

An mHealth Lifestyle Intervention Service for Improving Blood Pressure using Machine Learning and IoMTs

Jared Leitner
Department of Electrical &
Computer Engineering
University of California, San
Diego
San Diego, USA
jjleitne@eng.ucsd.edu

Po-Han Chiang
Department of Electrical &
Computer Engineering
University of California, San
Diego
San Diego, USA
pochiang@eng.ucsd.edu

Brian Khan
Department of Internal
Medicine
University of California, San
Diego
San Diego, USA
briankhan@gmail.com

Sujit Dey
Department of Electrical &
Computer Engineering
University of California, San
Diego
San Diego, USA
dey@eng.ucsd.edu

Abstract— In this paper, we present an AI-driven lifestyle intervention service for patients with hypertension. The automated intervention platform consists of a remote monitoring system that ingests lifestyle and blood pressure (BP) data and builds a personalized machine learning (ML) model to generate tailored lifestyle recommendations most relevant to each patient’s BP. Lifestyle data is collected from a wearable device and questionnaire mobile app which includes activity, sleep, stress and diet information. BP data is remotely collected using at-home BP monitors. With this data, the system trains random forest models that predict BP from lifestyle features and uses Shapley Value analysis to estimate the impact of features on BP. Precise lifestyle recommendations are generated based on the top lifestyle factors for each patient. To test the system’s ability to improve BP, we enrolled hypertensive patients into a three-armed clinical trial. During the 6-month trial period, our system provided weekly recommendations to patients in the experimental group. We evaluate the system’s effectiveness based on multiple BP improvement metrics and comparison with a control group. Patients in the experimental group experienced an average BP change of -4.0 and -4.7 mmHg for systolic and diastolic BP, respectively, compared to -0.3 and -0.9 mmHg for the control group. Our results demonstrate that the platform can effectively help patients improve their BP through precise lifestyle recommendations.

Keywords— *Blood pressure, hypertension, machine learning, personalized modeling, smart healthcare, remote patient monitoring, digital health, lifestyle medicine.*

I. INTRODUCTION

High blood pressure (BP), or hypertension, is one of the most prevalent chronic diseases in the world [1]. Hypertension management begins with lifestyle modification which, alone, can be effective in controlling BP [2-5]. Traditionally, relationships between lifestyle factors (e.g., activity, sleep, diet, stress) and BP have been investigated through large-scale Randomized Controlled Trials (RCTs). However, the aggregate findings from RCTs are generalized insights and not tailored for individuals. That is, these insights do not capture how the impact of specific lifestyle factors on BP may differ between individuals due to variations in physiology and genomic makeup. In addition, data from RCTs are usually collected in

healthcare settings or in a self-reported fashion. It is well-established that BP measurements obtained in healthcare settings are often unreliable [6], while self-reported data often falls short of accuracy and granularity.

In contrast, wearables such as Apple Watch, Fitbit and Samsung Galaxy Watch collect a great amount of lifestyle data in high granularity and continuity. With this data, personalized analytics can be carried out to elucidate relationships between BP and lifestyle factors at the individual level. In our previous work [7], we proposed a personalized, ML-based method to determine the top lifestyle factors impacting an individual’s BP. We utilized these top factors to provide personalized and precise insights to users, as opposed to general lifestyle recommendations. We conducted an experiment in which participants were randomized to either receive personalized lifestyle recommendations based on their data (experimental group) or not receive lifestyle recommendations (control group). We observed a significantly greater improvement in BP for the patients in the experimental group, demonstrating the potential of our recommendations to improve BP through precise lifestyle changes.

While the initial results are promising, there are 3 main limitations to our previous work: 1.) No diet or stress data was collected. Since diet and stress can have a significant impact on BP [8], it is important to consider these factors when providing personalized recommendations. 2.) Patients in the experimental group were only provided a one-time recommendation. Since an individual’s physiology changes over time, the top lifestyle factors impacting their BP, and therefore the correct recommendation, may also change. In addition, more frequent outreaches can keep patients better engaged with their health and lifestyle choices. 3.) We did not collect compliance data, which meant we could not determine whether or not patients followed our recommendations. This is a necessary step to validate if our personalized recommendations have a beneficial impact on BP.

In order to address these limitations, we propose an automated service, namely P3.AI, that uses remotely collected lifestyle and BP data to provide personalized, precise and proactive (P3) lifestyle interventions using artificial

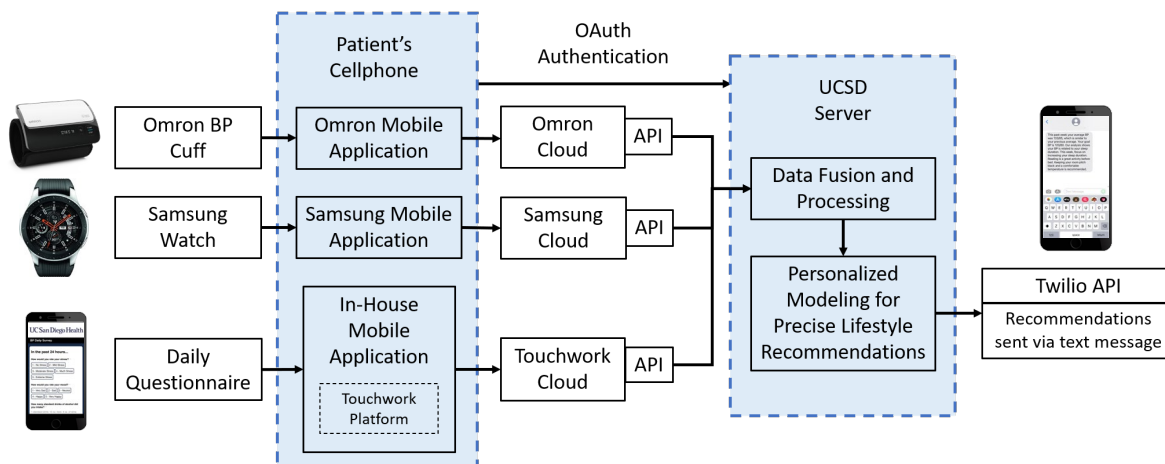


Figure 1. P3.AI remote monitoring and lifestyle intervention system architecture.

