WANDA: An End-to-End Remote Health Monitoring and Analytics System for Heart Failure Patients

Mars Lan[1], Lauren Samy[1], Nabil Alshurafa[1], Myung-Kyung Suh[1], Hassan Ghasemzadeh[1],
Aurelia Macabasco-O'Connell[2], Majid Sarrafzadeh[1]

[1]Computer Science Department
University of California, Los Angeles
Los Angeles, CA 90095

[2]School of Nursing
University of California, Los Angeles
Los Angeles, CA 90095

{marslan, lsamy, nabil, dmksuh, hassan, majid}@cs.ucla.edu   amoconnell@sonnet.ucla.edu

ABSTRACT

Recent advances in wireless sensors, mobile technologies, and cloud computing have made continuous remote monitoring of patients possible. In this paper, we introduce the design and implementation of WANDA, an end-to-end remote health monitoring and analytics system designed for heart failure patients. The system consists of a smartphone-based data collection gateway, an Internet-scale data storage and search system, and a backend analytics engine for diagnostic and prognostic purposes. The system supports the collection of data from a wide range of sensory devices that measure patients' vital signs as well as self-reported questionnaires. The main objective of the analytics engine is to predict future events by examining physiological readings of the patients.

We demonstrate the efficiency of the proposed analytics engine using the data gathered from a pilot study of 18 heart failure patients. In particular, our results show that the advanced analytic algorithms used in our system are capable of predicting the worsening of patients' heart failure symptoms with up to 74% accuracy while improving the sensitivity performance by more than 45% compared to the commonly used thresholding algorithm based on daily weight change. Moreover, the accuracy attained by our system is only 9% lower than the theoretical upper bound. The proposed framework is currently deployed in a large ongoing heart failure study that targets 1500 congestive heart failure patients.

Categories and Subject Descriptors

J.3 [**Computer Applications**]: Life and Medical Science – *Health*

General Terms

Design, Human Factors, Measurement

Keywords

Wireless health, Remote health monitoring, Medical data mining, Machine learning

1. INTRODUCTION

Heart failure (HF) is one of the leading causes of death in the US and around the world for adults over 65 years of age [1]. Nearly one fourth of the patients treated for HF are re-hospitalized within 30 days, and almost half of them are readmitted within 6 months [2]. These unplanned readmissions are estimated to cost the American healthcare system more than $17 billion annually, or 15% to 20% of the total Medicare expenditure in acute hospital care [3].

To reduce the morbidity, mortality, and economic cost associated with HF, remote health monitoring appears to be a promising solution that can work at scale. However, a major challenge to the realization of a large-scale remote monitoring system is the ability to collect, store, and process the large amount of data gathered from the sensors in an effective, robust, and automated fashion.

Furthermore, a truly successful remote health monitoring system should also be able to perform intelligent analysis on collected data. This means the system should provide a wealth of analytical algorithms that can infer useful information or discover predictive patterns from the data. Armed with the information, early interventions can be made to prevent the medical event from happening. Unfortunately, most current remote monitoring systems lack such advanced analytics capabilities.

In this paper, we present the design and implementation of WANDA, a remote health monitoring system for HF patients. The system is an end-to-end solution that covers all key aspects of remote health monitoring, from data collection, data storage and access, to data analytics. It also features several auxiliary tools such as an administrative portal, a questionnaire system, and a social network component. The system has been deployed in three heart failure clinical trials over the past 4 years.

We also showcase the analytics engine of WANDA by processing real clinical data and generating statistical models. The data for our experimental analysis is based on a pilot trial of 18 heart failure patients. Our analysis shows that the prediction models significantly improve the accuracy of predicting the worsening of HF patients' symptoms when compared to the traditional thresholding algorithm based on daily weight changes.

2. Related Work

The use of technology to monitor patients at a distance is increasingly gaining attention as a strategy to improve the care of patients with chronic diseases. Remote patient monitoring systems provide the opportunity for a better follow-up, early detection of signs of clinical deterioration, and early intervention to prevent hospitalization or even death. Several studies targeting the remote monitoring of heart failure patients have been conducted in the past 5 years.

