Carnegie Mellon University
Disruptive Health Technology Institute
Hamburg Hall | Carnegie Mellon University | 5000 Forbes Avenue | Pittsburgh, PA 15213 | 412-268-6920

ANNUAL PROGRESS REPORT

Project Title: Combining External and Implantable Sensors with Machine Learning to Detect changes in the Health Status in patients with Systolic Heart Failure

PI: Professor Asim Smailagic

Grant Award: SPEX 151506

Project Period: 10/1/2015 – 6/30/2017

I. Introduction

We hypothesize that external (wearable) sensors in combination with implantable sensors can provide better and sooner indication of changes in both physiologic and functional parameters in patients with systolic heart failure which will allow the medical team to make clinical decisions earlier and reduce the utilization of care by the patient. Towards this end, we are going to study the cohort of Congestive Heart Failure patients that have certain inclusion criteria and being able to complete a quality of life survey. The outcome of our study will provide better and sooner indication of changes in both physiologic and functional parameters in patients with systolic heart failure which will allow the medical team to make clinical decisions earlier and reduce the utilization of care by the patient. We aim to developa system of instrumentation and data analysis software that alerts clinicians to patient decline well before she reaches the point that hospitalization is called for. The underlying premise is that by applying modern analytics to those new and potentially voluminous data streams, clinicians can make good use of the data with only a very small increase in workload.

II. Daily Telehealth Readings to Predict CHF Exacerbations

A. Motivation

More than a quarter of a million Americans die from complications resulting from Congestive Heart Failure (CHF) every year[1]. Previous research has shown that clinical intervention during initial presentation of symptoms can significantly reduce the risks of a cardiac event [1]. Early intervention requires careful monitoring of the patient's vital signs using in-home telehealth monitoring. However, by the time symptoms manifest, it may already be too late for effective intervention. In this work, we present a model to predict sudden weight changes in CHF patients, which is an indication that the subject has begun retaining fluid. Integrated Autoregression is a popular class of algorithms for predicting future time series values, however with Congestive Heart Failure patients, we may see sudden changes in the time series distribution that traditional autoregressive models cannot anticipate. To account for this, we introduce Markovian latent variables to monitor the patient's global health trends. This allows us to quickly introduce new autoregression parameters to account for rapidly changing conditions in the patient's health.

[1] http://www.cdc.gov/dhdsp/data_statistics/fact_sheets/fs_heart_failure.htm

Using a latent variable autoregression model, we are able to predict sudden weight changes between one and two days before they occur, with a sensitivity of .8313. This framework could allow for clinical interventions to take place days before current systems, potentially saving the lives of many individuals living with Congestive Heart Failure.

B. Related Work

There has been significant clinical research aimed at targeted intervention of Congestive Heart Failure patients to reduce the risks of readmission. One study suggested that continual monitoring of a patient in home can reduce the risk of hospital readmission by as much as 86% [2]. However, this same study concluded that telehealth systems, while extremely effective, did not lead to better outcomes than daily phone calls between a nurse and a patient. This evidence could suggest that by the time medical complications become apparent in telehealth data, the subject is already sufficiently uncomfortable to seek intervention when it is available.

Data driven methods offer us the possibility to detect more subtle, multimodal changes in a patient's physiology that may not be immediately apparent to a clinician. It has become more common for CHF patients to be evaluated by a machine learning model upon hospital discharge to determine the risk of readmission. These methods can achieve above 80% predictive accuracy, exceeding that of a physician in some instances [3, 4]. In particular, these models are very useful for detecting the highest risk patients, that can then be monitored more frequently in order to prevent readmission. Machine learning methods have also been designed to diagnose CHF [5].

Very little research has been done to extend these approaches to use machine learning with daily telehealth datastreams. This is particularly important because some patients may be discharged from the hospital as low risk, only to have their health deteriorate without clinical monitoring. Many studies continue to suggest that telehealth monitoring leads to lower rate of death and re-hospitalization for CHF patients [6], however, hiring clinicians to monitor incoming data streams is extremely cost-prohibitive. This has led some medical centers to develop rule-based systems for monitoring telehealth data. We believe this presents an opportunity for more sophisticated, data-driven, automated monitoring tools.

Latent variables have been used in conjunction with autoregression models in the past to allow for more rapid and dynamic responses to changing conditions. However, these efforts have mostly been in the field of economics [7, 8], with much less effort being given to medical applications.

Fig. 1. A Telehealth Station for at-home Monitoring

C. Dataset

The data for this project was collected by Meadville Medical Center[2] using Authentidate telehealth systems, shown in figure 1. We used other datasets as well. The data includes 9 volunteers living with Congestive Heart Failure over an average period of 6 months. Each subject provided daily or twice daily measurements that include systolic blood pressure, diastolic blood pressure, heart-rate, and current weight. Figure 2 shows an example of data taken from a 10 month period. We can clearly see that weight is the most stable of the time series, however, we see a sudden increase in weight near the beginning of September that was the result of sudden fluid resulting from poor kidney health as a result of CHF.

Each subject also includes a set of rules, assigned by a clinician, that will alert the medical center when certain conditions are met, for instance if a patient gains more than 2 pounds of weight in a single day. This particular rule produced notifications in 11 instances, and we use these instances to evaluate our methods. Our goal is to predict these sudden weight gains using the autoregressive model with as much advance notice as possible.

[2]http://meadvillemedicalcenter.com/

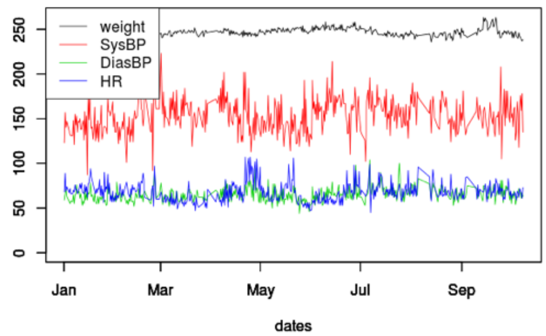

Fig. 2. Example Time Series from Dataset

D. Methods

When monitoring large populations of patients, the cost of clinical supervision becomes a prohibitive factor. Telehealth systems can provide useful means for monitoring high-risk patients, but we cannot depend on annotations supplied by clinical experts. Instead, we must infer what we can from the signals themselves. Autoregression is a very popular school of methods that allows us to model the dynamics of a time series and predict future values, without requiring supervised training. In this class of algorithms, future values of a time series are estimated using a function that takes the most recent time series values as inputs. Multivariate autoregression, which operates over multiple time series simultaneously, has been successfully used in medical applications in the past [9], and these methods are also very popular for financial applications [10]. Unfortunately, many autoregression models treat the time series as stable, meaning that the statistical moments of the distribution do not change over time. With congestive heart failure data, a patient with declining health will produce a time series with changing moments in the distribution, so we do not meet the stability requirement of many approaches. Autoregressive Integrated Moving Average (ARIMA) models are built to address just such a problem, allowing the first moment of the time series distribution to change over time. These models have been used successfully in applications such as electricity cost prediction, in which the average of the time series can change significantly with the market [11].

An ARIMA model is denoted ARIMA(p, d, q), with p representing the degree of the autoregression, d is the degree of difference required for stationarity, and q represents the degree of the moving average model. This leads to a forecasting model as follows:

$$\hat{y}_t = \mu + \sum_{i=1}^{p}(\psi_i y_{t-i}) - \sum_{j=1}^{q}(\theta_i e_{t-j})$$

In this instance, e_j represent the moving average differences of the time series, which can be thought of as a discrete variant of a d^{th} degree derivative. When $d = 2$, these values can be thought of as the local acceleration of the time series.

This type of moving average model works well for non-stationary time series, however it often requires that the moments of the time series distribution change slowly over time. For congestive heart failure patients, a sudden medical emergency can lead to very sudden changes in the time series distribution. As such, we model these sudden changes using Markovian hidden states, similar to a Hidden Markov Model. If our model has hidden states $h_i \in h_1...h_N$, each hidden state will have its own ARIMA parameters θ and ψ (though p, d, and q remain the same). We then have an initial state distribution π, and $N \times N$ transition matrix T denoting the probability of moving between states, and an observation matrix O denoting the probability of an observed value given a hidden state. By modeling the latent states of the patient's health, we are able to make rapid transitions between ARIMA models when it appears that the subject's condition is beginning to change.

We use a small parameter validation set from the patient data, and train the latent variable model and a multivariate ARIMA model on the weight time series, using diastolic blood pressure as a side input. This validation set is used to select the hyper-parameters of the model, results in $N = 3$ hidden states, with each state representing a $ARIMA(5, 2, 3)$ model. The hidden state parameters are trained using gradient descent, while the ARIMA parameters are estimated using the Box-Jenkins model.

Once the model is trained, we can use it to instantiate alerts for clinicians about the possible decline of patient health. In particular, we evaluate the average slope of the predicted time series curve over the next two days (denoted ψ). We can then issue an alert any time this value surpasses a given threshold:

$$\psi \geq \delta$$

We can then control the sensitivity and specificity

Model	Residual Mean
BL (Current Value)	2.1
BL (Mean)	1.61
ARIMA	1.08
Latent Variable ARIMA	0.92

TABLE I
MEAN RESIDUAL ERROR COMPARISON

of the model according to a medical center's available resources. In particular, the cost of a false negative (a patient with declining health that is not detected by the system) is much higher than the cost of a false positive. As such, we would select the lowest value of δ possible, such that a medical center has the resources to personally monitor and intervene with the number of patient flagged by the system for that threshold value. In this way, we can achieve the highest sensitivity possible, at the cost of lower specificity.

E. Empirical Results

Figure 3 shows the mean residual error of the latent variable ARIMA model as a function of the percentage of training data used. The minimum average error when predicting weight two days in advance was 0.92 pounds.

Additionally, table II-E shows the mean residual error of a traditional ARIMA model, without latent variables, as well as two simple baselines. In the first baseline, the current weight of the patient is simply used as the predicted value, while the second baseline uses the average weight of the patient across the entire time series. A 95% confidence bound for this values is ± 0.15 pounds, Meaning that the improvement seen over the baseline by both ARIMA models is statistically significant.

While the mean residual error of the latent variable ARIMA model shows only a modest improvement over the traditional model, we see significantly better prediction for the latent variable model when dealing with sudden changes in weight. Figure 4 shows an example of a patient's sudden weight gain, likely as a result of fluid retention. This figure also shows that the latent variable ARIMA model correctly predicts the weight gain two days in advance. What is particularly interesting is that the latent state value changed from h_1 to h_2 on day 9, likely due to the erratic fluctuation of the weight in the days prior.

As with most auto-regression applications, prediction becomes much more difficult as the time horizon increases. Figure 5 shows the effects of conducting prediction one month in advance. We see that the model begins to settle into predicting the moving mean value of the time series, missing the larger deviations in the actual data. Evidence suggests that 2 or 3 days is the ideal horizon in which clinical decisions should be made as a result of the predicted values.

Of the 11 instances in the dataset in which a patient gained more than 2 pounds in one day, using a threshold value of $\delta = 1.0$ allows the latent variable ARIMA model to correctly predict the weight gain in 9 instances, resulting in a sensitivity value of 0.8313, along with 8 false positive examples.

