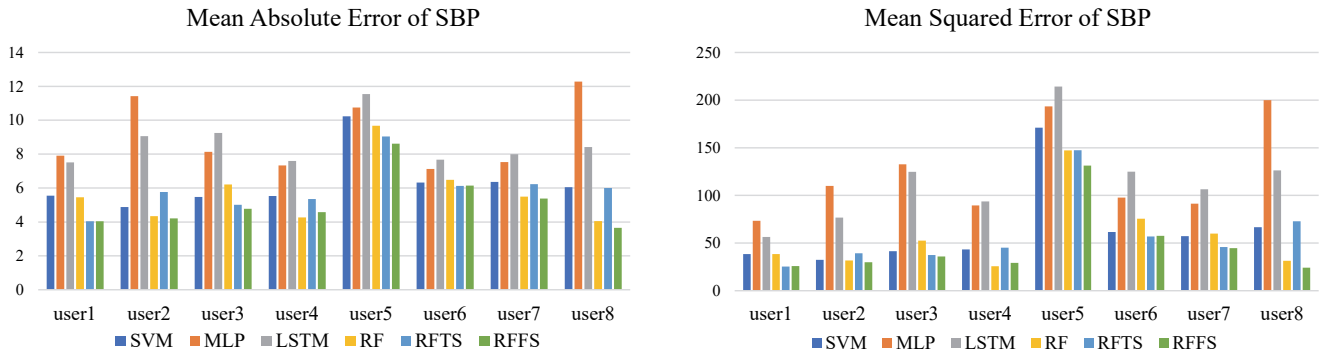


Figure 2. (a) left, MAE of SBP different participants and (b) right, MSE of SBP different participants

is worse than RF in terms features of SBP. However, our proposed RFFS (using both additional time-series features and feature selection) performs better than RF by 18.1% and 10.2% in terms of MSE and MAE of SBP; 18.6% and 15.4% in terms of MSE and MAE of DBP respectively. The above two observations show that including time-series related features in prediction is beneficial only with a proper feature selection technique.

Since SBP and DBP share similar trends of prediction performance, we use MSE and MAE of SBP to illustrate the personal prediction performance of each user, as shown in Fig. 2. Compared with the average prediction performance, we have the following observation: 1) RFFS consistently gives the best prediction except on user 4 where RFFS performs worse than RF by 12.4% and 6.8% in terms of MSE and MAE, 2) the prediction performance significantly varies with different participants. For example, the MAE of SBP ranges from 3.65 to 8.62 with RFFS, 3) including time-series related features has mixed effects among participants if feature selection is not used. For example, RFFS for user 1, 3, 5 and 6 is better than RF in terms of MSE and MAE, but it is worse for user 2, 4, 7 and 8. The above observation indicates that the effect of temporal correlation of BP time series is different from person to person. This also motivates us to propose our feature selection technique to remove redundant and irrelevant features.

TABLE III. Comparison of MSE and MAE models using personalized and aggregated dataset

|                    | SBP   |      | DBP   |      |
|--------------------|-------|------|-------|------|
|                    | MSE   | MAE  | MSE   | MAE  |
| Personalized model | 47.33 | 5.18 | 37.45 | 4.30 |
| Aggregated model   | 74.18 | 6.75 | 43.28 | 5.03 |

Finally, we compare the prediction performance between Personalized models and the Aggregated model using RFFS, as shown in Table III. Note that the MSE and MAE of personalized model are the average of all participants. The only difference of aggregate model is that the data of all participant are concatenated into a single dataset, which was then trained by one RF model. Personalized model performs better than Aggregated model by 36.2% and 23.3% in terms of MSE and MAE of SBP; 13.5% and 14.5% in terms of MSE and MAE of DBP respectively. Although Aggregate model has larger dataset and hence theoretically should have performed better than the Personalized model for each individual trained using his/her smaller dataset, the latter outperforms the Aggregate model since the effect of health behavior features on BP level varies across individuals.