Chaudhry's telemonitoring study [4] required participants to make daily phone calls to an automated telemonitoring system (provided by Pharos Innovations [5]) for a period of 6 months. Each call played a prerecorded voice message that consisted of a series of questions about symptoms and weight for which the participants had to provide answers using the keypad on the phone. The responses were then downloaded from the telemonitoring system to an Internet website for daily review by clinicians. However, compared to standard care, this study proved to be unsuccessful in reducing the risk of readmission or death in heart failure patients.

Another heart failure study conducted by Soran [6, 7] included an electronic scale and an individualized symptom response system (Alere DayLink monitor [8]) connected to a computer database via a standard phone line. Patients were instructed to weigh themselves and answer a series of heart failure questions daily. Nurses reviewed the transmitted data on a daily basis and immediately contacted patients whenever the data fell out of a healthy range. After contacting the patient, the nurses immediately notified the patient's primary physician of any symptom changes by means of a fax report. Again, this study showed no improvement in clinical outcomes when compared to standard care of heart failure.

The Home or Hospital in Heart failure (HHH) study is another study that was conducted to evaluate a home telemonitoring system to supervise heart failure patients outside the hospital setting [9]. In this study, patients took weekly measurements of their weight, heart rate and blood pressure. Measurements were then entered via a phone keyboard, in reply to questions from a computerized interactive voice response system. Electrodes were attached to the patients' bodies to allow for a 24-hour cardiorespiratory monitoring. Since the raw ECG signal could not easily be transferred through the standard telephone lines, only the RR time series were actually transmitted to a database and then to an analysis center, together with the other measurements. Despite high patient compliance (81% of vital signs transmissions and 92% of cardiorespiratory recordings were completed) results showed no significant effect of remote health monitoring in reducing bed-days occupancy, cardiac death or hospitalization when compared to usual care of HF.

According to Desai [10], an effective home monitoring system must contain the necessary elements that together complete the circle of heart failure management. Some of the important circle elements are the reliable measurement of physiological variables that can help in the early detection of adverse events, efficient transmission of data to make possible a timely response, the direct reception of data by personnel qualified to recommend an effective intervention, and patient adherence.

Compared with routine care, the telemonitoring solutions used in Chaudhry's, Soran's and the HHH studies show little improvements in readmission rates or death. This can be explained by several breaks in this circle. In Chaudhry's and Sorans' studies, the lack of patient adherence as well as the requirement for decision making by midlevel providers to be supervised by physicians may stall timely responses to detected early signs, especially in large-scale studies. More importantly, however, weight and symptoms alone do not suffice as accurate and responsive measures for the early prediction of heart failure. Rapid weight gain is a relatively specific predictor of heart failure decompensation. However, it is not a very sensitive marker because weight can vary with changes in caloric intake, and the quantity of weight gain before hospitalization is typically fairly modest, with fewer than 50% of patients gaining more than 2 lb (0.9 kg). Therefore, using weight as the only measure is inadequate to recognize impending decompensation in sufficient time to intervene to prevent hospitalization [11]. These reasons, among other things, disrupt Desai's circle of HF management, preventing even an effective intervention from improving outcomes in practice. Although the HHH study measures three different markers (weight, heart rate and blood pressure) as well as continuous cardiovascular signals, it severely violates the efficient data transmission and timely response element of the circle. Because the measurements are only taken and transmitted once every week over a slow telephone line (as is the case in [4] and [6]), rich raw ECG data can not be transmitted and the granularity of intervention becomes very coarse for effective outcomes to be observed.

It is therefore necessary to design a remote monitoring system that can effectively close or at least tighten the circle of management in practice. This includes encouraging patient compliance, offering a reliable measurement of accurate physiological variables that make early detection possible, and transmitting data efficiently to ensure a timely response. WANDA is an automated real-time scalable remote health monitoring system developed with these results and findings in mind and designed to address all of these requirements.

Today many existing remote monitoring system for heart failure patients suffer the following two shortcomings. Some are invasive, requiring trained technicians to attach sensors to a patient's body [12] or implanted inside of it [13]. Other systems are reactive in nature [14] where patients are contacted only when their symptoms are already exacerbated or they are not feeling well. WANDA, in contrast, is a noninvasive, easy-to-use, proactive remote monitoring system for heart failure patients that can detect early vital signs and predict adverse events using a set of powerful algorithms.