F. Discussion

As we have seen in this work, we can quite accurately model the dynamics of a weight time series for patients with congestive heart failure, and we see some promising results for predicting sudden exacerbations of health. While the number of occurrences of fluid retention in this dataset is too low to make confident claims about the efficacy exacerbation detection, these results, combined with the low mean residual error, suggests that we may be able to detect early warning signs of health decline in a very high percentage of occurrences, potentially giving us the means to cut costs and save lives. It is interesting to note that for the 2 instances of fluid retention not detected by the system, using lower values of δ do not produce improved results. This suggests that daily weight and blood pressure readings alone may be insufficient to detect all causes of cardiac failure. By using more sophisticated monitoring and analysis techniques, we have the opportunity to reduce the quarter of a million annual deaths that result from chronic congestive heart failure.

III. MONITORING PATIENT HEALTH WITH WEARABLE DEVICES

A. Overview

CHF patients may not show symptoms of heart failure at rest, but they may show them on exertion. Peak oxygen consumption (peak VO2) of patients are an index of the functional capacity of the heart. Traditional tests to assess patients require directly measuring the peak VO 2 levels while exercising, which requires expensive gas analyzers. Instead, patients can perform simple exercises to achieve peak VO and their performance can be correlated with how their disease is progressing. One such exercise test 2 is the Six Minute Walk Test (6MWT). We intend to use externally wearable sensors such as the Apple Watch to measure a patients performance in 6MWT. The data collected will be used to train predictive models using Machine Learning techniques.

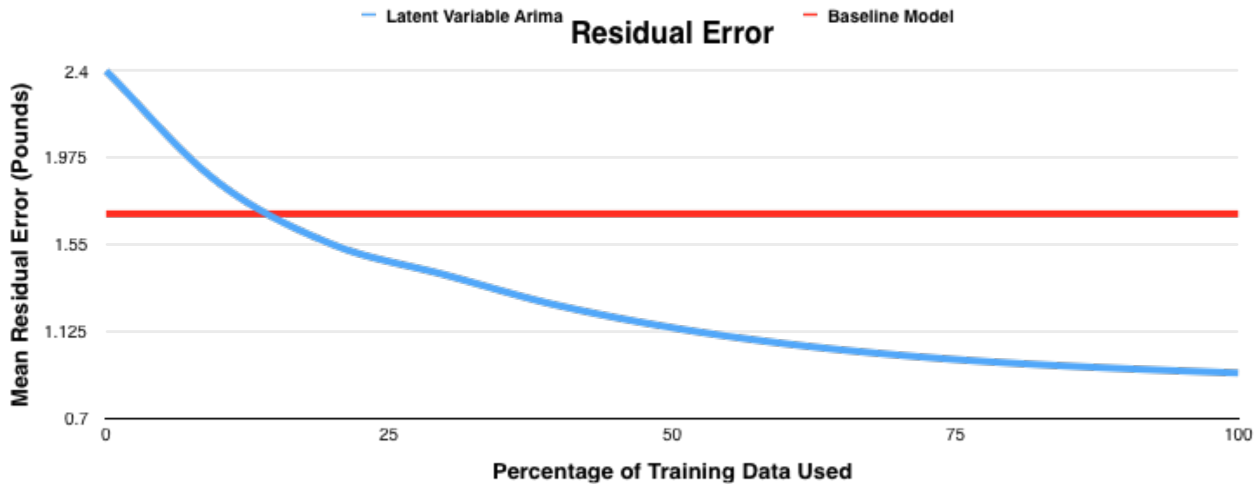

Fig. 3. Mean Residual Error of Latent Variable ARIMA

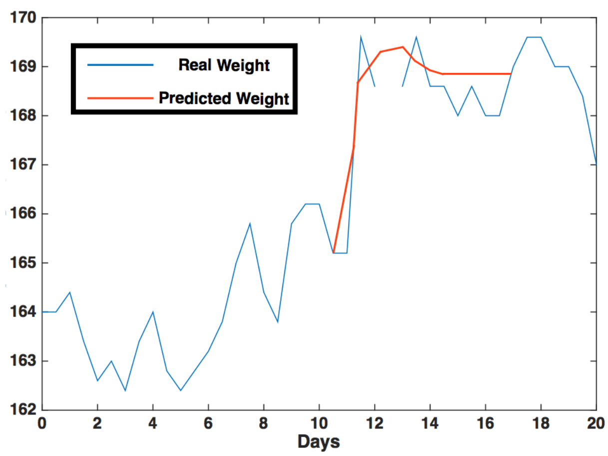

Fig. 4. Example of Predicted Subject Weight Gain

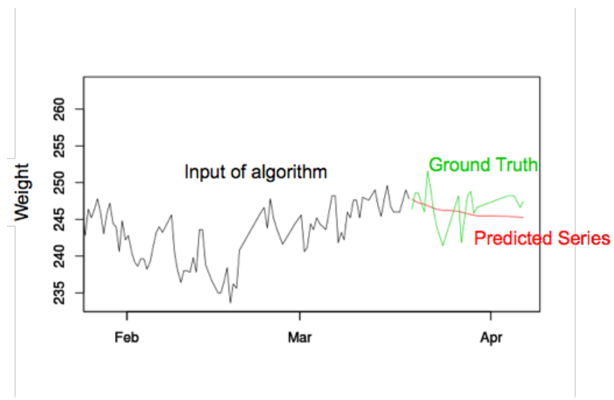

Fig. 5. Long Term Weight Prediction

Remotely monitoring progress of patients using wearable devices and the Six Minute Walk Test will reduce time and costs of monitoring CHF patient recovery.

The peak VO 2 of a patient provide useful information for prognosis of heart failures. The most accurate measurement for peak VO 2 can be made using gas analyzer masks which are able to measure breath by breath levels of O and CO. Exercise testing is a common technique for monitoring the peak VO levels of a patient. The peak exercise capacity is the maximum ability of the cardiovascular system to deliver oxygen to the human body. Often patients are required to simple exercises such as brisk walking, climbing stairs, etc. to their maximum capacity. Their vital signs such as heart-rate and blood pressure are measured before, during and/or after the test to make inferences about their current health and predicting failures. Surveys may be administered to monitor other parameters such as:
1. Number of times patients stopped in the duration of the exercise
2. Mood of the patient before and after the exercise test
3. Whether encouragement was needed to complete the exercise
4. Any unusual observations that may introduce variability
Body weight, body height, gender, age, etc. are also

combined with all the above factors to infer about the patients health and a doctor may intervene if necessary.

B. Six Minute Walk Test (6MWT)

The 6MWT is the most popular exercise test used with cardiac patients. It is the most practical and simple test to be administered to a patient. Walking is an activity performed by all patients except those who are severely impaired. In this test, patients are required to quickly walk on a flat, hard surface for a period of six minutes. Important factors such as the distance covered in six minutes, the heart-rate before and after the test, heart-rate recovery period and number of stops made are considered to conclude about the results of the test. Most patients do not achieve maximal exercise capacity during a 6MWT. However, since most regular activities are performed at submaximal levels of capacity, 6MWT is still a better indicator of capacity than regular activities.The 6MWT being administered in a long hallway. The patient is asked to walk quickly while a nurse times the test and records observations

C. Shuttle Walk Test

The purpose of the shuttle walk test is to measure how far and how fast the patient can walk. The test requires two cones to be placed on the ground about ten meters apart. Patients are instructed to walk back and forth between the cones. They start at a slow pace and then incrementally increase their speed in one minute intervals. They must keep walking until they are completely exhausted and cannot go on further. The important factors recorded are the number of laps (distance covered), heart-rate, blood pressure, number of stops, etc. similar to the 6MWT.

Since the patients are required to continue until their maximal exercise capacity is exhausted, the shuttle walk test is a more accurate indicator of peak VO than the 6MWT. However, if the patients exert themselves beyond capacity, it raises concerns about the tests safety. Also, the shuttle walk test requires cones or beacons to be placed on the ground, so the test can be administered only in places with such an arrangement. Therefore, the 6MWT is preferred over shuttle walk test in most cases.

D. Other Tests

There are several other tests found in literature such as 200 meter fast walk, stair climbing test, etc. However, the goal of these tests remains the same: to achieve the patients maximal exercise capacity or peak VO intake. The parameters recorded may differ according to each test. These tests are not as popular 2 as the 6MWT because they are harder to administer and the results are more variable across patients.

E. Selecting a Test

We selected the 6MWT tests as our primary tool of analysis in this project. Our goal is to use externally wearable and implantable sensors to track the patients performance during these tests. Due its simplicity, the 6MWT will be easiest to monitor through such sensor devices. Since the tests might be performed independently by patients, the 6MWT protocol is the easiest to follow and understand. Also, other tests might require larger open spaces, presence of beacons on the ground, etc. for the exercises to be performed.

F. Wearable Heart Rate Sensors

There are many fitness trackers and wearable sensors available today that can be used to improve healthcare systems and quality of life. We are particularly interested in selecting a wearable heart-rate sensor which can provide relevant data of the 6MWT with reasonable accuracy. This section outlines the selection process of wearable heart-rate sensors for the project.

G. Wrist Bands and Smart Watches 3.1.1. Apple Watch

The Apple Watch was the wearable device chosen for this project based on the sensors available, accuracy, storage, and battery life compared to other devices. We surveyed 11 wearable devices based on available onboard sensors, storage capacity, and accuracy. The Apple Watch emerged to be the best considering all of these factors as compared to other wearable devices. The Watch provides accelerometers and gyroscope for activity analysis, a heart-rate sensor, GPS data (from iPhone), 8 GB of storage and 18 hours of battery life. The Watch has a builtin workout app which can be used to perform 6MWT. Data such as heart-rate, steps count, etc. can be obtained through the iPhones Health App. All the data from the app can be exported in a single XML file.

One study compared the accuracy of the step counts and heart-rate sensors of multiple wearables we considered [1]. In this study, the Apple Watch showed accuracy of 99.1% (SD = 16.6) for 200 step counts, rising to 99.5% (SD = 25.8) for 1,000 step counts. It showed the most precise results for 1,000 steps (CV = 2.6). It also had the highest accuracy (99.9%) and precision (5.9%) for heart-rate compared to a clinical grade heart-rate monitor. Based on these results, the Apple Watch was determined to be the best wearable for this project.

Fig. 6. Zephyr Chest Band

H. Chest Band Heart Rate Sensors

Chest band heart-rate sensors are generally more accurate than wrist-based sensors. Sensors worn around the chest measure heart-rate with electrodes placed close to the heart. However, chest band sensors are more intrusive and wearing them throughout the day is impractical.

In order to determine the relative accuracy of the Apple Watch sensor as compared to other wearable sensors, we decided to conduct experiments with a chest band heart-rate sensor. The data collected from the chest band will also be useful in estimating values missed by the Apple Watch.

We selected the Zephyr Chest Band among the four sensors we surveyed. The Zephyr HxM smart is a Bluetooth Low Energy (BLE) based device that can pair with iPhone or Android devices. The device turns on when it is worn and automatically sync with the Health app on the iPhone. The data collected from the Zephyr is also obtained in an XML file along with other health data. This allowed us to follow our originally proposed data collection protocol with the exception of adding the Zephyr chest band.

