This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2016.2518673, IEEE Journal of Biomedical and Health Informatics

Remote Health Monitoring Outcome Success Prediction using Baseline and First Month Intervention Data

Nabil Alshurafa*, *Member, IEEE,* Costas Sideris, *Student Member, IEEE,*
Mohammad Pourhomayoun, *Member, IEEE,* Haik Kalantarian, *Student Member, IEEE,*
Majid Sarrafzadeh, *Fellow, IEEE,* and Jo-Ann Eastwood, *Fellow, IEEE,*

***Abstract*—Remote health monitoring (RHM) systems are becoming more widely adopted by clinicians and hospitals to remotely monitor and communicate with patients while optimizing clinician time, decreasing hospital costs, and improving quality of care. In the Women's Heart Health Study (WHHS) we developed Wanda-CVD, where participants received healthy lifestyle education followed by six months of technology support and reinforcement. Wanda-CVD is a smartphone-based RHM system designed to assist participants in reducing identified cardiovascular disease (CVD) risk factors through wireless coaching using feedback and prompts as social support. Many participants benefitted from this RHM system. In response to the variance in participants' success we developed a framework to identify classification schemes that predicted successful and unsuccessful participants. We analyzed both contextual baseline features and data from the first month of intervention such as activity, blood pressure and questionnaire responses transmitted through the smartphone. A prediction tool can aid clinicians and scientists in identifying participants who may optimally benefit from the RHM system. Targeting therapies could potentially save healthcare costs, clinician and participant time and resources. Our classification scheme yields RHM outcome success predictions with an F-measure of 91.9%, and identifies behaviors during the first month of intervention that help determine outcome success. We also show an improvement in prediction by using intervention-based smartphone data. Results from the WHHS study demonstrates that factors such as the variation in first month intervention response to the consumption of nuts, beans and seeds in the diet help predict patient RHM protocol outcome success in a group of young Black women ages 25-45.**

***Index Terms*—Remote Health Monitoring; Prediction and Modeling; Outcome Success; Machine Learning**

I. Introduction and Motivation

REMOTE health monitoring (RHM) systems are increasingly proving to be effective in saving costs, reducing illness, and prolonging life [1]. However, studies on the efficacy of RHM systems continue to produce controversial results, as some studies report the benefits of RHM [2], [3], while others center on RHM failures [4]. Desai explains that it is likely that no single approach to health monitoring will be effective, and rather more patient personalization in RHM systems is necessary [4]. To be able to design systems that are more personalized, we need to identify predictive behaviors or patterns that determine RHM outcome success. As a result we're interested in predicting who will benefit from an RHM system, whether baseline contextual features are adequate for predicting outcome success, and how first month intervention data can improve prediction. Such a prediction tool can aid clinicians in identifying at an early stage participants more likely to capitalize on the technology provided, while at the same time identifying participants that might need further support.

In this paper, we describe an enhanced RHM system, Wanda-CVD, that is smartphone-based and designed to provide wireless coaching and social support to participants. CVD prevention measures are recognized as a critical target by health care organizations worldwide i.e. the World Health Organization, the Institute of Medicine and a primary goal for Healthy People 2020 [5]. In a six month study designed to reduce CVD risk factors in young black women, Wanda-CVD was deployed to about half (39) of the total study population (90).

In a previous paper [6] we described how to predict adherence in an RHM system exclusively using baseline contextual features. However, given the expense of purchasing smartphones and their resultant data/service plans, we were motivated to identify rationale for the successes and failures regarding participant outcomes. Based on this pilot study, we were interested in identifying a method to predict successful participants early on in the study. This effort could save time and resources but also help us learn how to mold our current health monitoring systems to better suit different populations. Because dropout rates increase with questionnaire length, developing such a prediction model could also aid in reducing the burden of participants by identifying important questions that relate to the success criteria of a study [7].

This paper is organized as follows. Section 2 describes the related work. Section 3 presents the Wanda-CVD system and the Women's Heart Health study in which it was tested. Section 4 describes the baseline questionnaires used in the study and Section 5 describes the features used on the sensor data from the first month of the intervention. Section 6 provides a definition of outcome success. Section 7 discusses our prediction methodology and how we go about testing it.

N. Alshurafa is with the Department of Preventive Medicine, Northwestern University, Chicago, 60611.

C. Sideris, M. Pourhomayoun, H. Kalantarian, and M. Sarrafzadeh are with the Wireless Health Institute, Department of Computer Science, University of California, Los Angeles, 90095.

J. Eastwood is with the School of Nursing, University of California, Los Angeles, 90095.

*Email correspondence to nabil@northwestern.edu

Manuscript received June 3, 2014; revised June 3, 2014.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2016.2518673, IEEE Journal of Biomedical and Health Informatics

Section 8 presents the results on the essential features selected by our framework and the accuracy of each predictive model. Finally, we conclude in Section 9.

II. Related Work

According to Bui et al., the potential for RHM systems to improve the management of heart failure patients is substantial [2]. A meta-analysis of RHM found that patients with heart failure that receive RHM has a 42% reduction in hospitalizations [8], and another study concluded that RHM can prevent up to 627,000 heart failure-related hospital readmissions each year [9].

Chaudhry et al. shows a high correlation between increases in body weight and hospitalization for heart failure beginning at least 1 week before hospital admission, however he concludes that remote health monitoring does not provide benefit [10]. However, Koehler et al. shows that compared to usual care, remote health monitoring had no significant effect on all-cause mortality or on cardiovascular death or HF hospitalization [11]. Desai also argues that home monitoring for heart failure does not necessarily improve patient readmission rates and outcomes [4].

Despite the increasing research in RHM systems, it remains to be seen whether the technical feasibility and effectiveness of such systems can generate optimal patient outcomes and prevent chronic disease in a cost effective manner [12]. In order to better assess who benefits from a RHM system, we must be able to predict which individuals are most likely to succeed. We must also identify the most effective features for predicting success in order to better understand the behaviors of the population we are learning.