3. System Architecture

WANDA is a three-tier, end-to-end remote monitoring system with extensive hardware and software components designed to cover the broad spectrum of the telehealth and remote monitoring paradigm. The overall architecture is summarized in Figure 1.

The first tier of the architecture consists of a data collection framework, which is formed from a heterogeneous set of sensing devices that measure various bodily statistics such as weight, body fat, body water, blood pressure, heart rate, blood glucose, blood oxygen saturation and body movements. The data from these sensors are collected, processed, and transmitted via a smartphone-based gateway to the cloud—the second tier of the WANDA architecture. The large amount of data are stored and indexed using a scalable database and are easily accessible using a RESTfulstyle web interface [15]. The last tier of the WANDA architecture is a backend analytics engine capable of continuously generating statistical models and predicting outcomes using various machine learning and data mining algorithms. In the following sections, we describe each component of the system briefly; however, the main focus of this paper is on the analytics engine and the new features that it provides for healthcare management.

3.1 Data Collection

Due to the increasingly ubiquitous nature of smartphones and their portability and connectivity, we decided to utilize an Android smartphone as a central hub for receiving patient

measurements and communicating them to the WANDA server. All the devices we utilized are Bluetoothenabled. Users can readily collect data on their WANDA smartphone application (WANDA App) which provides a user-friendly data management tool for the patient, while transmitting data to a central WANDA server for storage and backend data analytics.
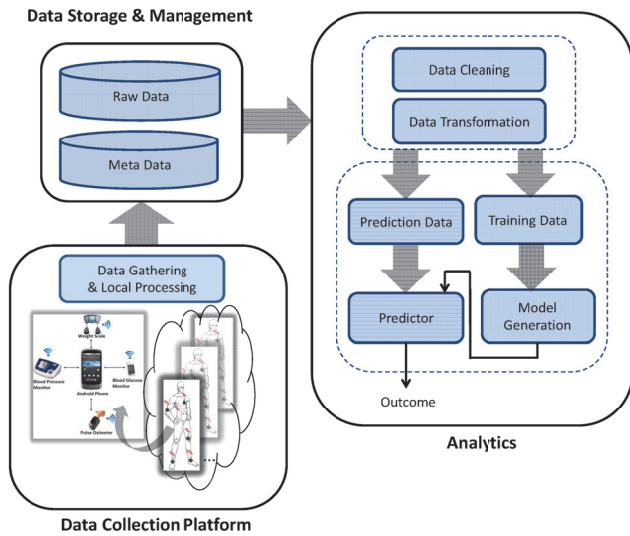

Figure 1. Overall architecture of WANDA

The WANDA App was designed to run on any device running Android OS version 2.0 or above and is compatible with multiple sensing devices. Before the phone can establish a connection with any device, a "pairing" process is required. When two Bluetooth nodes are paired, they are aware of each other's existence, have a shared-link key that can be used for authentication, and are capable of establishing an encrypted connection. Once two devices are paired, the basic information about the devices such as device name, class, and MAC address is saved and no more pairing is needed in the future. For security reasons, we require users to specify the sensors that they intend to use and the phone only accepts the connection request from these devices. Depending on the device, the phone can act as either slave or master. The device data are received by the Bluetooth service module before they are processed by a measurement module. The measurement module then stores the information through a record module. The history module provides a data management tool for users to analyze their own data, and the services module performs activity recognition and other sensor services in a background thread. All the data are then displayed to the user through a graphical user-interface (GUI) module that interacts with the user.

Since the majority of HF patients are older than 65 years, the WANDA smartphone application was designed to be a user-friendly and intuitive. In the main page, the user has the choice to record a new measurement or view a history of the measurements, as illustrated by Figure 2(a). If the user selects the option to record a new measurement, a record page is displayed with five options. The options the user can choose from are the following: weight, blood pressure, blood glucose, blood oxygen saturation, voice, and questionnaire (Figure 2(b)). If a medical device is selected, a short pre-measurement instruction page is displayed for the user as shown in Figure 2(c). When ready to take a measurement, the