I. Data Collection Protocol

11 subjects were given Apple Watches and asked to wear them daily for an extended period. Directions for setting up the Watch and creating a daily reminder on their iPhone to perform the Six Minute Walk were provided. The step by step protocol for the Six Minute Walk is listed below.

Six Minute Walk Test

1. Open Health app on iPhone. If no data is displayed on the dashboard, restart the iPhone. This is to avoid a known connection issue between the iPhone and Apple Watch.

Fig. 7. Apple Watch screens for selecting workout mode and setting goals. Steps 3, 4, 5 and 6.

2. Manually record heart-rate on the Watch for 20-30 seconds.
3. Start an Outdoor Walk workout in the Workout app on the Watch.
4. Set the goal of the workout to six minutes
5. Bring the iPhone on the walk for the most accurate distance data from the iPhones GPS.
6. Start the workout on the Watch and begin walking.
7. Stop walking upon completion of the workout.
8. End the workout and save data.
9. Manually record heart-rate on the Watch at 1, 2 and 3 minute intervals after the test.

Subjects were asked to export data from the Health app on days 1, 2, and 3 and email to the email address given. Once it was confirmed that the export process was working as expected, subjects were asked to export data from the Health app weekly and email to the email address given.

J. Issues in Data Collection with Apple Watch

We faced minor issues while collecting data for the purposes of this project using the Apple Watch. Before our custom iOS app had been developed and while we were using Apples existing Health app we experienced two issues which made the exportation of apple watch data inconvenient.

1. The Health app has a bug which causes it to not sync with Apple Watch at times.
2. The Health app exports all data from its beginning of time which can cause the data exports to become too large in size. If the size of the export is beyond the limit of the mail server, the export fails.

To remedy these issues we ensured our iOS application automates the process of data collection and labels

	No. of days for which the Apple Watch was used	No. of workouts performed
Subject #1	12	1
Subject #2	27	5
Subject #3	29	0
Subject #4	4	4
Subject #5	8	7
Subject #6	8	6
Subject #7	4	2
Subject #8	1	1
Subject #9	1	1
Subject #10	1	1
Subject #11	1	1
Total	**96**	**29**

Fig. 8. Number of days for which Apple Watch was used and Number of workouts for each subject

collected data such that 6MWT data can be separated from the rest of the data.

There were two issues which interfered with the time series nature of the apple watch data.

1. The Apple Watchs heart-rate sensor has different sampling frequencies when in Workout mode as compared to its Default mode. In Default mode, the watch samples heart-rate every 10 minutes while in Workout mode, heart-rate is sample every 5 seconds.

2. The Watch rejects some values for reasons such as jerky movements of the hand, poor light conditions, poor contact with the user's skin, etc.

Robust interpolation techniques were sufficient to deal with the missing values in the heart- rate data. This was verified by comparing interpolated data originally gathered by the Apple Watch with data gathered using the Zephyr device which didnt have such missing values.

K. Data Analysis

We collected data from 11 subjects consisting of 4 doctors and 7 students. The subjects wore the Apple Watch for at least one day and performed workouts during that time. This table shows the number of days for which the subjects wore the Apple Watch and the number of workouts they performed in this time. In total, we have 96 days of data from 11 subjects with 29 workouts to analyze. When the watch is in workout mode, heart-rate is sampled every 5 seconds while in Default mode heart-rate is sampled every 10 minutes. Some values may be rejected by the Watch if it cannot get a reliable reading.

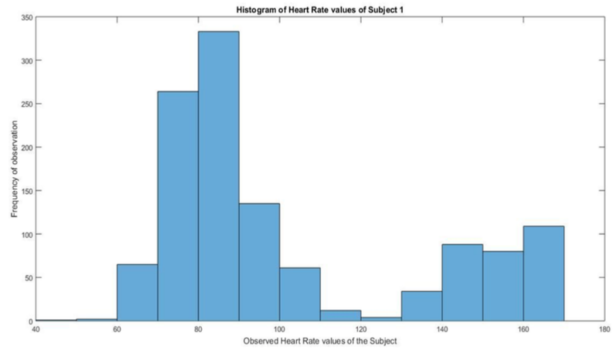

Fig. 9. Histogram of Heart Rate Data collected from Apple Watch for Subject 1

L. Histograms

Histograms provide a graphical display of the overall distribution of the data. The first step of analysis on the heart-rate data we collected was to plot histograms to gain insight into their distribution. On analyzing the histograms we found that the data is distributed across Mixture of Gaussians. The first Gaussian seems to be centered at the Mode of the resting heart-rate while the the second Gaussian is centered at the mode of the Workout data. This figure shows the histogram plot of heart-rate collected from Subject#1s Apple Watch. This subject wore the Watch for 12 days and performed one Workout during this time. One Gaussian can be seen to be centered around a resting heart-rate of about 80-90 bpm while during Workout the Gaussian is centered around 160 bpm. Since the subject performed only one Workout, the shape of the Gaussian in Workout is not perfect but the distribution would be more distributed if more workouts had been performed. The elevation in heart-rate in Subject #1s data does not seem consistent with a 6MWT. Subject #1 is known to be healthy and we suspect that the subject engaged in physically intensive workout for their heart-rate to elevate as much as compared to their resting heart-rate.

Subject #2 wore the watch for 27 days and performed 5 workouts in that duration. We observe a similar Mixture of Gaussians distribution for Subject#2s heart-rate data as well. One of the Gaussians is centered 65 bpm which is suspected to be the resting heart-rate while another Gaussian is centered around 180 bpm. The elevation in heart-rate in Workout as compared to Resting heart-rate for Subject#2 is also not consistent with 6MWT. Subject#2 is known to be healthy and

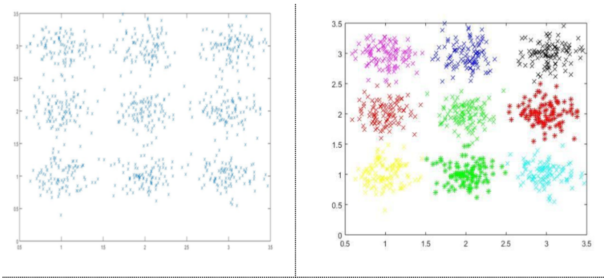

Fig. 10. Result of k-means Clustering on Test Data Set. The same color indicates one cluster.

would only achieve sub maximal peak exercise capacity during a 6MWT.

M. Key Observations

Some of the subjects seem to have performed Workouts other than a 6MWT for their heart-rates to elevate as much as compared to their resting heart-rate. However, we can expect that the distribution of resting and workout data will be similar if data were to be collected from CHF patients. Overall, the distribution seems to be a mixture of two Gaussians centered around the Resting and Workout heart-rate modes,

k-means clustering is an unsupervised learning algorithm that aims to group data points into clusters. The number of clusters is determined a priori and is a hard bias of the algorithm. The algorithm computes the distance of each data point from randomly selected center. Data points close to a selected center are classified into the same cluster. The k-means algorithm updates the selected centers in order to minimize the objective function, i.e. the euclidean distance of all points in a cluster from the selected center.

The figure to the left shows the 2D representation of the data. The figure to the right shows the data clustered into 9 clusters represented with different colors and symbols. We applied k-means clustering algorithm on heart-rate data collected from Apple Watch to classify Workout points from non-workout points based on the observations made in the histograms. We applied k-means clustering with a priori value of k set to 2. The preceding figure shows the k-means clustering results for heart-rate data collected from Subject#2. The Resting heart-rate values are clustered together in the blue cluster while the Workout data points are clustered together in the red cluster. As observed in Subject#2s histogram, there was clear distribution of two Gaussians.

Observations from Clustering for Subject #2:

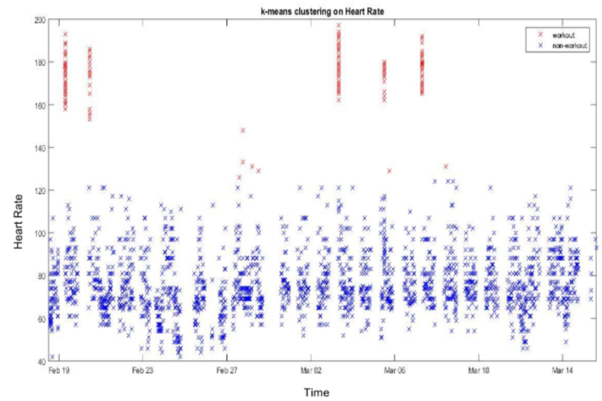

Fig. 11. K-means clustering with K=2 on Heart Rate data collected from Apple Watch for Subject 2

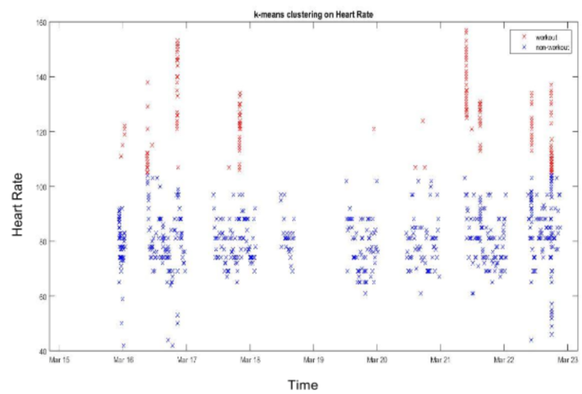

Fig. 12. K-means clustering with K=2 on Heart Rate data collected from Apple Watch for Subject 6

· Total number of sample points: 2644
· Points identified in Cluster 1(Workout): 405
· Points identified in Cluster 2: 2239

For subject 6, k-means clustering identifies the Workout heart-rate points in the red cluster and the non-workout heart-rate datapoints in the blue cluster. Subject#6s histogram also showed a distribution consistent with mixture of two Gaussians. Observations from Clustering for Subject#6:

· Total number of sample points: 1277
· Points identified in Cluster 1(Workout): 473
· Points identified in Cluster 2: 804

Since Subject#6s histogram indicated that the Gaussian of heart-rate datapoints for 6MWT was overlapped with the resting heart-rate workout, k-means clustering

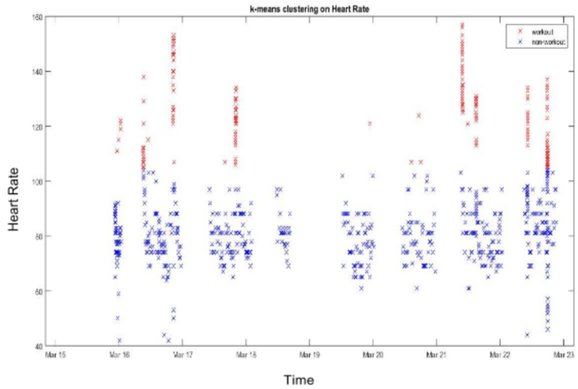

Fig. 13. K?means clustering with K=2 on Heart Rate data collected from Apple Watch for Subject 2.