Chronic conditions have been perceived as a unique market for the use of smartphone applications [13]. A recent review of over 60 studies found chronic conditions such as diabetes mellitus and cardiovascular disease in particular have always been perceived as a special ‘niche market’ for smartphone apps [14]. Krishna et al. found significant improvements in compliance to medical regimens using cell phones and text messaging interventions [15]. However, while smartphones are found to be useful in providing continuous data from a participant in a RHM system, we attempt to use this information to better understand our participant population.

Several RHM studies report patient characteristics of successful participants. However there is no research in the area of predicting outcome success based on a subset of patient contextual characteristics or intervention-based data from the first month. In this paper we identify key contextual features that help predict participant outcome success based on body mass index (BMI), waist circumference, low-density lipoprotein (LDL), and high-density lipoprotein (HDL), triglycerides (TG), and total cholesterol (TC). In a previous effort we attempted to predict adherence [6] and outcome success based on contextual-features [16]. However, this effort attempts to improve prediction of outcome success using smartphone data collected during the first month of the study intervention. In reviewing current literature, we find extensive research regarding adherence to medication prescription [17], however, we believe this is the first work in predicting outcome success to RHM systems.

III. Remote Health Monitoring System

A. Wanda-CVD

The Wanda-CVD RHM system, illustrated in Figure 1, is an advanced version of a previous RHM system we developed named Wanda [18], that targets patients at risk of CVD through Wireless Coaching. Wireless Coaching is in the form of automated messages prompting the user to take certain actions, such as measuring their blood pressure or reminding them to increase exercise intensity. There are several key components in the complete design of the Wanda-CVD system. The first component is the Android-based smartphone application designed as a means to collect data from the user, while displaying clinician feedback. Using embedded sensors, Bluetooth and Wi-Fi/Cellular network technology, the smartphone application can be programmed to connect to many stand-alone patient monitoring systems. The application then transmits this information to a backend server, where it is stored and machine learning algorithms process the data to identify patterns and learn patient models. The server provides a graphical user interface in the form of both a web- and tablet-based portal to the nurses, to provide a visual cue and summary of what is happening with each patient, alerting them when a matter requires their attention [19], [20], [21].

B. Women’s Heart Health study

The Women’s Heart Health Study is an IRB-approved clinical trial of 90 young black women aged 25-45 years that have a minimum of two risk factors for CVD. The goal of this study is to prevent or halt the progression of CVD in young black women through education regarding CVD risk reduction via lifestyle changes utilizing Wireless Coaching. In this study, 39 participants in the intervention group completed baseline screening and six month’s worth of intervention data (two were omitted due to significant missing data in month one). They received nutrition and lifestyle education, along with a Bluetooth blood pressure monitor and a smartphone worn around the waist to detect physical activity. The control group received usual care which included identification of CVD risk factors. However, this paper focuses solely on the intervention group. The system transmits participant measured data in real-time using Wi-Fi and 3G/4G technology. The intervention group received four educational classes focused on self-management of diet, nutrition, physical activity and stress reduction. After completion of baseline screening of cholesterol levels, blood pressure, BMI, demographic and psychosocial questionnaires and completion of the educational classes, the participants were taught how to wear and manage the phones and blood pressure monitors. They were told that the primary purpose of the smartphone was to track their physical activity while providing a user interface and a mechanism for automated feedback. The subjects were able to send/receive unlimited text messages along with unlimited data plans.

IV. Discrete Data: Baseline Contextual Features

During the face to face baseline, 3 and 6 months visits, physiological as well as psychological outcomes are measured via anthropometric measures, questionnaires and a software

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2016.2518673, IEEE Journal of Biomedical and Health Informatics