user selects the "Start Measurement" button to establish the Bluetooth connection with the selected device. After the measurement is complete, the phone is designed to receive a complete measurement packet from the Bluetooth medical device and displays the measurement values to the user, as depicted in Figure 2(d). The system immediately returns to the record page allowing the user to record another measurement. In the main screen of the smartphone application, users can select "History Data," to view their device-specific history of measurements. A similar page to Figure 2(b) is displayed for the user to select the intended measurement device. After a device is selected, the user can select the time period of interest as illustrated in Figure 2(e). The device-specific history of data measurements is then displayed in a graphical format for the user to view, as shown by Figure 2(f).
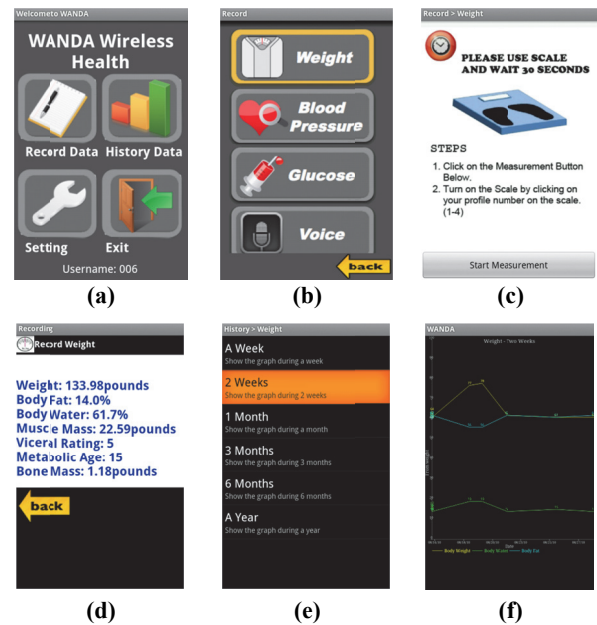

Figure 2. WANDA Android App screen shots

3.2 Storage and Access

WANDA takes advantage of Amazon's highly scalable cloud data storage to handle its growing storage needs. Amazon S3 provides data security through multiple access control mechanisms and encryption for both secure transit and secure storage on disk. Data durability and reliability are guaranteed by Amazon's redundant storage and regular verification and repair of corrupted stored data. Security and reliability are by far the most important considerations for a data storage and management system used for storing confidential medical patient information.

We also use Amazon's SimpleDB to store structured data such as patient and sensor information, data indices, adverse events, and data annotations. SimpleDB is a NoSQL database management system as opposed to the traditional relational database. It offers superior scalability and an extremely flexible database schema. The flexible schema is particularly important as it allows data stored in WANDA to be both backward and forward compatible without undergoing painful schema migrations.

However, the NoSQL nature of SimpleDB presents several technical challenges when it comes to data access. For example, SimpleDB lacks many features commonly found in a relational

database such as the JOIN operation and full ACID guarantee. This makes it complicated to develop software that interacts with the database directly. As a result, we have developed an intermediate layer that provides a web-based RESTful interface for the developers. Not only does this interface free the developers from learning the specifics of SimpleDB, it also gives a pseudo object-oriented structure to the database. Table I gives a few examples of RESTful resources with available methods and expected results.

The WANDA database also features a Publisher-Subscriber (PubSub) interface for applications that need to be notified of changes to the database in real-time. For example, a simple safety system may wish to check every blood pressure reading and alert the medical staff when the reading exceeds or falls below a certain value. Instead of polling the database at a regular interval, the system can subscribe to the blood pressure topic published by the database. It then receives a notification via HTTP requests or email whenever a new reading is added to the database. The database publishes a wide range of topics on addition, editing, and removal of objects at both fine- or coarse-grained levels, e.g. for a particular patient or for all patients from a particular study.

Table I. Example WANDA RESTful resources

Resource	Methods	Result
/measurements	GET	Return a list of latest measurements
/subjects/add	PUT	Add a new subject
/sensors/<Sensor ID>	GET PUT DELETE	Return info of a specific sensor Update info of a specific sensor Remove a specific sensor

3.3 Data Analytics
One of the strengths of WANDA over other remote health monitoring systems is its analytics engine. Based on the data and annotations collected, the analytics engine can generate multiple statistical models using various machine learning and data mining algorithms, including classification, clustering, association rule mining, etc. These models can then be used for both diagnostic and prognostic purposes. For example, in the case of HF patients, it is highly desirable to be able to predict the worsening of symptoms before the patient is actually hospitalized.