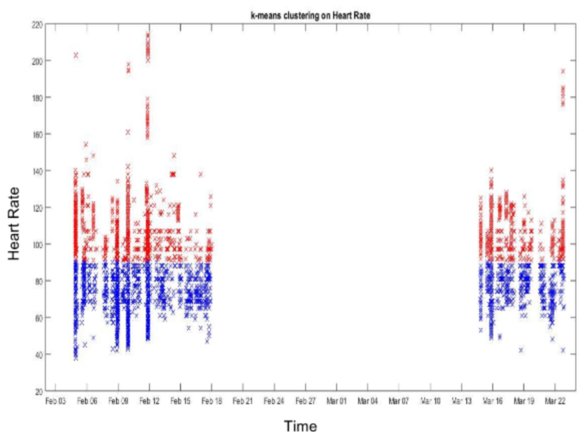

Fig. 14. K?means clustering with K=2 on Heart Rate data collected from Apple Watch for Subject 5.

cannot effectively classify the two clusters with random centers.

Observations from Clustering for Subject#5:

· Total number of sample points: 6490

· Points identified in Cluster 1(Workout): 2374

· Points identified in Cluster 2: 3576

With subject 5, we observe that some subjects have a wider separation between resting heart-rate and workout heart-rate which makes it easier to classify them with simple algorithms such as k-means clustering. Subject#2 is a doctor, so we assume that this subject must not be engaging in a lot of physical activity throughout the day except when making rounds, going to surgery, etc. Subjects 5 and 6 are students and we expect them to be more active regularly (walking to and from class, playing sports, etc.). Differences in lifestyles of the patients might be an explanation for the variability in results for the k-means clustering algorithm.

N. k-means objective analysis

The k-means clustering algorithm updates the centers of the clusters in order to minimize the objective function. The objective function is basically an index of the variance within each cluster. This makes sure that only points that are sufficiently close to the selected center are classified in that cluster. Since the value of k or the number of clusters has to be determined a priori, we used the value of k-means objective function to determine if the dataset can be clustered into one or two clusters.

The k-means objective value is calculated for K=1 and K=2. This dataset also showed a mixture of Gaussians distribution in the histogram. The k-means objective function shows that the dataset is more suitable to be clustered into two clusters. We can use the k-means objective function values to dynamically determine the number of clusters to be set when applying the k-means clustering algorithm.

O. Hidden Markov Models

A Hidden Markov Model (HMM) models a system assuming that it is a Markov process with unobserved or hidden states. In a hidden Markov model, the state is not directly visible and can be estimated on the basis of transition probabilities between states and the probability of emission of the observed parameter. Our next approach to classify workout and non-workout datapoints was to use Hidden Markov Models to decode a given heart-rate time series sequence as workout or non-workout. The model was trained using part of each subjects dataset.

P. Key Observations

The model assigns Workout state to the Workout heart-rate points in the dataset. However, Workout state is falsely assigned to some outliers as well. The blue crosses represent the data points plotted against time axis. The red line superimposed on the plot shows the decoding results of the HMM. A high state shows Workout while a low state shows non-workout.

Subject#3 did not perform any workouts, the HMM model is not able to model the state sequence for Subject#3s heart-rate. In case of Subject#3, we did not have two states to identify in the first place due to which the model failed.

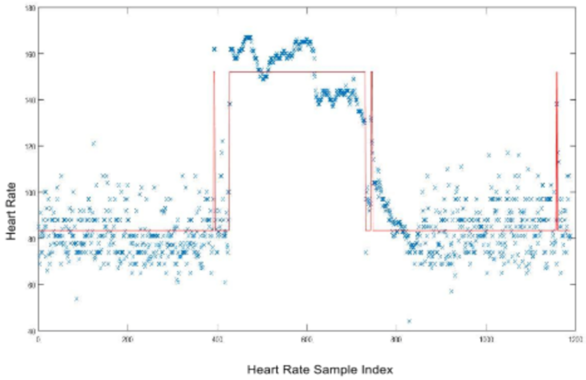

Fig. 15. Hidden Markov Model for Subject 6s Heart Rate data from Apple Watch. The Model identifies two states: Workout and Non-workout

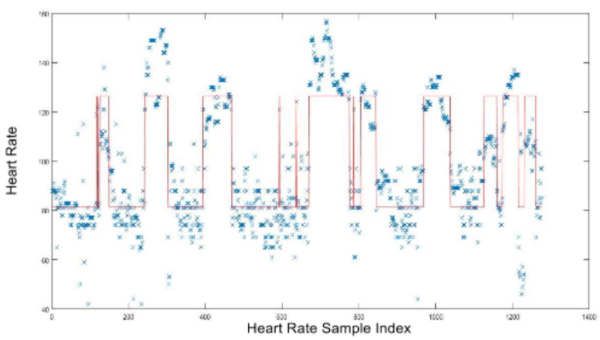

Fig. 16. Hidden Markov Model for Subject 5s Heart Rate data from Apple Watch. The Model identifies two states: Workout and Non-workout

Subject#5 has performed workouts, however, the model cannot correctly identify the Workout state from the non-workout state. This is because the dataset is more random as compared to other datasets. The transition and emission probabilities cannot be reliably determined from such a dataset.

When trained with only workout data, the HMM can assign different states to the workout itself and the recovery of heart-rate. Heart rate recovery is the time period during which the elevated heart-rate decreases steadily to the resting heart-rate when the exercise is complete. Heart rate recovery is an important tool to analyze a subjects performance in the 6MWT.

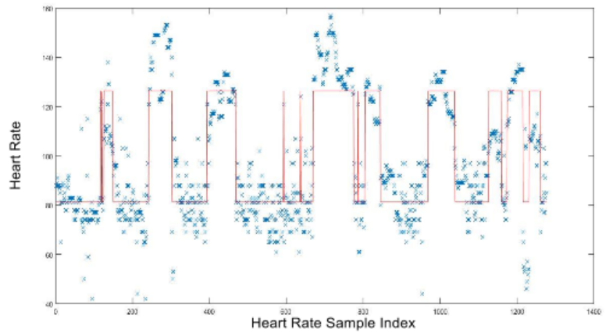

Fig. 17. Hidden Markov Model based Sequence Decoding for Subject 6s Heart Rate data from Apple Watch.

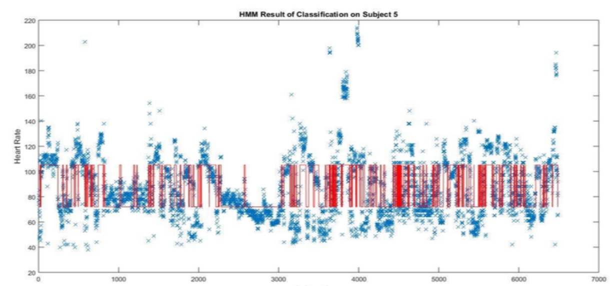

Fig. 18. Hidden Markov Model based Sequence Decoding for Subject 5s Heart Rate data from Apple Watch. The Decoding fails when the dataset is too noisy.

Q. Lag Plots

A lag plot is used to evaluate whether the values in a dataset or time series are random. If the data are random, the lag plot will show no identifiable pattern. If the data are not random, the lag plot will demonstrate a clearly identifiable pattern. Lag plots can also help to identify outliers. A lag plot plots pairs of consecutive datapoints in single dimension to two dimensions. Thus, lag plots help visualize transitions in data that are not apparent in a single dimension. Consider a time series [1,1,2,3,4,4]. The lag plot will be generated as follows:

R. Lag Scatter plots

We can see two evident clusters in the lag plot. One on the bottom-left is the more scattered resting heart-rate cluster while the cluster to the top-right is the workout heart-rate cluster. For Subject#1, we can

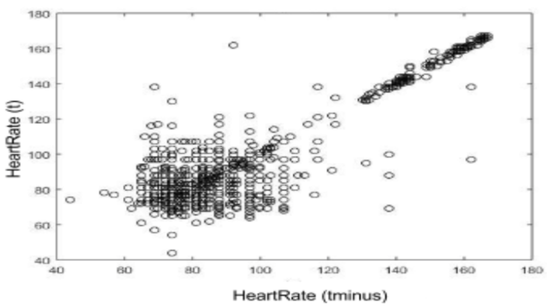

Fig. 19. Lag plot representation of heart rate data collected from Subject 3s Apple Watch. The lag plot shows two separate clusters of workout and non-workout

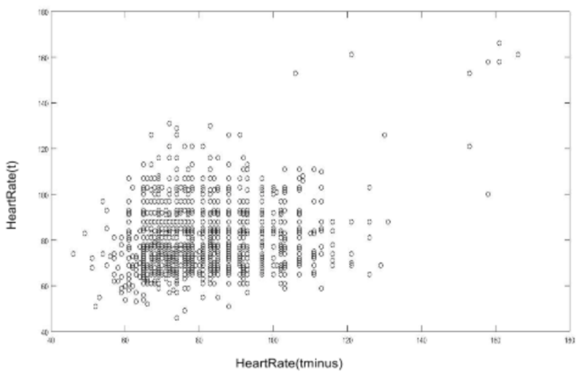

Fig. 20. Lag density plot representation of heart rate data collected from Subject 2s Apple Watch. The datapoints along a positive slope 45 degree line are most densely populated

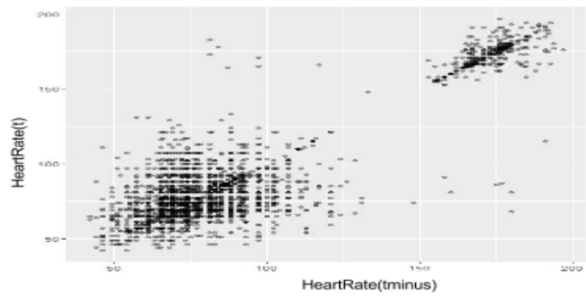

Fig. 21. Lag density plot representation of heart rate data collected from Subject 2s Apple Watch. The data points are densely populated along a positive slope 45 degree line

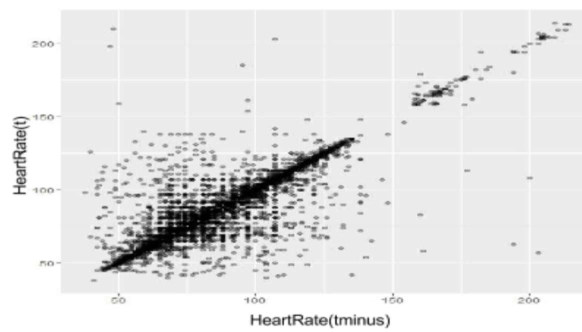

Fig. 22. Lag density plot representation of heart rate data collected from Subject 6s Apple Watch

identify a clear boundary to classify Workout heart-rate from non-workout heart-rate.

For subject 6, this data is distributed along a 45 degree line in the plot. For this Subject, the workout heart-rate is quite elevated as compared to the resting heart-rate. The sampling rate during Default mode is every 10 minutes which explains the randomness in the lower cluster of the lag plot. Heart rate is more variable throughout the day. However, heart-rate is less likely to change every 5 seconds than it is every 10 minutes. This lag plot doesnt put resting heart-rate data and workout heart-rate data in two different clusters. Rather, the heart-rate is distributed over a continuous range between 400-190 bpm.