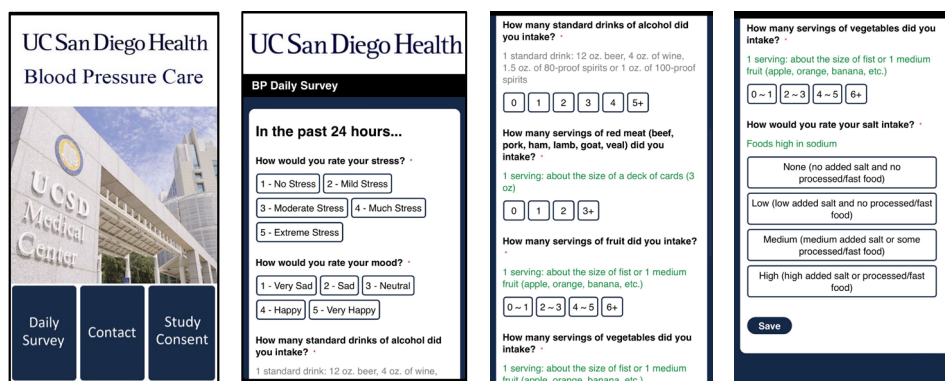


Figure 2. Questionnaire mobile app.

intelligence (AI) to patients with hypertension. In addition to BP and wearable device data, our remote monitoring system collects stress and diet information through a mobile questionnaire app. This data is provided as additional lifestyle features to improve the comprehensiveness of the personalized model. In order to provide ongoing recommendations and increase patient engagement with their health, the P3.AI service automatically sends a new lifestyle recommendation to patients every week based on their updated data. Every week, each patient’s personalized model is retrained with their new data using a 30-day rolling window and the top lifestyle features impacting their BP are updated. Based on their updated top lifestyle features, a new recommendation is sent.

To evaluate the effectiveness of the P3.AI service, we conducted a clinical trial in which patients with hypertension were enrolled. The trial was designed in a fully remote manner so that patients could participate from anywhere in the USA. We collected lifestyle and BP data from each patient and provided weekly lifestyle recommendations based on their personalized model. We monitored whether patients were compliant with our recommendations based on their lifestyle data before and after each recommendation. We compare the BP improvement results for patients that received weekly recommendations (experimental group 2) to that of the patients from our previous work who received a one-time

recommendation (experimental group 1) and no recommendations (control group).

The rest of the paper is organized as follows. In Section II, we present the design of our remote monitoring system and describe our machine learning approach for generating personalized and precise lifestyle recommendations for improving BP. In Section III, we detail our clinical trial design including the cohort statistics, control and experimental groups, and outcome measures. In Section IV, we present the clinical trial results, including BP improvement and recommendation compliance. We draw comparisons between the control group and two experimental groups. Finally, we conclude the paper in Section V.

II. SYSTEM DESIGN

In this section, we first detail our remote monitoring system architecture and the data collected. We then present our ML approach for generating personalized and precise lifestyle recommendations for improving BP.

A. System Architecture

Figure 1 displays the overall architecture of the P3.AI remote monitoring system. The system consists of a Samsung Galaxy Watch, an Omron Evolv wireless BP monitor and a questionnaire mobile app. The mobile app was developed using the Touchwork platform [9] and is displayed in Figure 2. Data

Table I. Daily questions in our mobile app. Diet questions are tailored for measuring information relevant to hypertension.

Question	Answer Options
Please answer the questions based on your past 24 hours of activities	
1. How would you rate your stress?	a. <i>No Stress</i> b. <i>Mild Stress</i> c. <i>Moderate Stress</i> d. <i>Much Stress</i> e. <i>Extreme Stress</i>
2. How would you rate your mood?	a. <i>Very Sad</i> b. <i>Sad</i> c. <i>Neutral</i> d. <i>Happy</i> e. <i>Very Happy</i>
3. How many standard drinks of alcohol did you intake?	a. 0 b. 1 c. 2 d. 3 e. 4 f. 5+
4. How many servings of red meat did you intake?	a. 0 b. 1 c. 2 d. 3+
5. How many servings of fruits did you intake?	a. 0-1 b. 2-3 c. 4-5 d. 6+
6. How many servings of vegetables did you intake?	a. 0-1 b. 2-3 c. 4-5 d. 6+
7. How would you rate your salt intake?	a. <i>None (no added salt and no processed/fast food)</i> b. <i>Low (low added salt and no processed/fast food)</i> c. <i>Medium (medium added salt or some processed/fast food)</i> d. <i>High (high added salt or processed/fast food)</i>

Table II. Lifestyle features used for personalized modeling.