The analytics process normally consists of two stages. In the offline stage, the data are downloaded and analyzed offline based on various hypotheses. Once a strong model has been generated and validated, it can then be uploaded to the server to perform online prediction in the second stage. One of the challenges here is to optimize the algorithm so that it can be executed in a real-time fashion. This often requires using the PubSub data access interface, in conjunction with a temporal storage for pre-computed historic data. Finally, when the algorithm detects a pattern that is strongly associated with an undesirable outcome, an alarm is triggered and the personnel in charge are contacted immediately.

The analytics engine integrates nicely with the Weka framework [16]. This allows the engine to support a wealth of commonly used machine learning and data mining algorithms. Furthermore, an API is also provided for adding new algorithms to the engine.

3.4 Supporting Tools
We have also developed a number of tools that work closely with the core WANDA system to support the need of various clinical trials and studies. These tools include an administrative portal, a questionnaire system, and social network applications.

3.4.1 Administrative Portal
A Web-based tool serving as a study and information management platform was developed in-house. The tool is a web application that facilitates labeling, annotating and entering data into the database. Medical clinicians and nurses can easily enter information including patient demographics, like gender, age and ethnicity, as well as notes of events like hospital discharge, hospitalization or death. The data is stored in the second tier of the system described in Section 3.2. In addition to data entry and data annotation, the tool also offers a rich and customizable interface that provides advanced search functionality for stored data. The search results can also be visualized in the form of graphs and charts to display information like patient measurements over time. Good data visualization can enhance fast diagnostic decision-making based on simple trends seen in the displayed data, like a constantly increasing weight, for example. Data annotation also serves our analytics engine by providing our algorithms with more training data. The tool provides easy navigation through different panels to deliver an intuitive user experience.

3.4.2 Questionnaire System
Another supporting tool for WANDA is the questionnaire system. Questions about symptoms are entered by clinicians into the administrative portal described in Section 3.4.1. These questions are then shown on the patient's Android phone at a predefined frequency. The patients can answer the heart failure symptom questions using the phone's keyboard or touch screen. The user responses are then uploaded to the WANDA storage platform described in Section 3.2. Medical staff can then access these responses using the administrative portal.

3.4.3 Social Network Applications
WANDA Social is WANDA's social component which can be a powerful tool for encouraging healthy behavior as well as improving patient compliance. A Facebook application was developed in-house to offer two motivating factors: cooperation and competition. By connecting with people who have similar conditions and health concerns, users can work together and encourage each other to achieve a common goal like losing weight, for instance. The application also provides competitive events such as online tournaments, where users can challenge each other to accomplish certain health-related tasks like lowering their blood pressure to a healthy level within a certain time frame. A leaderboard and nominal prizes can motivate users to achieve the greatest improvement in a given health indicator. The WANDA Social platform was fully integrated with Facebook to take advantage of its large user base. To respect the user's privacy, the application will allow authenticated and registered participants of the study to post their information, e.g. sensor readings and achievements, using their real identity or as anonymous users. By tapping into cooperative and competitive relationships through WANDA Social, we can motivate patients to adhere to a treatment regimen or adopt a healthier lifestyle.

4. Clinical Trials
Since November 2009, the WANDA system has been used for health data collection on the intervention arm including 26 different congestive heart failure patients [17]. The population of this first clinical trial is approximately 68% male; 40% White, 13% Black, 32% Latino, and 15% Asian/Pacific Islander, with a mean age of approximately 68.7 ± 12.1. The gender distribution

and anticipated age of participants are representative of the incidence and natural history of congestive heart failure. Study participants were all provided with Bluetooth-enabled weight scales, blood pressure monitors, and personal activity monitor devices.

The second clinical trial of WANDA started in February 2011 with 18 lowliteracy Latinos with heart failure. This population is disproportionately affected with HF, is more likely to be hospitalized with HF, and is at greatest risk for re-hospitalization and death from HF. The population of the participants in this study was approximately 89% male with a mean age of approximately 54. Study participants were provided with Bluetooth-enabled weight scales, blood pressure monitors, landline or Ethernet gateways, and Android activity monitoring applications.