With subject 3, we can observe that the heart-rate data forms a cluster on the lower-left corner of the plot. This supports our hypothesis for the other Subjects that the rest heart-rate forms one cluster in the lag plot. Few outliers are also seen on the lag plot. We believe that such outliers might be a result of some physical activity during the day, such as, climbing stairs, walking to the car, etc.

Our next approach to analyze lag plots was to visualize density of the heart-rate values in the lag plot. The heart-rate values are dense along the 45 degree line and become less dense farther away from the line. In some of these figures, we observe a similar hotspot of heart-rate along the 45 degree line. Since the scale on both axes is the same, this implies that heart-rate remains more or less the same in 10 minute intervals during rest and 5 second intervals during workout. The outliers can be accounted for the ramp up in heart-rate and subsequent ramp down before and after the workout respectively. Some of the outliers may also stem from unaccounted physical activities which are not the 6MWT

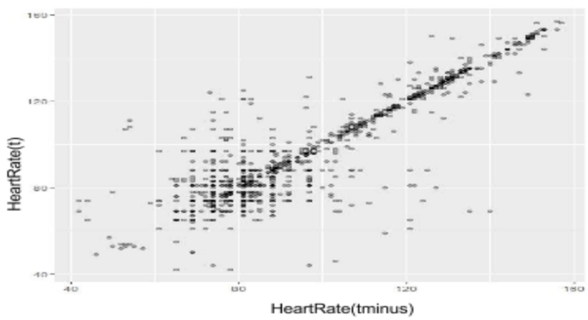

Fig. 23. Lag density plot representation of heart rate data collected from Subject 6s Apple Watch.. The heart rate is measured in BPM.

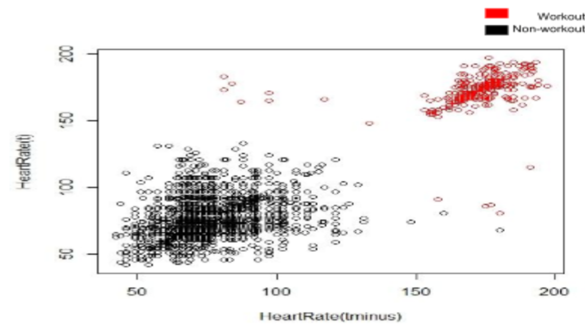

Fig. 24. K-means clustering on lag plots for Subject 2s heart rate data. The algorithm clusters workout from non-workout

or the Watch sensor reading incorrect values due to poor environmental conditions.

The preceding figure shows a lag density plot of the heart-rate data collected from the Zephyr chest band. The Zephyr chest band is extremely accurate and takes a reading every 1 second. There were no missing values found in the Zephyr chest bands data. If the band is unable to take a reliable reading, the missing value is handled internally. The sensor turns on only when the band is strapped around the chest and the electrodes make contact with the skin. This allows us to measure heart-rate only during workout.

We plotted a lag density plot from workout data obtained from Subject #6. There are very few outliers in the plot due to the high sampling rate of the Zephyr band.

S. Clustering on Lag Plots

To combine the results we obtained from k-means clustering and lag plots, we applied k-means clustering to lag plots. The goal was to separate the Workout points from the non-workout points. The a priori k value was set to 2. The k-means clustering algorithm identifies two clusters: a resting heart-rate cluster (in black) and a workout heart-rate cluster (in red). The outliers are very few and they are assigned to either of the two clusters.

We see from the figure that the k-means clustering algorithm attempts to divide the points along a median value rather than identifying two clusters. From the k-means objective analysis, we know that this subjects dataset is not suitable for k-means clustering.

This subject wore the watch only during workouts. The k-means clustering algorithm however has a hard bias for searching two clusters in the dataset. Therefore, k-means clustering to identify workout fails in this case. We can perform analysis similar to k-means objective function and dynamically select how many clusters should be identified in the dataset.

T. Summary

1. The Six Minute Walk Test is the most popular and simplest tool to analyze the functional capacity of a Congestive Heart Failure patients heart.

2. The Apple Watch has the most accurate heart-rate sensor among wrist based heart-rate sensors. However, the Watch rejects many readings and sometimes records incorrect values. The Zephyr chest band can be used for calibration and missing values can be estimated using interpolation techniques.

3. The data collected from subjects shows a distribution consistent with mixture of two Gaussians.The two Gaussians represent two states for each subject: Workout and Rest.

4. Techniques such as k-means clustering fail to separate the two Gaussians when they overlap or when the distribution is not a mixture of Gaussians (when subjects dont perform workouts or dont wear the watch at rest)

5. Lag plots help evaluate patterns in time series data. We observe that different subjects show different patterns with lag plots. We may need to build a separate model for each subject to identify subtle trends in the data.

IV. Experiments with the Apple Watch and Zephyr Chest band

Three different experiments were conducted in order to determine the possible causes of missing and inaccurate data. One subject wore two Watches, one on each arm, and performed workouts with different settings

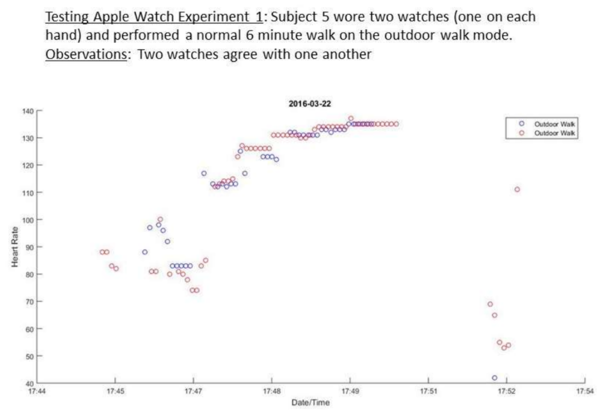

Fig. 25. Heart-rate data when two watches are worn by the same person.

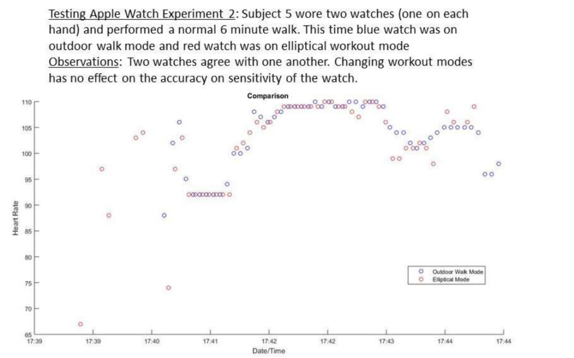

Fig. 26. Heart-rate data when two watches are worn by the same person with one on outdoor walk mode and one on elliptical walk mode

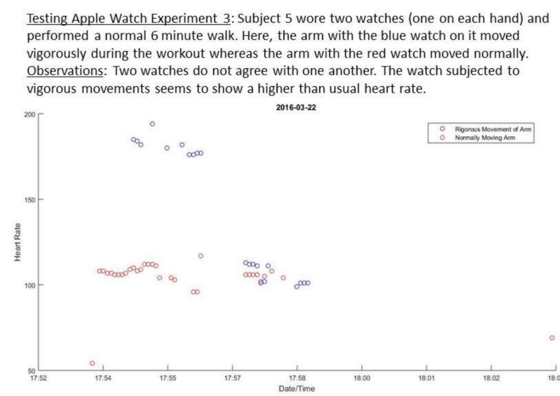

Fig. 27. Heart-rate data when two watches are worn by the same person with one hand doing vigorous movement and other normal movement

and actions that could provoke data loss. The three test descriptions are listed below.

1. Normal settings on each Watch.

2. One Watch on Elliptical mode, one Watch on Outdoor Walk mode.

3. One Watch with normal arm movement, other with vigorous movements (both on Outdoor Walk mode).

A. Observations and Results

Test 1 Two watches on the same person Subject wore one watch on each hand and conducted the Six Minute Walk Test normally as per the protocol. The goal was to observe if either one of the watches collected better or cleaner data than the other, suggesting difference in precision and accuracies between watches.

Both watches show a similar trend in heart-rate rise and fall during the workout and even have similar trends in missing data points. Both watches rejected data for about 2 minutes after the workout possibly due to abrupt change in heart-rate levels. There is no deviation in the data collected between the two watches.

Test 2 Effect of Workout Modes

Subject wore two watches. One on each hand. The left watch was on outdoor walk mode and the right watch was on elliptical mode.

The workout modes do not seem to affect the data. The sampling frequency, the precision and the accuracy of both the watches remains similar to each other for the Six Minute Walk even if they are on different workout modes. Another interesting observation is that the watch takes a few seconds to settle on the actual heart-rate reading.

Test 3 Effect of Rigorous Movements

Subject wore two watches one in each hand. The right arm was subjected to rigorous movements during the walk while the left hand moved back and forth normally. The goal was to observe the effects of jerky movements to the data collected.

The watch that was subject to jerky movements recorded higher readings than the one which was moving smoothly. There were more missing points in this watch. It is therefore evident that if the subject moves their arms rapidly during the workout, the accuracy of the data collected is affected.

B. Conclusion

The Apple Watch was determined to be the best wearable device for this project, but multiple issues came

Zephyr	Apple Watch
No. of data points - 752	No. of data points - 117
Mean - 107.1	Mean - 101.4
Median - 110.5	Median - 95
Mode - 117	Mode - 95

up during this phase of data collection. Experiments were performed to try to explain some of these issues. The following conclusions were made -

1. Different watches have the same level of precision, accuracy and vulnerability to noise.

2. Changing the workout mode of the watch does not have any effect on the data collected. The sampling rate and the tendency to reject noisy data points remains unchanged across different modes.

3. Excessive movement of the arm wearing watch might result in noisy data which the watch records as higher heart-rate. Subjects must therefore be careful not to do jerky movements of the arm while doing the 6 minute walk.

4. The watch sometimes rejects data after the workout when there is an abrupt drop in heart rate levels.

5. The watch requires a few seconds of settling time to get a reliable heart-rate reading.

C. Experiments Conducted with Apple Watch and Zephyr Chest band

In order to test the deviation from ideal collection and representation of heart-rate data, we comapared data collected from Zephyr chest band with that collected from Apple watch. We had 3 subjects wear both the Zephyr chest band and the Apple watch. Two of them continued to wear these devices for 3 hours before performing the 6 minute walk according to the improved protocol and one of them only wore it during the workout. The following section describes the observations and inferences from the experiments.

Zephyr has a sampling frequency of 1 Hz as opposed to Apple Watchs sampling frequency of 0.2Hz during workout modes. For some experiments we downsampled Zephyr to match Apple Watch. On a general scale, zephyr was always more accurate than the Apple watch since it is in contact with the subjects chest making it subjected to lesser motion atrifacts. 6.2.1. Observations and Inferences Test 1 Two devices only during workout mode Here are some statistics about the data collected. They reflect the effect of a higher sampling frequency in the chest band.