Feature Categories	Features
Activity (n=6)	Steps, Floors, Walking/Running Speed, Sedentary Time, Lightly Active Time, Very Active Time
Sleep (n=7)	Sleep Duration, Bed Time, Wake Up Time, Light Sleep, Deep Sleep, REM Sleep, Sleep Awakeness
Heart Rate (n=4)	Max Active HR, Mean Active HR, Sleep HR, Sleep HR Fluctuation
Stress & Diet (n=7)	Stress, Mood, Alcohol, Red Meat, Fruits, Vegetables, Salt

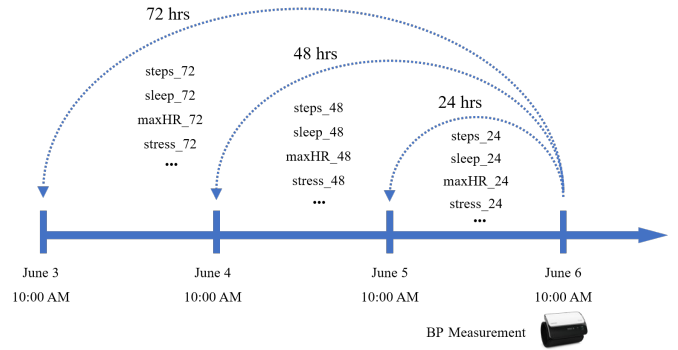


Figure 3. Lifestyle features are aggregated 24, 48 and 72 hours before every BP measurement.

was collected remotely through the application programming interfaces (APIs) provided by Samsung, Omron and Touchwork. Patients were asked to wear the Samsung device as often as possible, including during sleep, and take two BP measurements per day, once in the morning (8-10 am) and evening (7-9 pm). The primary metrics used to measure BP are systolic and diastolic blood pressure (SBP and DBP), which are defined as the maximum and minimum BP, respectively, during a heartbeat and measured in millimeters of mercury (mmHg). In addition, patients completed a daily questionnaire that asked about their stress, mood and dietary choices in the past 24 hours. These questions were developed in collaboration with physicians on our team and are detailed in Table I. The diet questions are tailored for measuring information relevant to hypertension including alcohol, red meat, fruits/vegetables and salt consumption. The P3.AI system integrates the data into a combined dataset for training personalized lifestyle-BP models and generating precise lifestyle recommendations for patients. These recommendations are sent to patients via text message using the Twilio API service [10] as displayed in Figure 1. Patients' BP and compliance data are collected after each recommendation for evaluation.

B. Data Description and Feature Engineering

The Samsung Galaxy Watch includes a heart rate monitor, accelerometer, ambient light sensor and barometer. The device uses these sensors to calculate various health parameters, including lifestyle and vitals measurements. Lifestyle features

include activity (steps, walking/running speed, floors climbed), sleep timing (duration, bed time, up time) and sleep stages (deep, light, REM, awake). The device records heartrate (HR) and steps data every minute which we use to calculate different levels of active time (sedentary, lightly active, very active). To do this, we first calculate three HR zones for each patient based on their maximum HR. Maximum HR (HR^{max}) of each patient is calculated as [11]:

$$220 - age \quad (1)$$

Three HR zones (zone 1, 2, and 3) are defined as [12]:

$$Z * HR^{max}, Z \in [.5, .7, 1] \quad (2)$$

We define the three active levels as follows: sedentary (steps < 10 or HR in zone 1), lightly active (steps ≥ 10 and HR is in zone 2), and very active (steps ≥ 10 and HR in zone 3) [7]. We also calculate the mean and standard deviation of HR during sleep. These features are designed to capture information about sleep quality. Stress, mood and diet features are calculated based on the daily questionnaire responses. Stress and mood features are encoded from 1 to 5 based on the Likert scale [13] response to questions 1 and 2 in the daily questionnaire. For example, "no stress" is encoded as 1 while "extreme stress" is encoded as 5. The alcohol, red meat, fruits and vegetables features represent the number of servings consumed for that day. The salt feature is encoded from 1 to 4 based on the response to question 7.

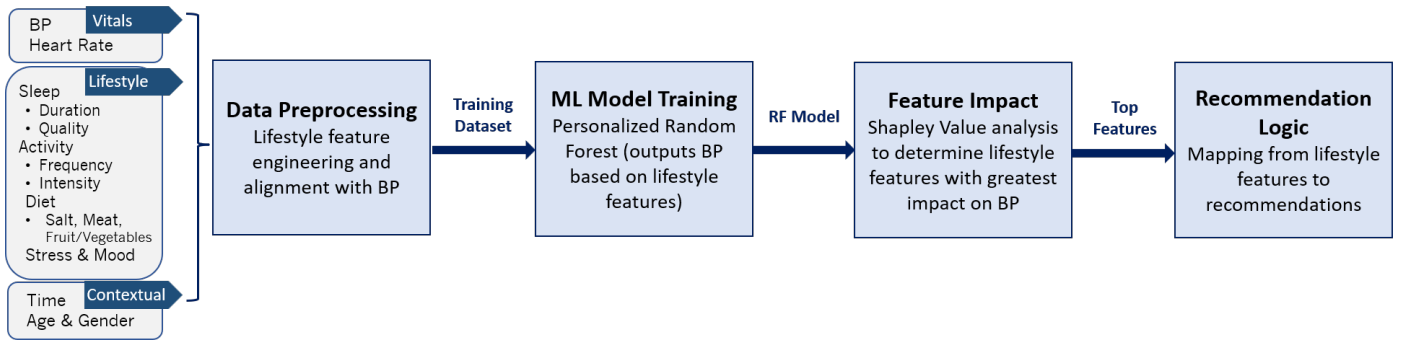


Figure 4. Block diagram of proposed method for generating personalized and precise lifestyle recommendations.

Table II presents all the lifestyle features along with which category each feature falls in.