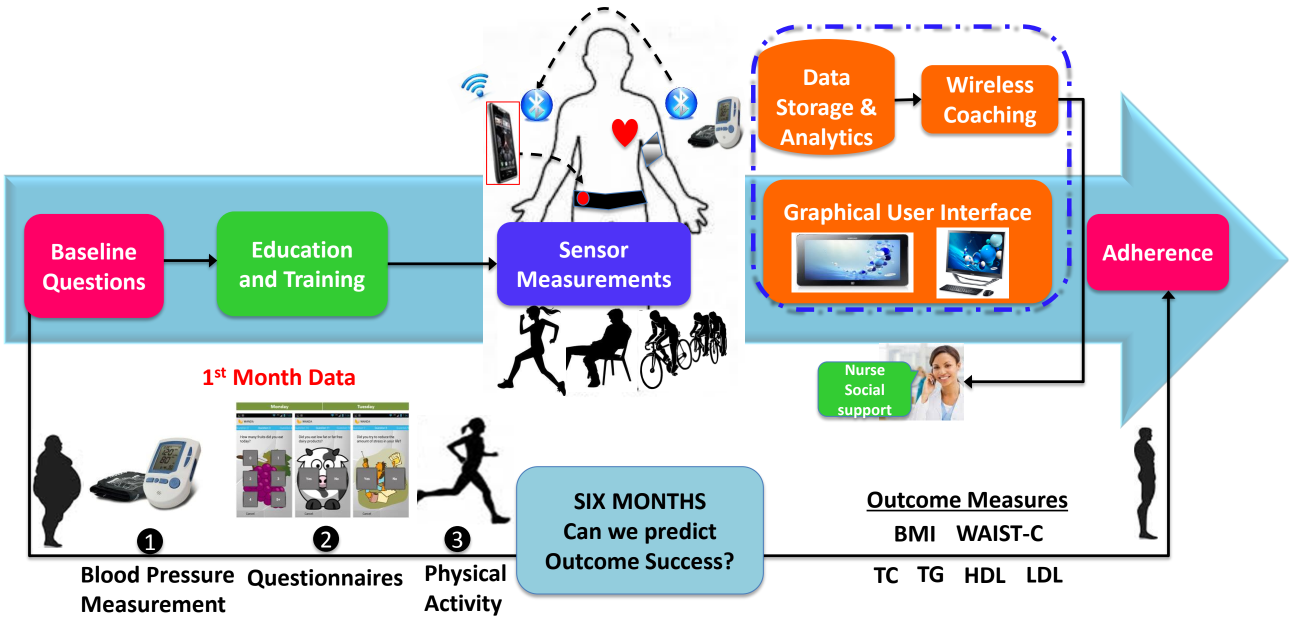

Fig. 1. Wanda-CVD System Architecture

dietary program. At baseline, the information in Table I was collected. The questionnaires are grouped into categories such as: family history (FAMHX), anxiety (BRIEFS), depressive symptoms (PHQ), quality of life (SF), stress levels (STRESS), perceived threat of heart disease (PMT), self-esteem (SLFEST), and social support group (SOCSUP). We also incorporate the following demographic information as features: age, education level, financial status, social status, employment status, and literacy. Our goal is to identify a subset of the measurements and questions that determine participant CVD study success.

TABLE I
BASELINE MEASUREMENTS AND QUESTIONNAIRES

Acronym	Measurements	Purpose
	Clinical Measures	Waist, BMI, BP, Lipids
FAMHX	Demographics-Health History	Family & Medical
BRIEFS	Brief Symptom Inventory	Anxiety
PHQ	Patient Health Questionnaire	Depressive Symptoms
MOSSAS	Medical Outcomes Study-SAS	Adherence
SLFEST	Rosenberg Self-Esteem Scale	Self Esteem
SF	MOS-SF-12	Quality of Life
PMT	Protection Motivation Theory	Health threat of heart disease self-efficacy
STRESS	INTERHEART STRESS	Stress
SOCSUP	Perceived Social Support Scale	Social Support

V. Continuous Data: Smartphone-based

We would also like to see whether the first month of smartphone intervention data can improve prediction of outcome success. We analyze the first month's data for the following categories on a weekly basis: physical activity, blood pressure, daily questionnaire and weekly questionnaire responses. Physical activity values included total activity level and time in low, medium and high activity. Blood pressure values included systolic and diastolic blood pressure. We then included the 12 daily and 12 weekly questionnaire responses. The daily questions were defined by a medical expert in order to reinforce the education received. The weekly questions follow the MOS-SF-12 weekly questions [22]. The total number of continuous variables per participant analyzed in this paper were 31. The percentage of missing data per person and per variable is less than 20%. There are several methods to handle missing data. The multiple imputation technique for example is one technique that minimizes the bias in the standard error. However, Downey et al. [23] has shown that since the percentage of missing data is less than 20% the single imputation method will not significantly deter the reliability of the data. As a result we handled the missing data using the single imputation (mean substitution) technique, where for each variable we averaged the values of both the outcome success and failure class to substitute in the missing data. During our cross-validation testing, we derive the mean values from the training set and apply them to the validation set. We then calculate for the first month of data 14 statistical functions as shown in Table II, totaling 420 features. Figure 3 illustrates a sample of the four different signals analyzed for each participant during their first month of intervention.

VI. Class Labels: Outcome Success

A total of 53 participants initially enrolled as part of the intervention group. However, 39 continued until the end of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2016.2518673, IEEE Journal of Biomedical and Health Informatics

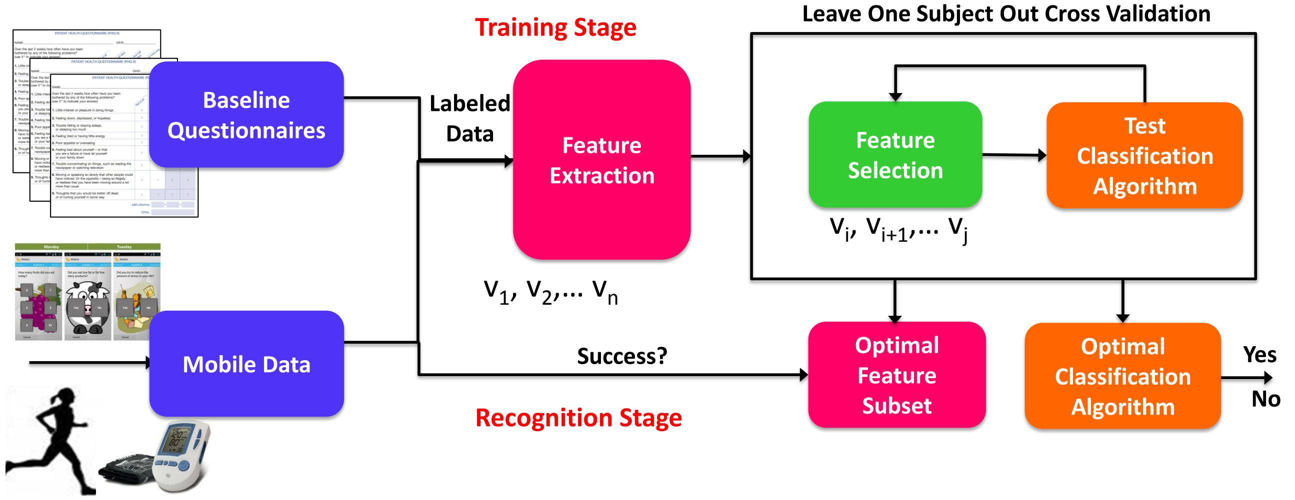

Fig. 2. Wanda-CVD Prediction Methodology

TABLE II
STATISTICAL FEATURES

Type	Statistical Functions	Acronym
Extremes	Minimum	min
	Maximum	max
Average	Mean	mean
	Root Mean Square	RMS
	Median	median
Quartiles	1st Quartile	qrtl1
	3rd Quartile	qrtl3
	Interquartile Range	iqrl
Moments	Standard Deviation	std
	Variance	var
	Skewness	skew
	Kurtosis	kurt
Peaks	Number of Peaks	noPeaks
	Mean Amplitude of Peaks	mampPeaks

the six month study, and 37 participants resulted in usable first month data. Out of the 37 that were studied, 18 were considered to be successful participants, and 19 were considered to be unsuccessful as defined by a medical expert.