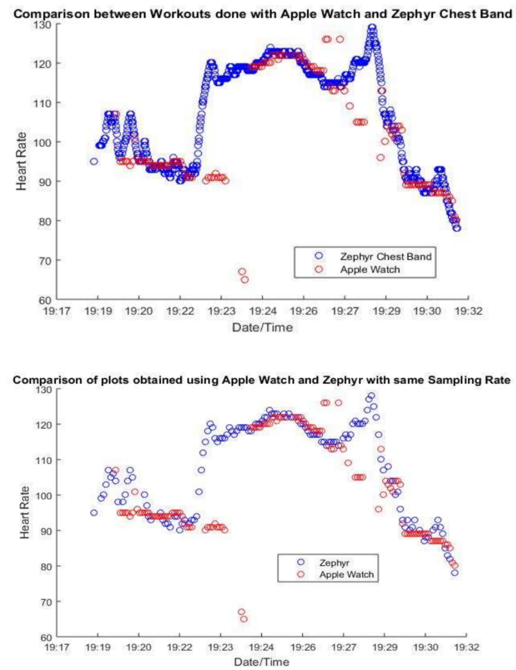

Fig. 28. Original comparison (top) and downsampled Zephyr comparison (bottom)

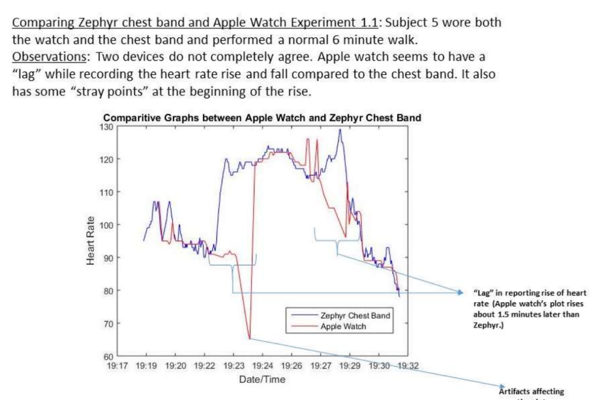

Fig. 29. Line plots of plots in previous figure. Lag and artifacts have been shown.

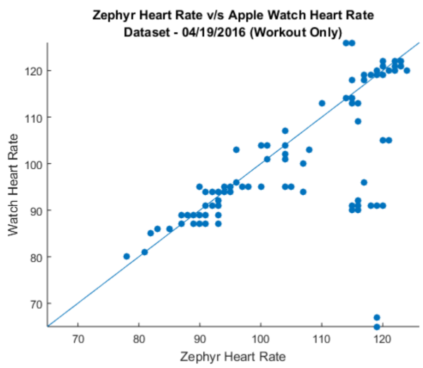

Fig. 30. Here zephyr and apple watch data points have been plotted against each other to highlight the fact that they agree for most parts except the artifact regions

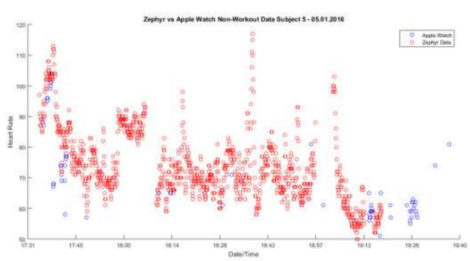

Fig. 31. Non-workout data from Apple watch superimposed on Zephyr data

* Consistent with our previous observations, Apple watch has many missing points - mainly during the heart-rate rise and fall. Zephyr measures better!

* Some missing points also observed in the middle of a stable workout heart-rate possible motion artifacts.

Test 2 Two devices worn for a period of 3 hours capturing workout as well as nonworkout data In this experiment, the subjects wore the chest band and Apple watch for 3 hours and performed the 6 minute walk at the end.

-During non-workout time periods, the Apple watchs samples of data are consistent with the corresponding Zephyr data points (most of the times)

-During workout, differences can be observed during heart-rate rise and fall.

-When the Apple watch finally stabilizes to a given heart-rate level, the watch and zephyr data points agree with each other (with a reasonably tolerant difference). In the second dataset, the watch and Zephyr data seem

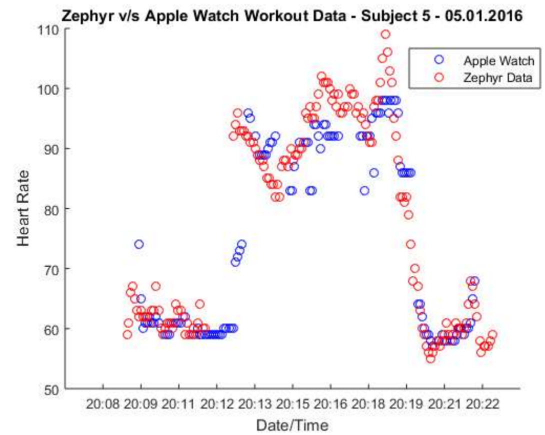

Fig. 32. The workout region of the datasets

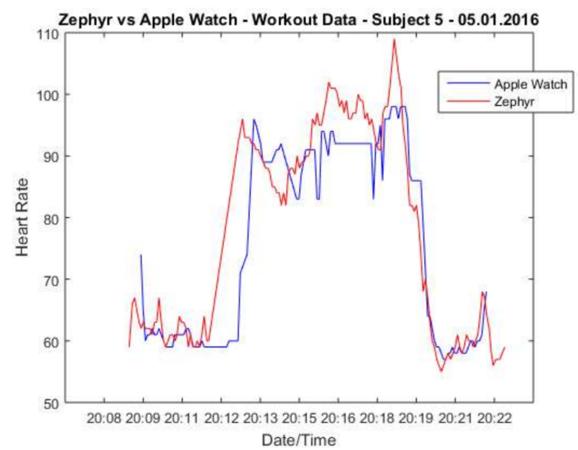

Fig. 33. Line plot of previous figure

to differ more than usual.

D. Conclusion

1. Apple watch data points agree with Zephyrs data points for most of the time.

2. The watch has some missing points during heart-rate rise and fall that can be fixed with interpolation techniques to condition the data. Using the ideal plots obtained by Zephyr for a 6minute workout, a model can be prepared (by fitting a polynomial curve to the plot) and this can be used to interpolate watchs data when there are missing points.

3. This technique can also be used to get rid of the artifacts at the beginning of the heart-rate rise.

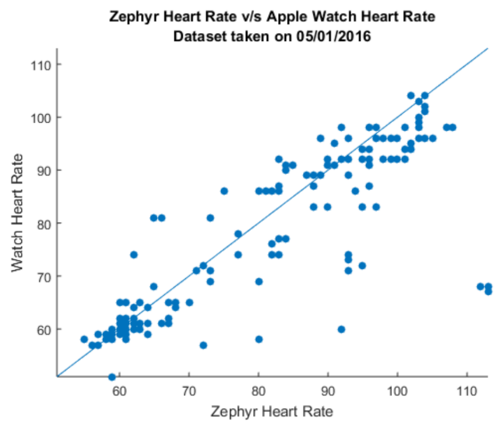

Fig. 34. Comparison of Zephyr and Apple watch data

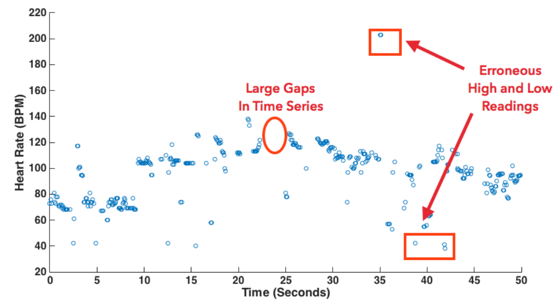

Fig. 35. Raw Apple Watch Data

4. Thus, by comparing Zephyr with the Apple Watch we could fix some issues with the watchs plot.

E. Summary

This phase of the project thus dealt with the initial data collection and analysis phase for our main study. With our work, we discovered some issues with data collection which we could resolve with a more streamlined data collection protocol (which will be used in the next phases of the study)

With the analysis of the data collected as well as the data from the experiments, we discovered general trends of the data in workout and non-workout modes. In the next phases of this study, these discoveries and inferences can be used to train a model out of the data collected so that detecting changes and identifying abnormalities with respect to the model becomes more straightforward.

V. Improving Apple Watch Data

As we have seen, data collected by the Apple watch has many problematic qualities compared to data collected by a chest-mounted device. In particular, the data includes many erroneous heart-rate readings with unrealistic values, such as 35 or above 200. Additionally, whenever the heart is undergoing a rapid change in heart-rate the watch ceases collecting data at all. Figure 35 shows an example of raw Apple watch heart-rate data. This being said, there are many attractive qualities to the watch. It is unobtrusive, inexpensive, and a desirable piece of consumer technology.

To address the problems with the Apple watch data, we have developed an algorithm to automatically remove erroneous readings, interpolate holes in the data, and make the sample frequency uniform. To begin, we train a linear Support Vector Machine with 30 example of erroneous heart-rate readings and 30 examples of correct heart-rate readings. This model is trained using the following features for given datapoint, y_i:

- The value of y_i in BPM
- The slope distance $\frac{\delta y}{\delta x}$ to y_{i-1}, y_{i-2}, y_{i+1}, and y_{i+2}
- The average heart-rate of all data within 5 seconds of y_i

For each of these features, a linear SVM parameter θ_i is learned. For each datapoint the following calculation is conducted:

$$F(Y) = \sum_{i=1}^{d} \theta_i f_i$$

Each datapoint, y, for which $F(y) \geq 0$ is denoted as erroneous and removed from the time series.

While we have no direct means of confirming errors in the data, filtering the data using a Support Vector Machine corrects the most obvious errors in a given time series. Figure 36 shows the same time series as figure 35, after SVM-filtering has been performed.

At this stage the gaps in the time series have actually widened due to the removal of the erroneous data. Thus, we now apply polynomial interpolation for the entire series, leaving us with a continuous function $H(x)$ that represents the heart-rate time series. This continuous function may then be sampled at any discreet frequency we desire, or we may simply use it as a continuos signal. The end-product of the filtering and interpolation is shown in figure 37

We have instances in which we have simultaneously collected Apple Watch and Zephyr heart-rate data, and we believe the Zephyr chest-mounted represents a very

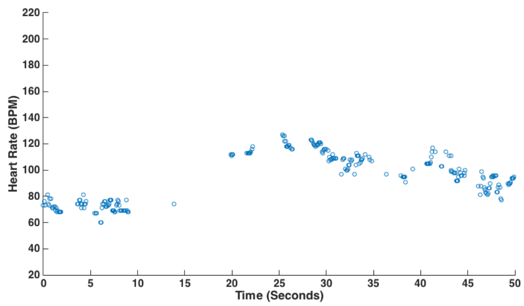

Fig. 36. SVM-Filtered Apple Watch Data

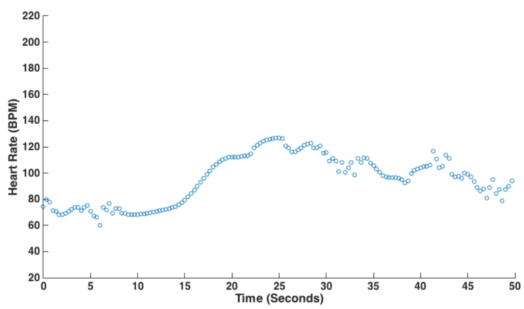

Fig. 37. Filtered and Interpolated Apple Watch Data

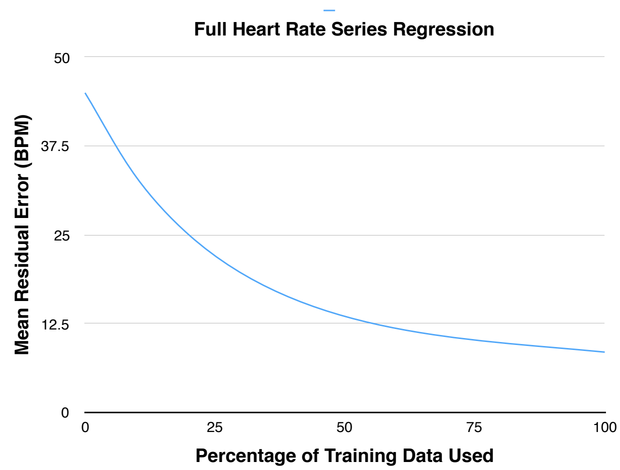

Fig. 38. Mean Residual Error for LV-ARIMA Applied to Heartrate Series

Method	Mean Residual Error
ARIMA(15, 2, 10)	9.1 BPM
LV-ARIMA(15, 2, 10)	7.2 BPM
Current Value Baseline	15.2 BPM
Mean Value Baseline	12.1 BPM

TABLE II
MEAN RESIDUAL ERROR FOR PREDICTED HEARTRATE

accurate representation of the user's heart-rate. Before data correction is applied, the mean difference between the Zephyr and Apple watch data was 6.9 BPM. After our algorithm is applied to the Apple watch data, this difference decreases to 2.1. This result allows us to get very close to the heart-rate accuracy of a professional grade chest-mounted device, with the more desirable Apple watch device.