Galaxy Watch data is recorded every minute while BP is measured by patients twice per day. As a result, the combined dataset consists of time series with varying frequencies. Moreover, although the guideline for this study is to measure BP in the morning (8-10 am) and at night (7-9 pm), there are missing measurements, variations in measurement time (e.g., measurements in the afternoon) and redundant measurements (e.g., two evening measurements at 8 pm and 9 pm). To address these varying frequencies and create our labeled dataset for model training, each feature is aggregated on a 24-hour, 48-hour and 72-hour basis before each BP reading as illustrated in Figure 3. Each feature is calculated as the average daily value in the 24/48/72 hours before each BP measurement. For example, the “steps_24”, “steps_48” and “steps_72” features are the average daily steps in the previous 24/48/72 hours before each BP reading. Similarly, the “sleep_24”, “sleep_48” and “sleep_72” features are the average sleep durations during the past 24/48/72 hours before each BP reading. The “stress_24”, “stress_48” and “stress_72” features represent the average stress score in the past 24/48/72 hours and the “red_meat_24”, “red_meat_48” and “red_meat_72” features represent the average daily red meat servings in the past 24/48/72 hours. The rationale for engineering features in this fashion is that a patient’s lifestyle choices in the days prior to each BP measurement are most relevant to that BP

measurement. We use this processed dataset to train our ML model and investigate which features have the most significant impact on BP prediction.

C. Personalized Modeling and Lifestyle Recommendations

Figure 4 presents an overview of our ML-based method for generating personalized and precise lifestyle recommendations. Random Forest (RF) is used as the ML model in our proposed method. In our previous work [7], we compared multiple ML models in terms of BP prediction error and demonstrate that RF achieves a low error for predicting BP based on lifestyle features. In this paper, we do not present the predictive performance of the RF model since we focus on the effectiveness of the intervention instead of numerical prediction of BP. RF is an ensemble model that aggregates a collection of decision trees in order to reduce overfitting and the resulting high variance in prediction [14]. RF is more robust to noisy features as compared to other models [15], meaning redundant or irrelevant features will not greatly impact performance. In addition, the RF model provides a high level of interpretability compared to other models, which is necessary for determining which lifestyle features have the greatest impact on the model’s BP prediction. A separate model is trained for each patient using their personal lifestyle-BP dataset.

We utilize Shapley Value analysis to determine which lifestyle features have the greatest impact on BP prediction. Shapley Value analysis is a model-agnostic interpretation



Figure 5. Lifestyle feature ranking and corresponding personalized and precise recommendation for two patients. Green/Red colored bars indicate that an increase in a feature’s value results in a lower/higher BP prediction.

method derived from game theory. Given a set of feature values and a trained ML model, the estimated Shapley value indicates how each feature contributes to the model’s prediction. We use the tree SHAP (SHapley Additive exPlanations) framework [16, 17], which is optimized for tree-based models, to interpret the predictions of the RF model. The SHAP values attribute to each feature the change in the expected model prediction when conditioning on that feature [18]. The recommendation we provide to patients is based on their personal model’s top Shapley Value feature. Shapley Value analysis also determines the directionality of each feature’s impact on BP prediction. Figure 5 displays the Shapley Value feature ranking and the corresponding lifestyle recommendation for two different patients. In Figure 5, a red colored bar indicates that an increase in the feature’s value results in a higher BP prediction. A green colored bar indicates that an increase in the feature’s value results in a lower BP prediction. The top feature for patient 1 is “speed_72” and increasing this feature value results in a lower BP prediction, as indicated by the green bar. As a result, our recommendation to the patient is to increase speed and intensity during activities. For patient 2, the top feature is “sleep_48” and increasing this feature value results in a lower BP prediction. The recommendation for this patient is to increase their sleep duration.

All lifestyle features map to a different recommendation based on the feature’s impact on BP prediction. Most lifestyle features are actionable, including activity, sleep timing, diet and stress related features. Certain sleep related features, including sleep stage (deep, light, REM) and sleep heartrate features are not directly actionable. For these features, our recommendation to patients was to focus on improving sleep hygiene. In addition, for some patients we observed counterintuitive relationships between activity related features and BP prediction. For example, the third top feature for patient 2 is “lightly_active_48” which has a red colored bar. This means that more light activity resulted in a higher BP prediction. For cases when more activity was associated with a higher BP, we recommended that the patient increase time spent doing restful activities. Note that the 24/48/72 features map to the same recommendation. For example, if either “steps_24” or “steps_48” is the top feature, the recommendation will be the same. This is because the multiple feature time frames are used for modeling purposes and not directly actionable for patients. For each recommendation, we monitor patient compliance based on their lifestyle data the week before and after the recommendation. In Sec. IV (B) we discuss patient compliance with our recommendations.

The significant difference in feature rankings in Figure 5 demonstrates that lifestyle factors have a varying impact on BP for different patients, motivating our use of personalized and precise lifestyle recommendations. The P3.AI service automatically retrains each patient’s personalized model using a 30-day rolling window and updates their top features and lifestyle recommendation on a weekly basis. In order to test the effectiveness of these recommendations, we recruited patients with hypertension for our clinical trial.

Table III. Cohort Statistics (n = 38)

Age (years, mean ± SD)	50.9 ± 13.1
# Men	23
# Women	15
Initial SBP (mmHg, mean ± SD)	127.1 ± 8.2
Initial DBP (mmHg, mean ± SD)	80.9 ± 6.9

III. CLINICAL TRIAL

In this section, we first describe our recruitment strategy and trial cohort. We then present the trial design including the experimental/control groups and outcome measures.