Table III shows the rules that were used as the definition of success for each of the six specific outcomes. There were several participants that benefitted in each category, however, to determine the ones that succeeded the most, we calculated the difference in each category between baseline and six months. Next, we averaged the values across each category (giving equal weighting to each of the 6 outcomes). The participants above 50% improvement were considered successful, while those below 50% were considered unsuccessful.

VII. Prediction and Modeling

In order to predict outcome success using data from 37 participants in the study, we test three different feature sets. The first feature set includes the contextual baseline features. The second feature set includes the statistical functions applied to participants' first month data. The third feature set includes both the first and second feature set.

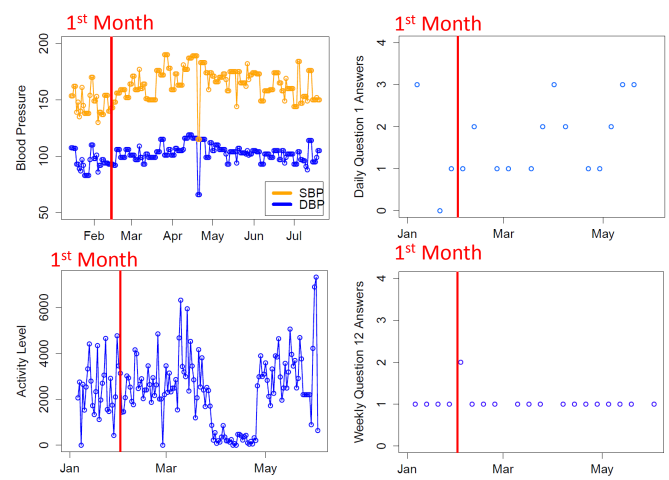

Fig. 3. Sample signals from sensor data. We use each participants first month data from blood pressure (top left), activity (bottom left), daily questionnaire (top right), and weekly questionnaire (bottom right) to see which signals best help predict outcome success.

TABLE III
DEFINING OUTCOME SUCCESS AND FAILURE

Outcome	Success
Body mass index (BMI)	If BMI Loss $> 1\ Ib/inch^2$
Waist Circumference (WC)	If WC Loss $>= 1\ inch$
High density lipoprotein (HDL)	If HDL increases
Low density lipoprotein (LDL)	If LDL decreases
Triglycerides (TG)	If TG decreases
Total Cholesterol (TC)	If TC decreases

We attempt to learn the optimal feature selection and classification combination for each classifier. The two well-known feature selection categories are the filter and wrapper methods. Filter methods use a specific metric to score each individual feature (or a subset of features together), and are typically fast and require less computational complexity.

Wrapper methods usually utilize a classifier to evaluate feature subsets in an iterative manner according to their predictive power [24]. Even in classifiers that perform feature selection, like decision tree classifiers, feature selection has been shown to improve the resulting models stability and accuracy. We applied the wrapper method, testing the following feature selection algorithms: Correlation-based (CFS), Information gain, Gain ratio, Chi-square, Principal Component Analysis (PCA). During our trials we consistently found the CFS algorithm to outperform others in reducing the dimensionality of the data while identifying a small enough subset of informative features for improving prediction performance. CFS is a filter method approach to feature selection which evaluates different subsets of features giving more value to the feature subsets that are highly correlated with the outcome of the classifier (making sure features are relevant), yet uncorrelated to each other (preventing redundancy).

Following the feature selection methodology we tested multiple discriminative classification algorithms: Logistic Regression (LR) [25], C4.5 Decision Trees (C4.5 DT) [26], kNN [27]; and one generative classification model: Naive Bayes [28]. We selected the five classifiers based on their comprehensiveness, complexity and accuracy.

Logistic Regression is a discriminative classification technique that is used in several medical applications that determines the relationship between the features (selected from CFS) and the outcome (success of failure) by calculating a probability using a logistic function [25]. To prevent overfitting of the training samples we applied the regularized least-squares constraint with the ridge factor $\lambda = 1 * 10^8$. We use a quasi-Newton method and do not limit the number of iterations for convergence. We also deploy a C4.5 Decision Tree algorithm which is a supervised classification algorithm that builds a decision tree from the training set based on the information gain of each feature. We used a confidence threshold for pruning set to 0.25, and the minimum number of instances per leaf is set to 2 [26]. While decision trees produce intuitive results and have a good combination of error rate and low computational complexity, it can easily overfit the data and is better designed to handle discrete as opposed to continuous features [26]. We deploy the k-nearest neighbor algorithm which is a non-parametric method used for classification which attempts to classify a participant based on a majority vote of the k nearest neighbors in the feature space of training samples, we tested $k = 1, 3, 5$. As the training samples increase, kNN is known to have good performance, but has larger storage requirements, and is susceptible to noisy data. We also test the Random Forest classifier, which is an ensemble classifier that generates a collection of n decision-tree predictors; we tested $n = 10, 50, 100$. To generate each tree we did not set a maximum depth on each tree and used a random selection of features $log_2 m + 1$, where m is the number of features. Random Forest is a way of implementing a stochastic discrimination algorithm that helps to avoid overfitting of the training samples exhibited in C4.5DT while improving stability and accuracy of the classifier. While discriminative classifiers are known to outperform generative classifiers in practice, we wanted to test the Naive Bayes generative classification algorithm which is a statistical method that is based on Bayes theorem and uses a strong independence assumption between the different features [28].