VI. Latent-Variable ARIMA Modeling of Continuous Heart Rate Signals

We now consider the task of utilizing the latent-variable ARIMA model described in section II-D to model the dynamics of continuous heart-rate data. Models were trained using data taken during workouts, data taken during resting periods, and a mixture of both. A total of 48 examples of each type of time series were collected, from 6 different volunteers. Figure 38 shows the training curve for this model, when predicting heart-rate 15 seconds into the future. Table II shows a detailed comparison of mean residual errors against baseline models.

For these experiments, we use an ARIMA(15, 2, 10) model with 3 hidden state values. This produces 75 ARIMA parameters that form a unique representation of the dynamics of an individuals heart. Applying Principle Component Analysis to this representation allows us to reduce the dimensionality down to 2, to form a visual representation of the models learned. Figure 39 depicts this 2-dimensional representation, and we can see a very clear separation between ARIMA models trained using resting data compared to active exercise data.

We believe that this PCA representation of heart dynamics may provide us with means to track changes in heart dynamics over time, potentially allowing us to detect declines in heart health before patient's become symptomatic. We are currently in the process of deploying this method on a dataset of ECG data from 15 patients with severe CHF during hospital stays.

Another particularly interesting application for this time series analysis is to model an individual's recovery period after physical activity. Clinicians believe that when a heart can quickly return to its resting heart-rate after a period of activity, this is an indication of strong heart health. Figure 40 depicts the recovery period

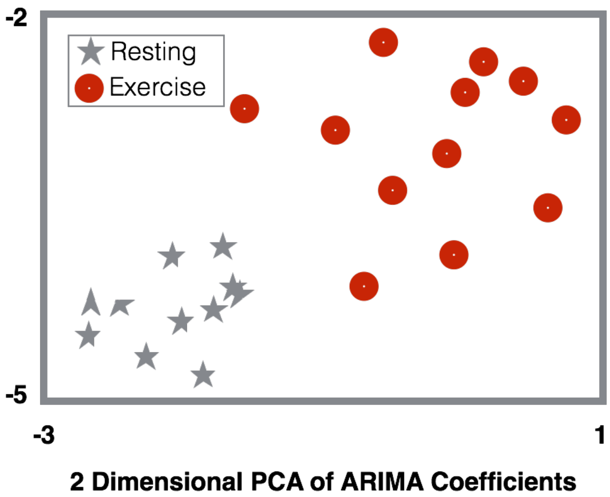

Fig. 39. 2-Dimensional PCA Representation of ARIMA Coefficients

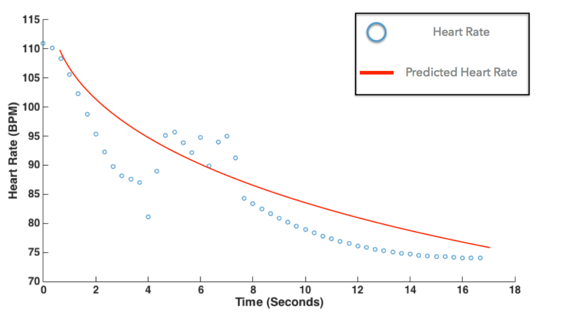

Fig. 40. ARIMA Prediction of Recovery After Exercise

predicted by the ARIMA model, overlaid with the actual observed recovery.

Table III compares various methods when predicted recovery time. In particular, the latent variable ARIMA model achieves a mean residual error of 6.4 seconds, with a ± 3.1 seconds .95 confidence bound. This indicates statistical significance when comparing again the baseline method of relying on an individual's average recovery time.

Clinically, this framework for estimating recovery time would allow us to detect deviations in an individual's heart health, particularly if the observed recovery time greatly exceeds the modeled recovery time. This provides us with another tool to detect changes in a CHF patient's condition before they become symptomatic.

Method	Mean Residual Error
ARIMA(15, 2, 10)	6.4 Seconds
LV-ARIMA(15, 2, 10)	4.9 Seconds
Mean Recovery Time	10.4 Seconds

TABLE III
MEAN RESIDUAL ERROR FOR PREDICTED RECOVERY TIME

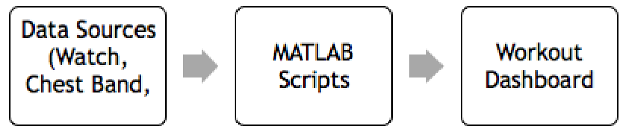

Fig. 41. Scraping Pipeline

VII. DATA VISUALIZATION

Data visualization is the graphical representation of statistical data that isnt easily perceivable to humans. Visualization is a key component with regard to data analysis. Associating abstract information and its variations with physical attributes enables easier perception and communication of the information [12]. Visualization enables the average human being to easily identify trends (if any) in the abstract information. Graphical data is also better imprinted in the human memory [13]. In the present context of health-related applications, graphical visualization of physiological data gathered from multiple sources can convey significant information about the health of an individual, the trends in the variation of certain physiological parameters and can also enable prediction of the future health of the individual. For patients of Congestive Heart Failure (CHF) in particular,

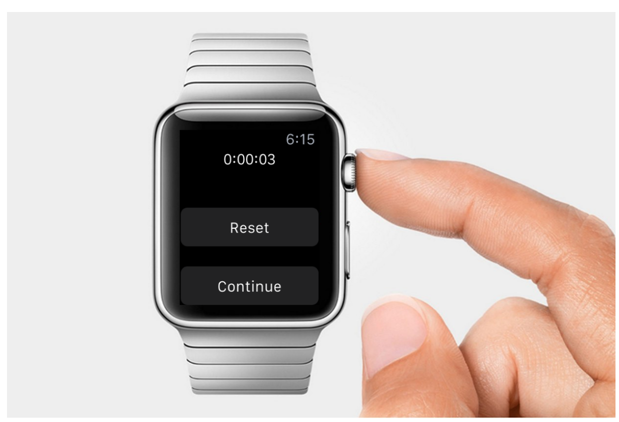

Fig. 42. Apple Watch Data Collection Application

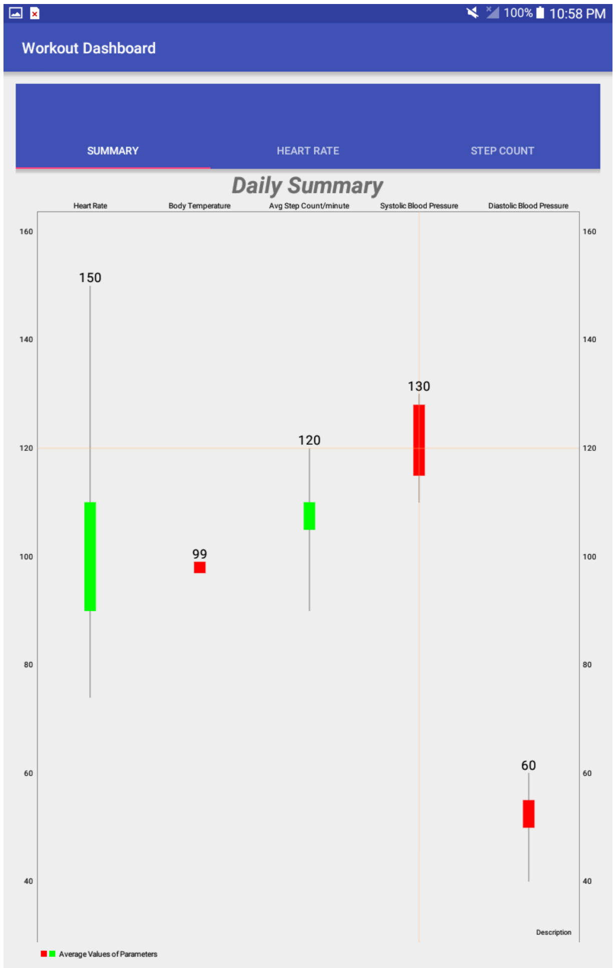

Fig. 43. Summary screen of the Workout Dashboard

visualization of data pertaining to heart-rate is imperative to determine patterns in the variation of resting and active heart-rate. A study conducted by a research group at UW Seattle showed that when CHF patients were presented with a score of their risk of re-admission, the actual number of re-admissions were drastically reduced [14]. Average heart-rates during workouts, the time taken to revert to resting heart-rate after a workout (ramp-down time), and variations in the resting and active heart-rate are all indicators of the status of the heart. In order to enable effective visualization of the relevant physiological parameters, the following two components need to be developed: An effective scraping technology for scraping data from the various devices of interest (Apple Watch, Zephyr Chest Band, etc.) A dashboard to display the collected data in an intuitive format.

A. *Data Scraping*

The scraping of data from the devices of interest is enabled by means of an iOS application that runs on the Apple Watch and an iPhone 5s (or above). This application enables the user to initiate data collection, stores the data in an XML file and allows the user to end data collection. Figure 41 shows the flow of data from source to visualization. Additionally, we have built an iOS and WatchOS application that automates the process of collecting heart-rate data. The application, depicted in figure 42, supervises the timing of a 6 minute walk, collected resting heart-rate data, and remotely sends the data to researchers automatically. This application will allow us to remotely collect data from hospitalized volunteers suffering from Congestive Heart Failure.

B. *The Dashboard*

The Workout Dashboard is an Android Application that runs on Android versions 4.4 (KITKAT) and above. It graphically reports the physiological status of the user using data scraped from various sources. The present version consists of three tabs: Summary: This tab displays the overall summary of the individual with a graphical representation of variations in height, weight, heart-rate, step count, distance covered, body temperature and energy burned. Variations observed in these parameters could be an early indicator of ill-health. Heart Rate: This tab graphs the variations in heart-rate over the past 7 days, the hourly heart-rate, heart-rate obtained from an Apple Watch vs another source and the ramp down pattern after a workout. The chart mapping heart-rate over the past 7 days enables the user to view the maximum, minimum and average heart-rate for each day in one consolidated view. The chart mapping the heart-rate through the day could be used to spot irregularities that occurred without a reason to explain it. Ramp down time is an indicator of the well-being of the heart. Longer ramp-down times are usually an indicator of poor heart health. Finally, a comparison between heart-rate data obtained from multiple devices enables the user to determine the accuracy of the projected results.