A. Recruitment and Cohort Statistics

Our clinical study was reviewed and approved by our university’s Human Research Protections Program, which operates Institutional Review Boards (IRBs). The study was in collaboration with our university’s healthcare system, with patient enrollment, onboarding and management conducted by the healthcare system’s clinical and translational research institute. The inclusion criteria required patients to be pre-hypertensive or have Stage I hypertension (SBP between 120-140/ DBP under 90 per ACC/AHA 2017 guidelines [19]) and not be taking any antihypertensive medications. The trial was designed in a fully remote manner so that patients could participate from anywhere in the USA. After patients were pre-screened through an online eligibility questionnaire, our research coordinator confirmed their BP and medication status with their healthcare systems. Eligible patients who consented were provided a Samsung Galaxy Watch and an Omron Evolv wireless BP monitor to collect their lifestyle and BP data for up to 6 months. In total, the study cohort consisted of 38 patients and Table III describes the cohort statistics. The average age of participants was 51 years and 61% were male. The average initial BP for patients was 127/81. The initial BP for each patient is calculated as the average of their first week of measurements. 25 of the 38 patients in our cohort are from our previous work [7]. Since then, we recruited an additional 13 patients to participate in our trial.

B. Trial Design

The objective of this trial is to assess the ability of the P3.AI service to improve patients BP through personalized and precise lifestyle interventions. We aim to compare the BP improvement

Table IV. Comparison of control vs. experimental groups.

	Control (n=19)	Experimental 1 (n=6)	Experimental 2 (n=13)
Provided devices and mobile app for data collection	X	X	X
Personalized lifestyle recommendation (one-time)		X	
Personalized lifestyle recommendations (weekly)			X

results for patients that received weekly recommendations to that of the patients from our previous work who received a one-time recommendation and no recommendations. In order to make this comparison, we divided the patients into a control group and two experimental groups, as summarized in Table IV. Experimental group 2 included the 13 new patients recruited since our previous work. These patients received weekly lifestyle recommendations based on their updated data and personal model. Our recommendations were sent via text message as shown in Figure 5. Each message included a summary of the patient’s BP progression for the current week in addition to the lifestyle recommendation. Experimental group 1 consists of 6 patients who received a one-time recommendation. The control group consists of 19 patients who did not receive any recommendations, but were provided with a BP cuff and Samsung watch for data collection. The control group received devices to ensure that the only difference between the control and experimental groups was our personalized lifestyle recommendations. Consistently measuring your BP and wearing the Samsung watch is in itself an intervention since it makes the patient more aware of their health and lifestyle choices. As a result, patients may implement beneficial lifestyle choices on their own. All patients received devices to control for this confounding factor and ensure a fair and accurate comparison between the control and experimental groups.

The primary outcome measure for this trial is the average change in SBP and DBP in the experimental vs. control groups. Average SBP and DBP during the first and last week of a patient's enrollment is used to calculate the SBP and DBP change for each patient. A secondary outcome measure is patient compliance with the personalized recommendations within experimental group 2. This is an important measure to validate if our personalized recommendations have a beneficial impact on BP. Data from the Samsung watch (steps, sleep duration, active minutes, etc.) and mobile app questionnaire (stress, mood, alcohol, etc.) is used to assess whether patients followed the lifestyle recommendations. We also sent patients a weekly questionnaire asking them to rate the difficulty of following their current recommendation on a Likert scale [13] of 1 (easy) to 5 (difficult). For certain recommendations, such as improving sleep hygiene and increasing time spent doing restful activities, measuring compliance solely based on device data is infeasible.

For these recommendations, we used patient responses to the recommendation difficulty questionnaire to assess compliance.

IV. RESULTS AND DISCUSSION

Next, we present the clinical trial results, including BP improvement and recommendation compliance. We draw comparisons between the control group and two experimental groups. Finally, we provide a discussion on future work.

A. Blood Pressure Improvement

Table V compares the BP improvement results for each trial group. These results include the change in mean BP, the percentage of patients with decreasing mean BP, the change in maximum BP, and the percentage of patients with decreasing maximum BP. The mean and maximum BP during the first and last week of the trial is calculated for each patient in order to determine these results. A mean change of -0.3 and -0.9 mmHg for SBP and DBP, respectively was measured in the control group, -3.8 and -2.3 mmHg in experimental group 1, and -4.0 and -4.7 mmHg in experimental group 2. Evidently, patients in experimental group 2 achieved the greatest improvement in mean BP during the trial period. Experimental group 2 experienced a 3.7 and 3.8 mmHg greater reduction in SBP and DBP, respectively, compared to the control group and a 0.2 and 2.4 mmHg greater reduction in SBP and DBP, respectively, compared to experimental group 1. While the difference in mean SBP change between the two experimental groups is small, the mean DBP change in experimental group 2 is significantly greater than experimental group 1. In addition to mean BP change, we compare the change in maximum BP since maximum BP has been shown to be a strong predictor of cardiovascular events, independently of mean BP [20]. The average changes in maximum SBP and DBP were -3.3 and -2.5 mmHg for the control group, -10.5 and -8.8 mmHg for experimental group 1, and -9.9 and -8.3 mmHg for experimental group 2. Experimental group 1 experienced a 0.6 and 0.5 mmHg greater reduction in maximum SBP and DBP, respectively, compared to experimental group 2. However, both experimental groups achieved a significantly greater reduction in maximum BP compared to the control group.

We also compare the percentage of patients with normal BP (SBP < 120 and DBP < 80 per ACC/AHA 2017 guidelines [19])

Table V. Comparison of BP change in the experimental and control groups.