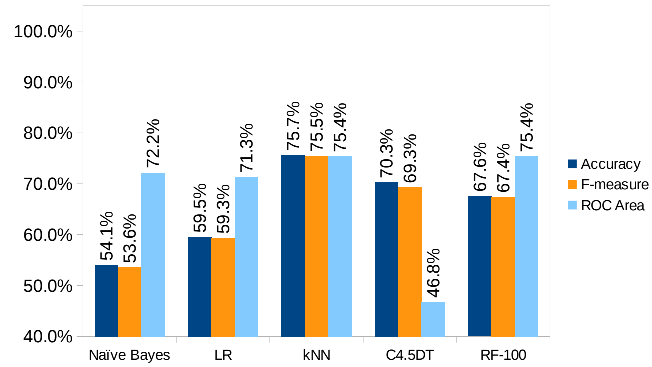

Fig. 4. Results of Classifier 1 which attempts to learn a model that distinguishes participants that were successful from those that were not using baseline contextual features. The kNN classifier outperforms the other four classification schemes.

To ensure confidence in our results, we perform Leave-one-subject-out Cross Validation (LOOCV). We also ensure that the feature selection is performed independent of the test subject to prevent biasing the results. We then calculate the total results from all 37 runs and report the following performance metrics overall accuracy, F-measure (harmonic mean of precision and recall), and the Receiver Operating Characteristic area under the curve (ROC AUC). We also identify highly predictive features and report our findings.

VIII. Results And Discussion

A. Baseline Contextual Features

The first classifier attempts to learn a model that divides the outcomes into participants that are successful and unsuccessful using only baseline contextual features. The optimal results are achieved under the kNN (k=1) classifier as shown in Figure 4, while Naive Bayes and Logistic Regression perform poorly. The accuracy, F-measure, and ROC AUC using the kNN classifier is 75.7%, 75.5%, and 75.4% respectively.

Four features were selected to distinguish between successful and unsuccessful participants. They include participant insurance levels, self-esteem and depressive symptoms. The following were the chosen features for the kNN classifer and consistently chosen across classifiers. They are also ranked according to their predictive power based on the Information Gain feature selection algorithm (with the corresponding pearson correlation coefficient and t-test value, and their corresponding p-value):

1) INSURA: (General Information) Are you currently covered by any of the following health insurances? Government insurance (Medicare, Medicaid, Veteran's Administration health plan, military medical plan, or other government-reimbursed care). Response is either "Yes" or "No." ($\rho = 0.45, t = 3.0, p = 0.005$)
2) SLFEST1: I feel that I'm a person of worth, at least on an equal plane with others. Response range: 1-

”Strongly Agree”, 2-”Agree”, 3-”Disagree”, 4-”Strongly Disagree.” ($\rho = -0.39, t = -2.53, p = 0.016$)

3) INSURB: (General Information) Are you currently covered by any of the following health insurances? Private/commercial insurance. Response is either ”Yes” or ”No.” ($\rho = -0.46, t = -3.04, p = 0.0045$)
4) PHQ7: Trouble concentrating on things, such as reading the newspaper or watching television. Response range: 1-”Not at all”, 2-”Several Days”, 3-”More than half the days”, 4-”Nearly every day.” ($\rho = -0.21, t = -1.25, p = 0.2196$)

It is interesting to note that all the participants that had government insurance succeeded in reducing their risk factors for cardiovascular disease. The majority of the participants that had private/commercial insurance (94.7%, n=18 participants) did not succeed in the study. This is an interesting finding especially since the majority of the participants that had government insurance exhibited financial struggles, showing the potential benefit of the intervention in this group. Also from the Rosenberg Self-Esteem scale we can see that the majority of the participants that succeeded (88.9%, n=16 participants) in the study had high levels of self-worth, strongly agreeing to the SLFEST1 question. Another interesting finding was that all subjects that had significant difficulty reading the newspaper or watching T.V. (value $\geq$ 3 or greater in PHQ7) did not succeed. Based on the t-test we can see that the null hypothesis can be rejected for the majority of variables except PHQ7. However, in classification, while one variable on its own may not be highly correlated with the outcome, the combination of variables may be meaningful.

B. Month One Intervention Features

The second classifier attempts to learn which participants succeeded based on first month intervention data transmitted from the smartphone. The optimal results are achieved under the Random Forest classifier with n=100 trees. However, comparable performance is achieved under the Naive Bayes, kNN, and C4.5 Decision Trees as shown in Figure 5. The worst performance is achieved using Logistic Regression. The accuracy, F-measure, and AUC using the Random Forest classifier is 83.8%, 83.8%, and 78.4% respectively.

Five features were selected to distinguish between successful and unsuccessful participants. The following were the chosen features for the Random Forest classifier and consistently across classifiers, ranked according to their predictive power based on the Information Gain feature selection algorithm (with the corresponding pearson correlation coefficient and t-test, and corresponding p-value):

1) Iqrl WQ-12: Interquartile Range for Weekly Question 12. ($\rho = 0.51, t = 3.53, p = 0.0012$)
2) Var WQ-12: Variance for Weekly Question 12. ($\rho = 0.46, t = 3.03, p = 0.0045$)
3) Var WQ-11: Variance for Weekly Question 11. ($\rho = 0.45, t = 3.02, p = 0.0047$)
4) Kurt DQ-1: Kurtosis for Daily Question 1. ($\rho = 0.41, t = 2.65, p = 0.0121$)
5) Kurt DQ-7: Kurtosis for Daily Question 7. ($\rho = -0.17, t = -1.03, p = 0.31$)

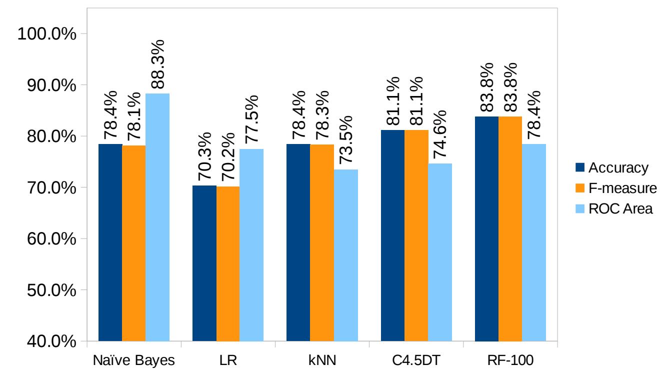

Fig. 5. Results of Classifier 2 which attempts to learn a model that distinguishes participants that were successful from those that were not using statistical features collected on the data during month one of the intervention. The Random Forest classifier outperforms the other four classification schemes.