Step Count: This tab graphically displays the total step count over the past 7 days and the step count per minute for the day. The consolidated view of the step count over the past week enables the user to keep track of the level of activity for each day. A comparison chart between heart-rate and the step count enables the user to spot irregularities in heart-rate that cannot be explained by physical activity.

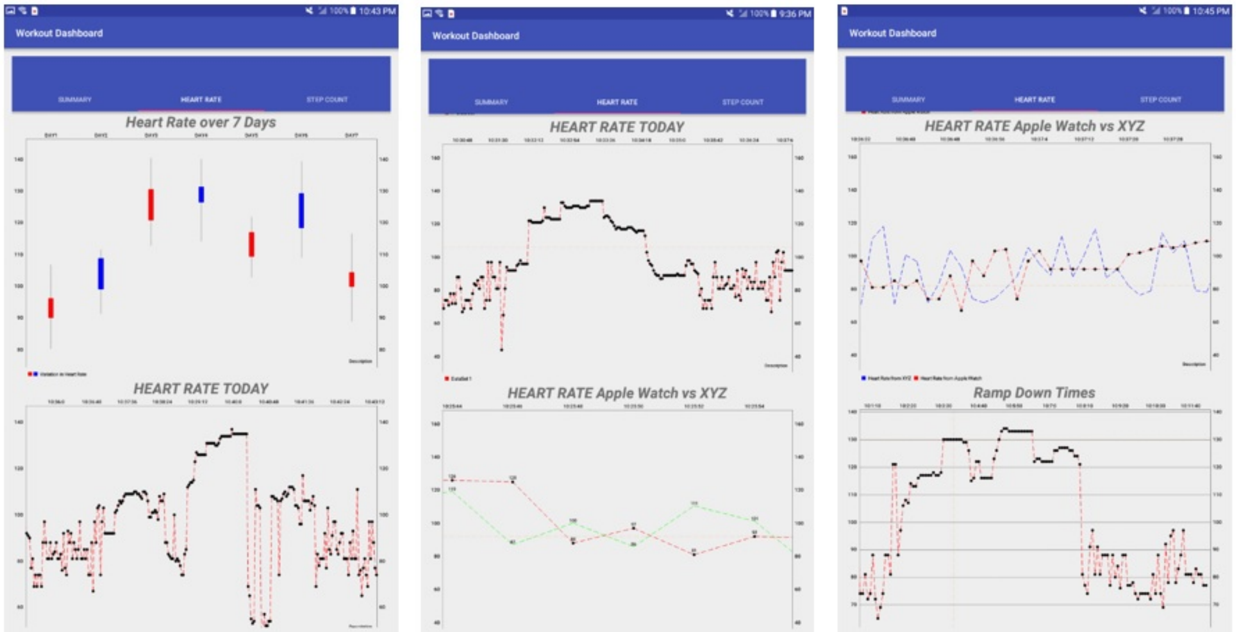

Fig. 44. Heart Rate Tab of the Workout Dashboard

The Workout Dashboard can be easily modified to include other physiological sensors and contains place-holders for the same. Better visualizations can be easily incorporated depending on user feedback. Future versions can include data from multiple devices of interest.

VIII. Key Research Accomplishments

- We have introduced a Latent Variable Autoregressive Integrated Moving Average model that can adapt quickly to rapid changes in the health of an individual living with Congestive Heart Failure.
- We have demonstrated preliminary results that suggest that we can use telehealth data to predict deteriorations in a CHF patient's health up to two days in advance, potentially leading to live saving interventions.
- We have collected real world heart-rate data, and analyzed this data using Hidden Markov Models, k-means cluster, lag plotting, and aggregate statistical analysis.
- We have demonstrated that we can use the latent variable ARIMA model to accurately model the dynamics of a heart-rate time series, either at rest or during exercise. This allows us to model activity recovery time, potentially giving us another means of detecting early warning signs of exacerbation in CHF patients.
- We have developed an iOS and WatchOS app that allows us to scrape data remotely from volunteer's devices, giving us the means to accumulate a larger dataset.
- We have built tools to visualize large, multi-dimensional datasets of medical data to make it easier for clinicians and patients to understand the state of their health.

IX. Reportable Outcomes

- Paper submission to the ICMLA conference
- Keynote Talk at the IEEE Mobile Data Management Conference, Pittsburgh, PA, 2015.
- Invention Disclosure
- Scraping tool able to collect data from different types of sensor devices and a dashboard for visualizing this data

X. Conclusion

We anticipate that the sensors data and our machine learning model will allow analysis of multiple data points. In reality, these would be far too many data points for the physician and medical team to decipher independently on a daily basis. Our team is using the

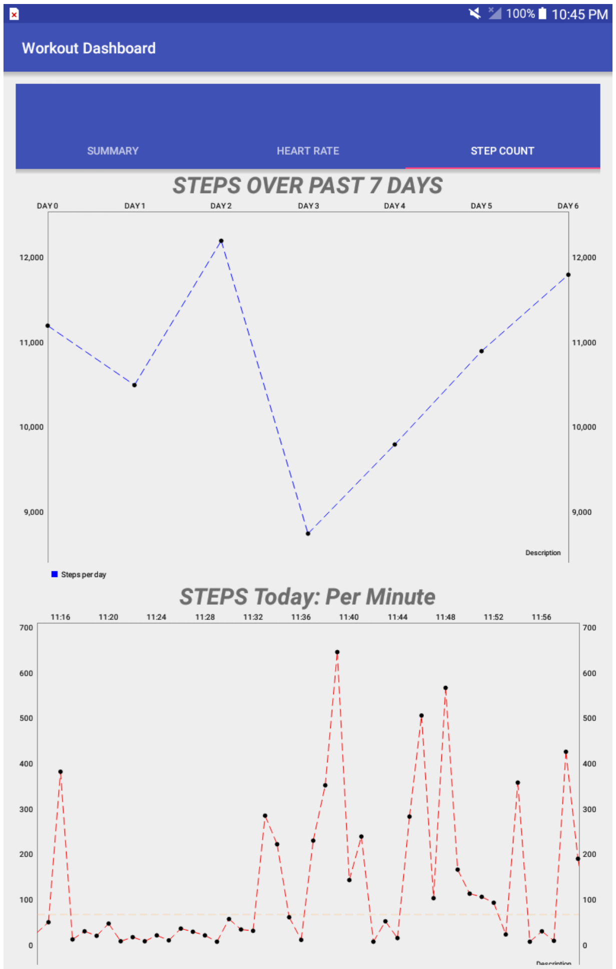

Fig. 45. Step Count Tab of the Workout Dashboard

large amount of data, analyze it and can give the medical team a concise reading of this data in an easily manageable format. Previous studies have focused on these individual parameters separately and have shown modest improvement in parameters such as hospitalization, but we believe including additional data in a collective format will allow for improved outcomes. We have demonstrated a Latent Variable Autoregressive Integrated Moving Average model that can adapt quickly to rapid changes in the health of patients with CHF. Our preliminary results indicate that we can use telehealth data to predict deteriorations in a CHF patients health up to two days in advance, potentially leading to life saving interventions. Our study showsthat wearable sensors can measure the level of physical activity in persons with cardiovascular disease, including patient workouts, as well as their heart rate. We have demonstrated that we can use the latent variable ARIMA model to accurately model the dynamics of a heart-rate time series, either at rest or during exercise. That way we can model activity recovery time, as another means of detecting early warning signs. We are now going to add sleep data to this study as well. We have developed an iOS and WatchOS app that allows us to scrape data remotely from users devices, providing us a capability to accumulate a larger dataset.

References

[1] S. Stewart, J. E. Marley, and J. D. Horowitz, "Effects of a multidisciplinary, home-based intervention on planned readmissions and survival among patients with chronic congestive heart failure: a randomised controlled study," *The Lancet*, vol. 354, no. 9184, pp. 1077–1083, 1999.

[2] A. F. Jerant, R. Azari, and T. S. Nesbitt, "Reducing the cost of frequent hospital admissions for congestive heart failure: a randomized trial of a home telecare intervention," *Medical care*, vol. 39, no. 11, pp. 1234–1245, 2001.

[3] K. Zolfaghar, N. Meadem, A. Teredesai, S. B. Roy, S.-C. Chin, and B. Muckian, "Big data solutions for predicting risk-of-readmission for congestive heart failure patients," in *Big Data, 2013 IEEE International Conference on*. IEEE, 2013, pp. 64–71.

[4] P. Melillo, N. De Luca, M. Bracale, and L. Pecchia, "Classification tree for risk assessment in patients suffering from congestive heart failure via long-term heart rate variability," *IEEE journal of biomedical and health informatics*, vol. 17, no. 3, pp. 727–733, 2013.

[5] G. Liu, L. Wang, Q. Wang, G. Zhou, Y. Wang, and Q. Jiang, "A new approach to detect congestive heart failure using short-term heart rate variability measures," *PloS one*, vol. 9, no. 4, p. e93399, 2014.

[6] J. Polisena, K. Tran, K. Cimon, B. Hutton, S. McGill, K. Palmer, and R. E. Scott, "Home telemonitoring for congestive heart failure: a systematic review and meta-analysis," *Journal of Telemedicine and Telecare*, vol. 16, no. 2, pp. 68–76, 2010.

[7] L. Bauwens and D. Veredas, "The stochastic conditional duration model: a latent variable model for the analysis of financial durations," *Journal of econometrics*, vol. 119, no. 2, pp. 381–412, 2004.

[8] A. Ang and M. Piazzesi, "A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables," *Journal of Monetary economics*, vol. 50, no. 4, pp. 745–787, 2003.
[9] A. E. Gelfand and P. Vounatsou, "Proper multivariate conditional autoregressive models for spatial data analysis," *Biostatistics*, vol. 4, no. 1, pp. 11–15, 2003.
[10] F. Z. N. Y. Z. Jinliang, "The impacts of monetary policy shock on stock market liquidity: An empirical study based on markov switching vector autoregression [j]," *Journal of Financial Research*, vol. 7, p. 006, 2011.
[11] J. Contreras, R. Espinola, F. J. Nogales, and A. J. Conejo, "Arima models to predict next-day electricity prices," *IEEE transactions on power systems*, vol. 18, no. 3, pp. 1014–1020, 2003.
[12] S. Few and P. EDGE, "Data visualization: past, present, and future," *IBM Cognos Innovation Center*, 2007.
[13] E. Segel and J. Heer, "Narrative visualization: Telling stories with data," *IEEE transactions on visualization and computer graphics*, vol. 16, no. 6, pp. 1139–1148, 2010.
[14] V. R. Rao, K. Zolfaghar, D. K. Hazel, V. Mandava, S. B. Roy, and A. Teredesai, "Readmissions score as a service (raas)."