TABLE IV. Top features and the weekly change of top 3 features and corresponding BP level

|                                                  | User 1                                                                                                                                         | User 2                                                                                                             |
|--------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------|
| Top 3 features (Normalized importance score)     | 1. Minutes sedentary (0.080611)<br>2. Calories burned (0.074208)<br>3. Minutes lightly active (0.070083)                                       | 1. Bed time (0.109519)<br>2. Minutes awake (0.102392)<br>3. Time in bed (0.090492)                                 |
| Change of the top features from week 1 to week 2 | 1. 21 minutes decrease in sedentary minutes<br>2. 152 calories increase in calories burned<br>3. 12 minutes increase in lightly active minutes | 1. 58 minutes earlier in bed time<br>2. 41 minutes increase in awake time<br>3. 66 minutes increase in time in bed |
| Average BP of the weeks (SBP/DBP)                | Week 1: 130.71 / 82.57<br>Week 2: 110.57 / 70.86                                                                                               | Week 1: 106.67 / 68.33<br>Week 2: 102.83 / 63.83                                                                   |

### C. Validation of Personalized Recommendation

To validate the effectiveness of feature importance of health behavior generated by RFFS model, we chose two participants and observed their feature importance score generated by RFFS. Our objective is to validate whether BP will change according to the change of top health behavior features (with the highest feature importance value) of users. Therefore, we exclude BP time-series features derived in Sec. II (C) since they are not health behavior features. We observe the BP level in two consecutive weeks for both participants who effectuate change in their top feature through change in their behavior.

In Table IV we list the top 3 features and their normalized importance score of user 1 and user 2. We can observe that the top 3 features are very different. BP is mainly correlated to exercise for user 1 and to sleep for user 2. Therefore, we suggest user 1 to increase exercise and suggest user 2 to go to bed earlier in week 2, as opposed to their normal exercise and sleep schedule in week 1. We summarize the changes of their top features in Table IV. The daily BP level in the two weeks is shown in Fig. 3, and the average BP levels for week 1 and week 2 are shown in Table IV. We can observe that the both SBP and DBP of user 1 decrease with more exercise, and with earlier bed time and more sleep for user 2. Although there is no exact conclusion indicating causal relationship of the top features of health behavior with the BP level of the participants since we use observational data instead of randomized controlled trials, the results indicate that identifying the top features that influence BP can be potentially used to provide participants with personalized recommendations of health behavior changes to improve and control their BP levels.

## IV. CONCLUSION

In this paper, we propose a data-driven model to investigate the personal effect of health behavior on BP using wearable devices and BP monitors. Our machine learning model can provide not only daily prediction of SBP and DBP but also importance score of health behavior factors on individual's daily BP. By extracting the time-series related data and integrating the RF based feature selection technique, we enhance the prediction performance of the original RF model. The experiment result shows that our technique outperforms other existing techniques in terms of MSE and MAE. Moreover, we show the significant change in BP after users changed their the most significant health behavior features suggested by our model, which validates the personalized

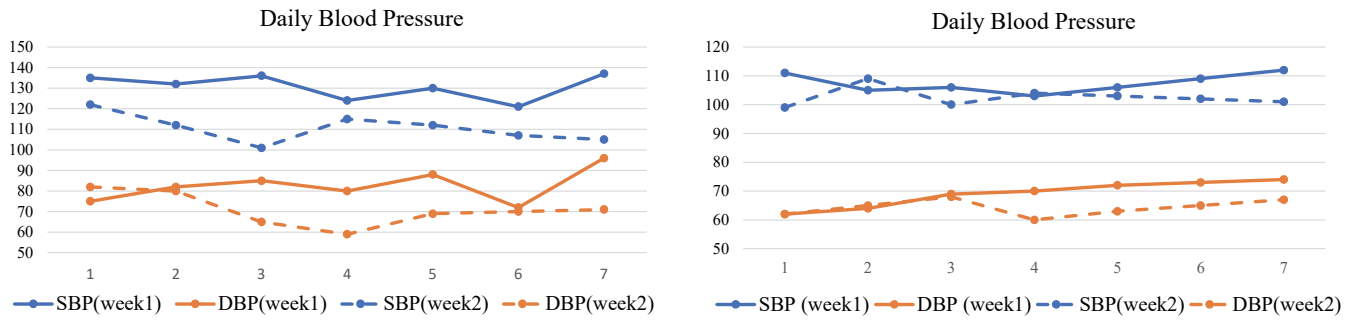


Figure 3. (a) left, daily BP of user 1 and (b) right, daily BP of user 2

recommendation on health behavior. In the future work, we aim to better utilize the high granularity of wearable data. This can be done either by raising the sampling frequency of BP or better representation of features to aggregate the wearables data. In addition to providing one day ahead prediction, we will also discover the long-term effect of health behaviors on BP.