	Control Group (n = 19)		Experimental Group 1 (n = 6)		Experimental Group 2 (n = 13)	
Age	52.9 ± 13.1		50.1 ± 15.0		48.4 ± 12.2	
Female ratio (%)	37%		33%		46%	
	SBP	DBP	SBP	DBP	SBP	DBP
Initial BP (mmHg)	127.2 ± 7.3	80.0 ± 5.9	125.4 ± 5.9	80.8 ± 6.7	127.5 ± 10.1	82.5 ± 7.8
Mean BP change (mmHg)	-0.3	-0.9	-3.8	-2.3	-4.0	-4.7
Subjects with decreasing mean BP (%)	9 (47%)	10 (53%)	5 (83%)	6 (100%)	9 (69%)	10 (77%)
Max BP change (mmHg)	-3.3	-2.5	-10.5	-8.8	-9.9	-8.3
Subjects with decreasing max BP (%)	12 (63%)	11 (58%)	6 (100%)	6 (100%)	11 (85%)	11 (85%)
Subjects with initial BP < 120/80 mmHg (%)	3 (16%)	14 (74%)	2 (33%)	2 (33%)	2 (15%)	6 (46%)
Subjects with final BP < 120/80 mmHg (%)	5 (26%)	14 (74%)	2 (33%)	5 (83%)	6 (46%)	9 (69%)

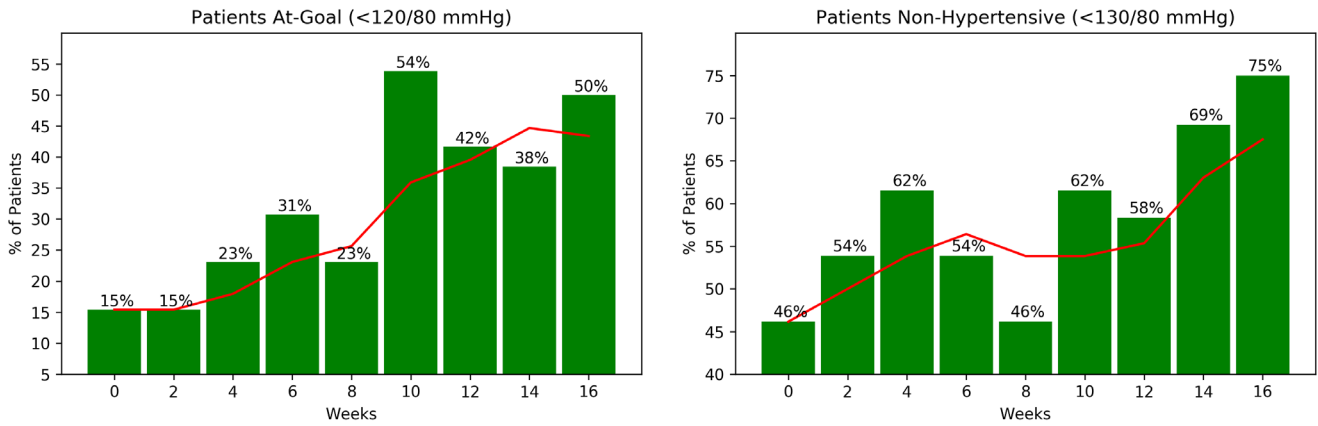


Figure 6. Progression of patients in experimental group 2 at-goal (<120/80 mmHg) and non-hypertensive (<130/80 mmHg). The red line represents a 6-week moving average (MA).

at the beginning and end of the trial in Table V. In experimental group 2, 15% and 46% of patients initially had their SBP and DBP, respectively, at normal levels. At the end of the study, 46% and 69% achieved normal levels for SBP and DBP, respectively. Evidently, there was a 31% increase in patients with normal SBP and a 23% increase in patients with normal DBP in experimental group 2. In experimental group 1, there was a 0% increase in patients with normal SBP and a 50% increase in patients with normal DBP. In the control group, there was a 10% increase in patients with normal SBP and a 0% increase in patients with normal DBP. Evidently, experimental group 2 experienced the greatest percentage increase in patients that reached normal BP levels during the trial. Only experimental group 1 experienced a greater percentage increase in patients with normal DBP. Since the number of patients in experimental group 1 is low, a small number of patients improving to normal BP levels will result in a large percentage increase. Overall, there is not one group that achieved the best results across all BP improvement metrics, Experimental group 2 achieved the greatest improvement in mean BP and percentage of patients achieving normal BP levels. We attribute this to the ongoing engagement with these patients through our weekly personalized lifestyle recommendations.

To investigate the BP progression for patients in experimental group 2, we calculate the percentage of patients at-goal (BP less than 120/80 mmHg) and non-hypertensive (BP less than 130/80 mmHg) on a bi-weekly basis. As displayed in

Figure 6, this analysis is carried out on the first 4 months of patient data since there are 3 patients who did not provide data for the last 2 months of the study. The red lines in Figure 6 represent a 6-week moving average (MA) and the first 2 days of patients' BP data are used to calculate the week 0 percentage. At the start of the trial, 15% and 46% of patients were at-goal and non-hypertensive, respectively. By the end of 4 months, these percentages increased to 50% and 75%, corresponding to a 35% and 29% increase in patients at-goal and non-hypertensive, respectively. While there are variations week to week in these percentages, the overall trend is positive as indicated by the moving average. These results demonstrate that the P3.AI service can help patients progress to healthier BP ranges through ongoing interactions with the patients.