Finally, the system is currently deployed in a much larger study that targets the remote monitoring of 1500 patients of 50 years or older with heart failure problems. This on-going project is conducted in collaboration with the UCLA Department of Medicine, UC-Davis, UCSF, UCI, UCSD, and Cedar Sinai Hospital. Heart failure patients who were hospitalized at any of the six participating medical centers are being considered to be recruited in either the control or intervention arm through a randomized trial process. Data collection for this study has started in November 2011 and is scheduled to complete in April 2013. Patients measure their weights and blood pressures and reply to questionnaires on a daily basis, and the collected data are transmitted to the database via a phone-line, an Internet connection at home, or through cellular networks.

5. Symptom Prediction

In order to demonstrate the capability of the WANDA analytics engine, we use the dataset gathered from the second clinical trial described in Section 4 to predict the worsening of HF symptoms using advanced machine learning algorithms.

So far, the most widely used metric to predict the worsening of HF symptoms is Daily Weight Change (DWC). It is even recommended by American College of Cardiology/American Health Association as a potential indicator for water retention, which leads to swollen ankles and other HF symptoms [18]. However, several studies have shown that DWC has relatively weak correlation with the worsening of HF symptoms [19-21]. We have also experienced the poor predictive power of DWC first-hand during our clinical trials. When the medical staff follow up observing a DWC of more than 2 lbs , the patient almost always attributes the change to food or normal weight fluctuation and denies any worsening of HF symptoms.

The case of using DWC as a predictor is further weakened from the example depicted in Figure 3. The figure shows systolic and diastolic blood pressure (SBP & DBP) and weight readings from the same patient over two separate 7-day periods. By the 7th day in Figure 3(a), the patient is known to have reported worsening of HF symptoms, whereas the patient reported no change in symptoms during the 7-day period in Figure 3(b). Although the weight has fluctuated in both cases, there is very limited change in weight (< 1lb) measured on the 7th day. If DWC were used to predict the symptom, we would most probably end up having either a large number of false positives, or miss the true positives altogether. On the other hand, it is clear from the figure that the systolic and diastolic blood pressures are less stable before the worsening of HF symptoms, suggesting that they may serve as better predictors in this case.

5.1 Experiment Setup

WANDA's second HF clinical trial is used as the data source for the experiments. We were particularly interested in patients' daily weight, systolic and diastolic blood pressure, and heart rate measurements as they are relatively free of missing data for most patients over the three-month period.

Patients also answer a series of 9 questions, as listed in Table II, on a daily basis. We interpret a positive response to question 8 and 9 as a worsening of a patient's HF symptoms, whereas a negative or a lack of response is interpreted as a stabilized HF condition. The answers are also crosschecked with nurses' call logs to remove any accidental false reporting. Based on these criteria there were a total of 34 instances of worsening of HF symptoms self-reported by 9 patients during the clinical trials.

Once instances of worsening HF symptoms have been identified, they are labeled as positive. All other instances that are at least 3 days away from the positive instances are considered negative. Instances that occurred within 3 days of positive instances are discarded. A total of 9 features, listed in Table III, are extracted at each instance and used by the prediction algorithms.
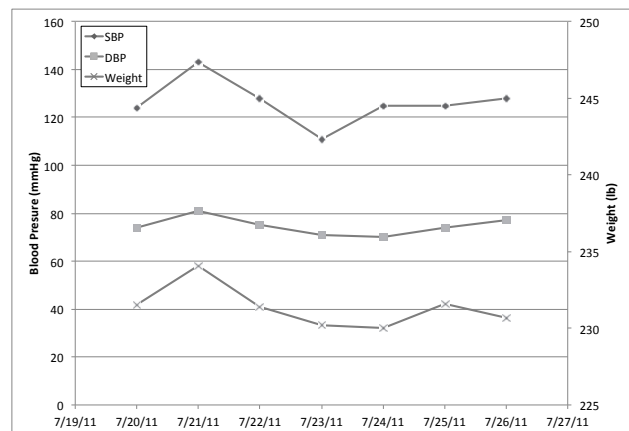
(a)
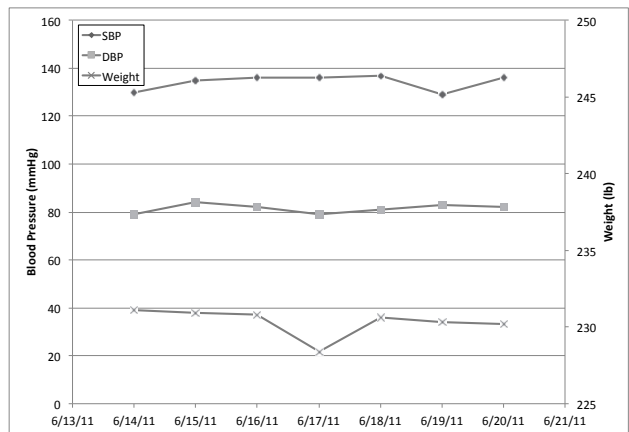
(b)