#### REFERENCES

- [1] G. Odedegbe and T. Pickering, "Principles and techniques of blood pressure measurement," *Cardiol. Clinics*, vol. 28, no. 4, pp. 571–586, 2010.
- [2] R. Mukkamala et al., "Toward ubiquitous blood pressure monitoring via pulse transit time: Theory and practice," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 8, pp. 1879–1901, Aug. 2015.
- [3] N. Covassin and P. Singh, "Sleep duration and cardiovascular disease risk: epidemiologic and experimental evidence," *Sleep medicine clinics*, vol. 11, no. 1, pp. 81–89, 2016.
- [4] V. Cornelissen and N. Smart, "Exercise Training for Blood Pressure: A Systematic Review and Meta-analysis," *Journal of the American Heart Association*, 2013.
- [5] T. Wu, G. Pang and E. Kwong, "Predicting Systolic Blood Pressure Using Machine Learning," in *Proc. of International Conference on Information and Automation for Sustainability*, Colombo, Sri Lanka, 2014.
- [6] B. Zhang, Z. Wei, J. Ren, Y. Cheng and Z. Zheng, "An Empirical study on Predicting Blood Pressure using Classification and Regression Trees," in *IEEE Access*, doi: 10.1109, 2017.
- [7] P. Su, X. Ding, Y. Zhang, J. Liu, F. Miao and Ni Zhao "Long-term Blood Pressure Prediction with Deep Recurrent Neural Networks," to appear in *IEEE BHI 2018*.
- [8] X. Li, S. Wu, & L. Wang, "Blood Pressure Prediction via Recurrent Models with Contextual Layer," In *Proc. of the 26th International Conference on World Wide Web*, Perth, Australia, 2017.
- [9] Wernick, Miles N., et al. "Machine learning in medical imaging." *IEEE signal processing magazine*, vol. 27, no.4, pp. 25-38, 2010.
- [10] Lipton, Zachary C., et al. "Learning to diagnose with LSTM recurrent neural networks." arXiv preprint arXiv:1511.03677, 2015.
- [11] N. Hammerla, J. Fisher, P. Andras, L. Rochester, R. Walker and T. Ploetz, "PD disease state assessment in naturalistic environments using deep learning," In *Proc. of the 29th AAAI Conference on Artificial Intelligence*, Austin, Texas, USA, 2015.
- [12] H. Golino, L. Amaral and S. Duarte, "Predicting Increased Blood Pressure Using Machine Learning," *Journal of Obesity*, 2014.
- [13] B. Ballinger et al., "DeepHeart: Semi-Supervised Sequence Learning for Cardiovascular Risk Prediction," In *Proc. of the 32th AAAI Conference on Artificial Intelligence*, New Orleans, Louisiana, USA, 2018.
- [14] L. Breiman, "Random Forests," *Machine Learning*, 45(1):5–32, 2001
- [15] R. Genuer, J. Poggi and C. Tuleau-Malot, "Variable selection using random forests." *Pattern Recognition Letters*, 2010.
- [16] B. Efron and R. Tibshirani, "An introduction to the bootstrap," *CRC press*, 1994.
- [17] A. Behnamian, K. Millard, S. N. Banks, L. White, M. Richardson and J. Pasher, "A Systematic Approach for Variable Selection With RFs: Achieving Stable Variable Importance Values," in *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 11, pp. 1988–1992, Nov. 2017.
- [18] "scikit-learn Machine Learning in Python," 2015, <http://scikit-learn.org/>.
- [19] F. Chollet, "Keras," 2015.
- [20] D. Kingma and J. Ba, "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980*, 2014