To investigate the BP trend for patients in experimental group 2, we calculate the 30-day moving average for SBP and DBP. We first calculate the 30-day MA for each individual patient then average these to get the combined result. In Figure 7, we display the SBP and DBP 30-day MA for all patients (left), patients with decreasing mean BP (middle) and patients with increasing mean BP (right). Based on the 30-day MA for all patients, SBP and DBP steadily decreased during the first 1.5 months (6 weeks) of the study. After 6 weeks, the SBP and DBP stopped decreasing and oscillated at around 124 and 79 mmHg, respectively. Furthermore, we separate the patients based on whether they experienced a decrease or increase in mean BP

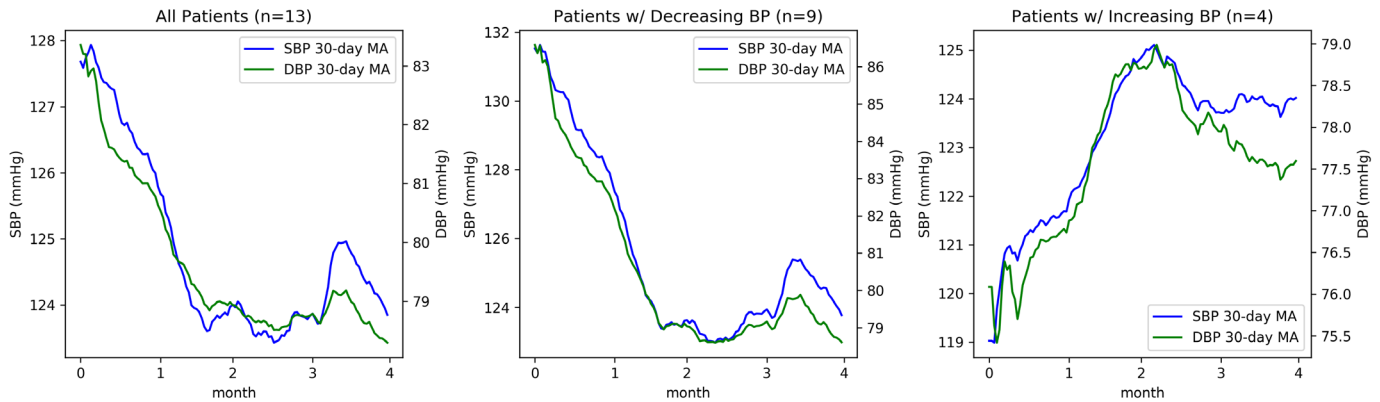


Figure 7. SBP and DBP progression for patients in experimental group 2. These plots display the 30-day moving averages (MA) for all patients (left), patients with decreasing mean BP (middle) and patients with increasing mean BP (right).

during the trial. For patients who experienced a decrease in mean BP, a similar trend to that of all patients can be seen. The average initial BP for these patients was 131/86 mmHg. On the other hand, for patients who experienced an increase in mean BP, the 30-day MA for SBP and DBP increased during the first 2 months of the study and then leveled off. Interestingly, the SBP and DBP for these patients stopped increasing and remained stable at around 124 and 78 mmHg after 2 months. The average initial BP was 120/76 mmHg for patients who experienced an increase in mean BP. Since an initial BP of 120/76 mmHg is already at normal levels, lifestyle intervention may not further reduce their BP. On the other hand, an initial BP of 131/86 mmHg is classified as Stage 1 hypertension [19]. These results indicate that patients who were initially hypertensive experienced an improvement in BP using the P3.AI service. Furthermore, the BP of these patients reached a healthy range in about 6 weeks and remained stable afterwards.

B. Recommendation Compliance

A secondary outcome measure for patients in experimental group 2 is their compliance with the personalized recommendations. Data from the Samsung watch (steps, sleep duration, active minutes, etc.) and mobile app questionnaire (stress, mood, alcohol, etc.) is used to assess whether patients followed the lifestyle recommendations. For each recommendation, we investigated the corresponding lifestyle feature the week before and after each recommendation. For example, if the recommendation was to increase daily steps, we calculated the average daily steps during the week before and after the recommendation. If step count was higher during the week after the recommendation, we marked the patient as compliant. For most recommendations it is possible to directly monitor compliance based on device data. However, for some recommendations it is not possible to monitor compliance based on a single lifestyle factor. These recommendations include improving sleep hygiene and increasing time spent doing restful activities. For example, many factors affect sleep hygiene (bedroom temperature, food/alcohol consumption before bed, daytime naps, etc.) therefore it is not possible to accurately measure based on one device feature. To assess compliance for these recommendations, we used patient responses to the recommendation difficulty questionnaire that was sent each week, as described in Sec. III (B). This questionnaire asked patients to rate the difficulty of complying with the current recommendation on a scale of 1 (easy) to 5 (difficult). If a patient responded with a 1 or 2, we counted the patient as compliant with the recommendation. The rationale is that patients are more likely to comply with recommendations that they find easier to follow.

A total of 204 lifestyle recommendations were sent to the patients in experimental group 2 during the trial duration. Of the 204 recommendations, 192 had sufficient data to assess compliance. In total, patients complied with 60% of these recommendations. Overall, this result indicates that the majority of our lifestyle recommendations were complied with. Sustained patient compliance and behavior change is a major challenge in healthcare. Precise lifestyle recommendations can improve patient compliance by focusing on a specific aspect of their lifestyle [21]. In order to investigate differences in compliance for different recommendations, we grouped recommendations into 5 categories: 1. Sleep, 2. Activity, 3. Restfulness, 4. Heart

Table VI. Compliance for different recommendation categories.

Recommendation Category	Count	Compliance
Sleep	69	58%
Activity	53	42%
Restfulness	39	74%
Heart Rate	21	71%
Stress & Diet	10	90%

Rate and 5. Stress & Diet. Stress and diet related recommendations are grouped into one category since this data is collected from the mobile app questionnaire. Table VI displays the number of recommendations sent and the compliance for each category. Evidently, the category with the greatest number of recommendations is sleep. Out of 69 total sleep related recommendations, 58% were complied with. The category with lowest compliance was activity, where 42% of recommendations were complied with. This may be due to the fact that increasing activity requires the greatest amount of effort compared to the other recommendations. The stress and diet category had the highest compliance of 90%.