The primary features involved statistical functions related to the following four (two weekly and two daily) questionnaires:

1) WQ-12: Ate beans, seeds or nuts 4 to 5 times per week. Response range: 0-”None of the time”, 1-”Some of the time”, 2-”Most of the time”, 3-”All of the time.”
2) WQ-11: Ate low fat or fat free dairy products (instead of whole milk, regular ice cream, butter, regular cheese or cottage cheese, yogurt or sour cream made with whole milk). Response range: 0-”None of the time”, 1-”Some of the time”, 2-”Most of the time”, 3-”All of the time.”
3) DQ-1: How much additional activity did you do today? Response range: 0-”0 minutes”, 1-”10 minutes”, 2-”20 minutes”, 3-”30 minutes or more.”
4) DQ-7: Did you follow a diet low in saturated and trans fats? Response is either ”Yes” or ”No.”

C. Combined Feature Set

The third classifier attempted to learn which participants succeeded based on both the baseline contextual features and the continuous data transmitted from the smartphone during the first month of the intervention. It is interesting to note that Logistic Regression typically outperforms other techniques. As shown in Figure 6, classifiers 1 and 2 perhaps had complimentary information necessary to strengthen the predictive power of the classifier. Not surprisingly the combined feature set included 15 features which were a combination of the baseline contextual features and the month one intervention features. The following is a list of the selected features: Iqrl WQ-12, Var WQ-12, Var WQ-11, Std WQ-11, Std WQ-12, Skew WQ-12, Kurt DQ-1, Kurt DQ-1, INSURA, INSURB, PHQ7, SLFEST1, and three family history related questions. Using a combined feature set improves overall results, but more importantly produces more stable results across different classifiers, signifying strength in the features selected for classification. The accuracy, F-measure, and AUC using the Logistic Regression classifier is 91.9%, 91.9%, and 95.9% respectively. It is interesting to note that a k-NN classifier outperforms or is comparable to a more complex Random Forest classifier when considering the contextual-based fea-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2016.2518673, IEEE Journal of Biomedical and Health Informatics

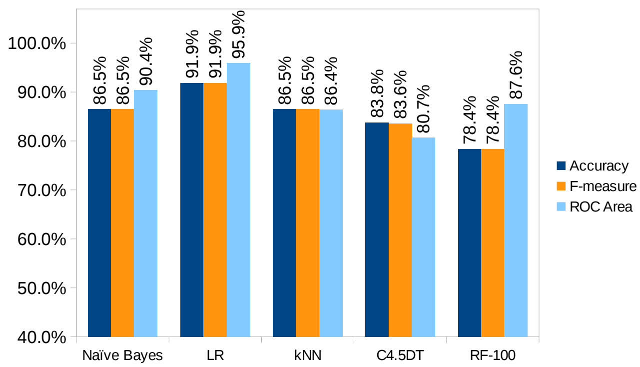

Fig. 6. Results of Classifier 3 which attempts to learn a model that distinguishes participants that were successful from those that were not, using both baseline contextual features and statistical features collected on the data during month one of the intervention. The kNN classifier outperforms the other four classification schemes.

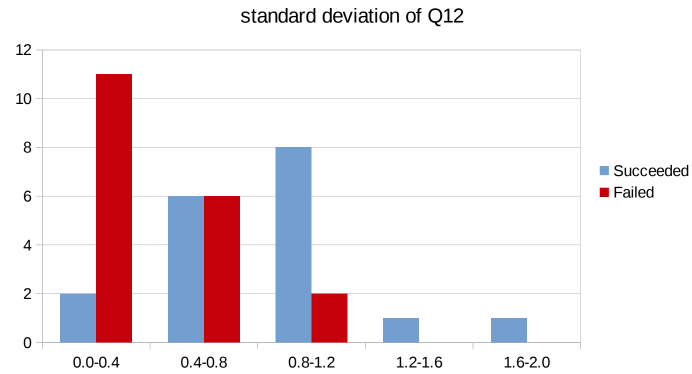

Fig. 7. Standard Deviation of Weekly Question 12: We can see that this feature is highly discriminative between successful and unsuccessful participants.

tures as shown in Figures 4 and 6. This finding suggests that there are perhaps more similarities across patients in terms of their baseline contextual features than in their weekly measureable habits. Alternatively, our sample distribution of weekly measurable habits is not wide enough to capture all possible month one intervention habits. The Naive Bayes generative model outperforms other discriminative models only in Figures 5 and 6, which shows us that given the weekly measurable habits perhaps a strong independence assumption between features can be safely assumed, or that it reaches its higher asymptotic error faster [29].