Figure 3. 7-day blood pressure and weight readings of a patient when (a) worsening of HF symptom is reported, and (b) no change in HF symptom is reported

Six different machine learning algorithms from the WANDA analytics engine have been evaluated in the experiments. They are listed in Table IV with a short description and the corresponding

parameters used for each algorithm. These algorithms are chosen as they utilize a wide range of fundamentally different statistical principles. This ensures that the results are based on the intrinsic discriminatory power of the extracted features rather than a particular algorithm overfitting the model. The same set of experiments is also conducted using simple thresholding based on a DWC of 2 lbs.

Several measures have been taken to ensure the fairness of the experiments. Firstly, equal numbers of positive and negative instances are used in each experiment to prevent biasing toward a particular class. Secondly, the results are computed using ten-fold cross validation to guard against overfitting given the relatively small number of instances. Finally, each experiment is conducted ten times and the average results are reported, thus reducing the effect of outliers.

Table II. Daily Questionnaire

#	Question
1	Are you coughing more than usual?
2	Are you more tired than usual?
3	Did you use an extra pillow last night?
4	Did you eat a low-salt diet?
5	Did you take your heart medications?
6	Did you take your water pill this morning?
7	Did you walk or exercise today?
8	Have you noticed increased difficulty in breathing?
9	Have you noticed increase in swelling of feet or ankles?

Table III. Classification Features

Feature	Description
dcs	Daily change in systolic blood pressure
dcd	Daily change in diastolic blood pressure
dcw	Daily change in weight
$sds.3d$	Standard deviation of systolic blood pressure over the past 3 days
$sdd.3d$	Standard deviation of diastolic blood pressure over the past 3 days
$sdw.3d$	Standard deviation of weight over the past 3 days
$sds.7d$	Standard deviation of systolic blood pressure over the past 7 days
$sdd.7d$	Standard deviation of diastolic blood pressure over the past 7 days
$sdw.7d$	Standard deviation of weight over the past 7 days

5.2 Results and Discussion

The algorithms are evaluated in terms of their accuracy, sensitivity, and specificity. They are computed based on the number of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) in the prediction results,

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$

$$Sensitivity = \frac{TP}{TP+FN} \qquad Specificity = \frac{TN}{TN+FP}$$

The results for each algorithm are plotted in Figure 4. Since the number of positive and negative instances is balanced, the minimal accuracy of any better-than-chance algorithm should be greater than 0.5. The figure shows that predicting the worsening of HF symptoms using DWC seems to be only marginally more accurate than guessing, with an accuracy of 0.519. This is not surprising considering that only 15.9% of the positive instances have been classified as such, whereas a substantial amount of negative instances (28.6%) turn out to have a DWC of more than 2 lbs.

On the other hand, the algorithms from the WANDA analytics engine attain much higher accuracy. Three algorithms, NBC, LR, and RIDOR, are able to correctly predict the worsening or stabilization of HF symptoms 74% of the time. Even though VFI has the lowest accuracy of 0.696, it is still nearly 20% more accurate than DWC. Furthermore, all six algorithms have sensitivity values that are at least 45% greater than that of DWC. This suggests that the features listed in Table III have a much stronger correlation to the worsening of HF symptoms than simple daily weight change, which can be further confirmed by the high specificity ranged from 0.696 (for kNN) to 0.87 (for RIDOR).