C. Future Work

Our future work will involve enhancing the P3.AI system to be device agnostic. This will expand accessibility to the service by enabling patients with different wearable devices to receive personalized lifestyle recommendations most relevant to their BP. Since not all devices will collect the same lifestyle information, different feature engineering strategies will be required for different devices. A limitation of this trial is the relatively small number of subjects who received our personalized lifestyle recommendations. In addition, patients in this trial were pre-hypertensive or had Stage I hypertension (SBP between 120-140 and DBP under 90). In order to obtain a more robust result, we plan to conduct a larger trial and enroll patients with Stage II hypertension (SBP greater than 140 or DBP greater than 90). While the majority of our recommendations in this trial were complied with, we plan to implement additional measures to further improve compliance. One possibility is to gamify our recommendations, where patients earn rewards for following our recommendations.

V. CONCLUSION

In this paper, we propose an automated service, namely P3.AI, that uses remotely collected lifestyle and BP data to provide personalized, precise and proactive lifestyle interventions using AI to hypertensive patients. The P3.AI system trains a personalized random forest model to predict BP based on lifestyle factors for each patient. Shapley Value analysis is used to identify the most important lifestyle attributes impacting a patient's BP. Based on the top lifestyle factors, the system provides precise recommendations to improve the patient's BP. We investigated the effect of the P3.AI service by enrolling 38 patients into a clinical trial. Patients were divided into a control group and two experimental groups to compare the BP improvement results for patients that received personalized recommendations to those who did not. The trial results show

that significant improvement in BP can be achieved with personalized lifestyle recommendations. After receiving recommendations, patients in experimental group 2 decreased their BP by 4.0 and 4.7 mmHg for systolic and diastolic BP, respectively, compared to a decrease of 0.3 and 0.9 mmHg for patients in the control group who did not receive recommendations.

REFERENCES

- [1] C. Fryar, Y. Ostchega, C. Hales, G. Zhang, and D. Kruszon-Moran, "Hypertension prevalence and control among adults: United States, 2015–2016," *NCHS data brief*, no.289, 2017.
- [2] L. Appel, C. Champagne, D. Harsha, L. Cooper, E. Obarzanek, P. Elmer, et al. "Effects of comprehensive lifestyle modification on blood pressure control: main results of the PREMIER clinical trial," *Journal of the American Medical Association*, vol 289, pp. 2083–2093, 2003.
- [3] K. Doughty, N. Del Pilar, A. Audette, D. Katz, "Lifestyle Medicine and the Management of Cardiovascular Disease," *Current cardiology reports*, vol. 19, no. 11, pp. 116, 2017.
- [4] 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.
- [5] V. Cornelissen and N. Smart, "Exercise Training for Blood Pressure: A Systematic Review and Meta-analysis," *Journal of the American Heart Association*, 2013.
- [6] A. Siu, "Screening for high blood pressure in adults: US Preventive Services Task Force recommendation statement," *Annals of internal medicine*. vol. 163, no. 10, pp. 778-786, 2015.
- [7] P. -H. Chiang, M. Wong and S. Dey, "Using Wearables and Machine Learning to Enable Personalized Lifestyle Recommendations to Improve Blood Pressure," in *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 9, pp. 1-13, 2021.
- [8] M. Siervo, J. Lara, S. Chowdhury, A. Ashor, C. Oggioni and J. Mathers, "Effects of the Dietary Approach to Stop Hypertension (DASH) diet on cardiovascular risk factors: A systematic review and meta-analysis," *British Journal of Nutrition*, vol. 113, pp. 1-15, 2015.
- [9] *Touchwork*, www.touchwork.com.
- [10] R. Alagappan and S. Das, "Uncovering Twilio: Insights into Cloud Communication Services."
- [11] American Heart Association, "Target Heart Rates Chart," <https://www.heart.org/en/healthy-living/fitness/fitness-basics/target-heart-rates>
- [12] R. Elshawi, M. H. Al-Mallah and S. Sakr, "On the interpretability of machine learning-based model for predicting hypertension." *BMC medical informatics and decision making*, vol. 19, no. 1, pp. 146, 2019.
- [13] A. Joshi, S. Kale, S. Chandel and D. Pal, "Likert scale: Explored and explained," *British journal of applied science & technology*, vol. 7, no. 4, pp. 396, 2015.
- [14] L. Breiman, "Random Forest," *Machine Learning*, vol. 45, no.1, pp. 5–32, 2001.
- [15] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 67, pp. 93–104, 2012.
- [16] S. Lundberg and S. Lee, "A unified approach to interpreting model predictions," in *Advances in neural information processing systems*, pp. 4765-4774, 2017.
- [17] S. Lundberg, G. Erion, H. Chen, et al., "From local explanations to global understanding with explainable AI for trees," *Nat Mach Intell* 2, pp. 56–67, 2020.
- [18] N. Guthrie, M. Berman, K. Edwards, K. Appelbaum, S. Dey, J. Carpenter, D. Eisenberg and D. Katz, "Achieving Rapid Blood Pressure Control With Digital Therapeutics: Retrospective Cohort and Machine Learning Study," *JMIR cardio*, 3(1), e13030, 2019.
- [19] P. Whelton, et al., "Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines," *Journal of the American College of Cardiology*, vol. 71, no. 19, pp. 2273-2275, 2018.
- [20] P. Rothwell, S. Howard, E. Dolan, E. O'Brien, J. Dobson, B. Dahlöf, P. Sever and N. Poulter, "Prognostic significance of visit-to-visit variability, maximum systolic blood pressure, and episodic hypertension," *Lancet*, 375, pp. 895-905, 2010.
- [21] G. Chevance, O. Perski and E. Hekler, "Innovative methods for observing and changing complex health behaviors: four propositions," *Translational Behavioral Medicine*, vol. 11, no. 2, pp. 676-685, 2021.