D. Best Feature

One of the features that was most highly correlated with predicting outcome success was the standard deviation and variance of WQ-12, which shows the participants struggling to improve their nut, bean and seed intake. Based on Figure 7, we can see that the majority of the successful participants exhibited higher variation in their response to consuming nuts, beans and seeds during the first month owing to their commitment in the struggle to improve their diet. We also see that the participants that succeeded were more likely to report greater consumption of the nut, bean and seed diet, as shown in Figure 8. Using the single standard deviation feature of Weekly Question 12, we obtain the following accuracy, F-measure and ROC Area: 78.4%, 78.3%, and 71.9% respectively. Such findings can help us better understand early in an intervention which participants may need further support. One reason why participants that succeeded exhibited higher variability in their response to the consumption of nuts, beans and seeds could be attributed to the honesty of the participants, which could be an indication of their dedication to following the study regimen. A second reason can be that these participants are reporting a necessary struggle that needs to take place to improve health and diet of young African-American women.

It is also interesting to note that in Figure 5 and Figure 6 the C4.5 decision tree algorithm does not significantly trail the results of the Random Forest classifier, which suggests that the underlying model for predicting and understanding outcome success may not be so complex, especially given a feature that is highly correlated with its outcome.

E. Discussion

Employing heart healthy habits such as a change in diet was a predictor of those who succeeded. Consuming nuts, beans and seeds was encouraged in the lifestyle change education provided to the women early on in the study. These nutrients are not normally found in this population, however it is most noteworthy that those women who incorporated this change in their diet are most successful in the program. Their willingness to purchase and consume more expensive snacks such as nuts and seeds rather than choose to ingest cheaper and more available snacks such as chips and baked goods may mean that they have taken the education seriously and have committed the tenets of the program into lifestyle changes for themselves and their families.

The intervention group was recruited from one church in an urban population in the LA area. This limits the generalizability of the findings. Even though education regarding healthy nutrition was imparted many factors limited the adoption of healthy changes in behavior due to the limited availability of fresh fruits and vegetables in the immediate area and also cost constraints. Many of the women were single parents and several had no means of transportation.

IX. Conclusions

Instituting RHM systems as a means of monitoring and predicting patient's outcome success to treatment regimens, coaching individuals to change behavior to adopt healthier lifestyles, and identifying dangerous habits is the wave of the future. Deploying RHM along with clinician resources can be costly. With the emphasis on cost consciousness and cost efficiency in mind, we report a pilot study presenting a useful framework for analyzing individual's baseline features or questionnaire responses and first month intervention data to predict their level of outcome success in reducing their risk factors for CVD. Along with providing a means for predicting outcome success, we also identify unique features collected during the first month intervention that are highly correlated with outcome success, such as participant variability

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2016.2518673, IEEE Journal of Biomedical and Health Informatics

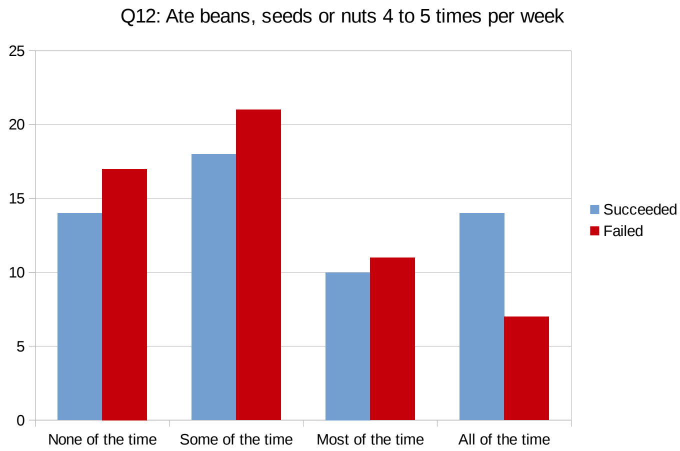

Fig. 8. Response to the following Weekly Question 12: Ate beans, seeds or nuts 4 to 5 times per week. We can see that the successful participants were skewed more towards eating nuts, beans and seeds.

in their response to their consumption of nuts, beans and seeds. This finding could be attributed to a serious struggle that is necessary for African American women to improve their health and diet.

Our enhanced Wanda-CVD RHM system was tested in the WHHS in a group of Black women ages 25-45. We analyze key contextual features that predict outcome success with an F-measure of 75.4%. We improve the prediction accuracy by incorporating features from the first month of the intervention, resulting in an F-measure of 83.8%. We also show that further improvement and stability can be obtained by combining both baseline contextual features and first month data, which achieves an F-measure of 91.9%. The results show promise in the ability of a classifier to predict RHM outcome success using a subset of features collected early on in the intervention. Limitations from this study are many. The study needs to be replicated in a larger, more diverse group of black women. Our group was in an urban setting and cannot be generalized to the entire population. Further study in larger and more diverse groups is warranted using similar analysis to further validate our results while identifying groups who could benefit from RHM systems.

X. Acknowledgements

This grant is partially supported by AHA: Clinical Research Grant $\#12CRP11910021$.

References

[1] J. Sarasohn-Khan. (2011) California healthcare foundation. the connected patient: Charting the vital signs of remote health monitoring. CIM.pdf. [Online]. Available: http://www.chcf.org/publications/2011/02/the-connected-patient

[2] A. L. Bui and G. C. Fonarow, "Home monitoring for heart failure management," *J. Am. Coll. Cardiol.*, vol. 59, no. 2, pp. 97–104, Jan 2012.

[3] M. A. Konstam, "Home monitoring should be the central element in an effective program of heart failure disease management," *Circulation*, vol. 125, no. 6, pp. 820–827, Feb 2012.

[4] A. S. Desai, "Home monitoring heart failure care does not improve patient outcomes," *Circulation*, vol. 125, no. 6, pp. 828–836, 2012.

[5] C. D. Gillespie, C. Wigington, and Y. Hong, "Coronary heart disease and stroke deaths - United States, 2009," *MMWR Surveill Summ*, vol. 62 Suppl 3, pp. 157–160, Nov 2013.

[6] N. Alshurafa, J. Eastwood, M. Pourhomayoun, J. J. Liu, S. Nyamathi, and M. Sarrafzadeh, "A framework for predicting adherence in remote health monitoring systems," in *Proceedings of the Wireless Health 2014, National Institutes of Health, Bethesda, MD, USA, October 29-31, 2014*, 2014, pp. 1–8. [Online]. Available: http://doi.acm.org/10.1145/2668883.2669586

[7] J. McCambridge, E. Kalaitzaki, I. R. White, Z. Khadjesari, E. Murray, S. Linke, S. G. Thompson, C. Godfrey, and P. Wallace, "Impact of length or relevance of questionnaires on attrition in online trials: randomized controlled trial," *J. Med. Internet Res.*, vol. 13, no. 4, p. e96, 2011.

[8] M. J. Wade, A. S. Desai, C. M. Spettell, A. D. Snyder, V. McGowan-Stackewicz, P. J. Kummer, M. C. Maccoy, and R. S. Krakauer, "Telemonitoring with case management for seniors with heart failure," *Am J Manag Care*, vol. 17, no. 3, pp. e71–79, Mar 2011.

[9] Remote physiological monitoring: Research update. [Online]. Available: http://www.nehi.net/publications/

[10] S. I. Chaudhry, Y. Wang, J. Concato, T. M. Gill, and H. M. Krumholz, "Patterns of weight change preceding hospitalization for heart failure," *Circulation*, vol. 116, no. 14, pp. 1549–1554, Oct 2007.

[11] F. Koehler, S. Winkler, M. Schieber, U. Sechtem, K. Stangl, M. Bohm, H. Boll, G. Baumann, M. Honold, K. Koehler, G. Gelbrich, B. A. Kirwan, and S. D. Anker, "Impact of remote telemedical management on mortality and hospitalizations in ambulatory patients with chronic heart failure: the telemedical interventional monitoring in heart failure study," *Circulation*, vol. 123, no. 17, pp. 1873–1880, May 2011.

[12] S. Mattke. (2010) Rand corp. health and well-being in the home. a global analysis of needs, expectations, and priorities for home health care technology. RANDOP323.pdf. [Online]. Available: http://www.rand.org/pubs/occasionalpapers/OP323.html

[13] M. Terry, "Medical Apps for Smartphones," *Telemed J E Health*, vol. 16, no. 1, pp. 17–22, 2010.

[14] J. J. Oresko, H. Duschl, and A. C. Cheng, "A wearable smartphone-based platform for real-time cardiovascular disease detection via electrocardiogram processing," *IEEE Trans Inf Technol Biomed*, vol. 14, no. 3, pp. 734–740, May 2010.

[15] S. Krishna, S. A. Boren, and E. A. Balas, "Healthcare via cell phones: a systematic review," *Telemed J E Health*, vol. 15, no. 3, pp. 231–240, Apr 2009.

[16] N. Alshurafa, J.-A. Eastwood, M. Pourhomayoun, J. J. Liu, R. Li, and M. Sarrafzadeh, "Remote health monitoring: Predicting outcome success based on contextual features for cardiovascular disease," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, July 2014.

[17] P. M. Ho, C. L. Bryson, and J. S. Rumsfeld, "Medication adherence: its importance in cardiovascular outcomes," *Circulation*, vol. 119, no. 23, pp. 3028–3035, Jun 2009.

[18] M. Lan, L. Samy, N. Alshurafa, M.-K. Suh, H. Ghasemzadeh, A. Macabasco-O'Connell, and M. Sarrafzadeh, "Wanda: An end-to-end remote health monitoring and analytics system for heart failure patients," in *Proceedings of the Conference on Wireless Health*, ser. WH '12. New York, NY, USA: ACM, 2012, pp. 9:1–9:8.

[19] N. Alshurafa, J.-A. Eastwood, M. Pourhomayoun, S. Nyamathi, L. Bao, B. Mortazavi, and M. Sarrafzadeh, "Anti-cheating: Detecting self-inflicted and impersonator cheaters for remote health monitoring systems with wearable sensors," in *BSN'14*, 2014, pp. 1–6.

[20] N. Alshurafa, J.-A. Eastwood, S. Nyamathi, W. Xu, J. J. Liu, M. Pourhomayoun, H. Ghasemzadeh, and M. Sarrafzadeh, "Battery optimization in smartphones for remote health monitoring systems to enhance user adherence," in *Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments*, ser. PETRA '14, 2014.

[21] N. Alshurafa, J. A. Eastwood, S. Nyamathi, J. Liu, W. Xu, H. Ghasemzadeh, M. Pourhomayoun, and M. Sarrafzadeh, "Improving Compliance in a Remote Health Monitoring System through Smartphone Battery Optimization," *IEEE J Biomed Health Inform*, Jun 2014.

[22] J. Ware, M. Kosinski, and S. D. Keller, "A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity," *Med Care*, vol. 34, no. 3, pp. 220–233, Mar 1996.

[23] R. G. Downey and C. King, "Missing data in Likert ratings: A comparison of replacement methods," *J Gen Psychol*, vol. 125, no. 2, pp. 175–191, Apr 1998.

[24] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Mar. 2003. [Online]. Available: http://dl.acm.org/citation.cfm?id=944919.944968

[25] A. Agresti, *Categorical Data Analysis, 2nd edition*. Hoboken, NJ: John Wiley & Sons, 1993.

[26] J. Quinlan, *C4.5: Programs for Machine Learning*. San Francisco, CA: Morgan Kaufmann Publishers, Inc., 1993.

[27] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.

[28] J. Pearl, "Bayesian networks," 1997.

[29] A. Y. Ng and M. I. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes," in *Advances in Neural Information Processing Systems 14*, T. Dietterich, S. Becker, and Z. Ghahramani, Eds. MIT Press, 2002, pp. 841–848. [Online]. Available: http://papers.nips.cc/paper/2020-on-discriminative-vs-generative-classifiers-a-comparison-of-logistic-regression-and-naive-bayes.pdf