However, not all the features are equally predictive. In fact, we already know that dcw has very limited predictive power based on the results for DWC. One way to rank the features is based on their relationships with the class labels, such as information gain, relevance, correlation etc. Many machine learning algorithms, including the one used in our experiment, make use of these relationships directly when constructing their statistical models by weighting the features differently. Out of the 9 features, we discover that the algorithms consistently identify $sds.3d$, $sdd.3d$, $sds.7d$, and $sdd.7d$ as the most predictive ones. Actually, C4.5, RIDOR, and LR even ignore the other features entirely when constructing the models.

Figure 5 shows the Receiver Operating Characteristic (ROC) curves for the WANDA analytics engine algorithms and DWC. These curves demonstrate the effect of varying the classification algorithm's discrimination threshold on the true positive rate (TPR) vs. the false positive rate (FPR), which are calculated as

$$TPR = \frac{TP}{TP+FN} \qquad FPR = \frac{FP}{TN+FP}$$

A perfect algorithm will produce a point at the top left corner of the figure where TPR = 1 and FPR = 0. On the other hand, an algorithm that guesses randomly will produce a curve that follows the dotted 45-degree line in the figure where TPR = FPR. Therefore, the area under the curve (AUC) is often used as a metric to gauge the performance of a classification algorithm.

It is clear from the figure that DWC is not significantly better than random guessing. In fact, the AUC for DWC is estimated to be 0.573, which is less than 8% higher than that of a completely random guess. In comparison, the other algorithms from the WANDA analytics engine perform significantly better. The AUC ranged from 0.817 for LR to 0.621 for VFI. It is worth noting that NBC can be tuned to achieve a decent TPR of 0.652 while minimizing FPR to 0 at the same time. This makes NBC the most
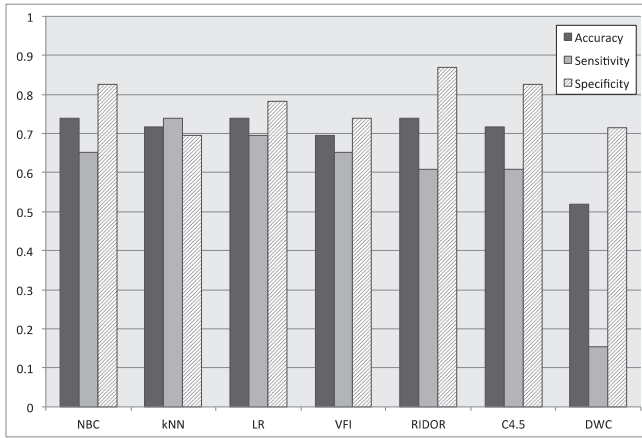
Figure 4. Accuracy, sensitivity, and specificity of NBC, 5NN, LR, VFI, RIDO, and C4.5 vs. DWC

Figure 5. ROC curve for NBC, kNN, LR, VFI, RIDOR, and C4.5 vs. DWC

suitable algorithm to use when the consequence of a false alarm is more costly than that of a missed prediction.

5.3 Error Bound Analysis

It is important to find out the lower bound of the classification error in order to put the accuracy of an algorithm into perspective. After all, a prediction algorithm is only as good as the features given to it. The lowest error rate achievable by any binary classifier is bounded by the Bayes Error Rate,

$$P(error) = \int_{R_2} p(x|w_1)P(w_1)dx + \int_{R_1} p(x|w_2)P(w_2)dx$$

where R_1 and R_2 are the regions where x is misclassified given class distributions w_1 and w_2. Unfortunately, the distributions are unknown in reality and can therefore only be estimated. Using the proof in [22], we can estimate the Bayes Error Rate as half of the error rate of 1-NN, which can be obtained empirically from the data.

As a result, the Bayes Error Rate is estimated to be 0.174, which implies a theoretical optimal accuracy of 0.826, which is only about 9% higher than the best accuracy achieved by the WANDA analytics engine. While this suggests that the algorithms can be better tuned, the improvement may come at the cost of reduced sensitivity or specificity. It is thus more fruitful to find features that have stronger predictive power.

6. Conclusion

In this paper we have introduced the design and implementation of a remote health monitoring system called WANDA. The system is an end-to-end solution comprising three parts: a) a smartphone-based gateway that wirelessly collects data from various sensor-based devices, b) a scalable cloud-based database with a RESTful web interface and PubSub system, and c) an analytics engine capable of prognostic prediction using machine learning and data mining algorithms.

Using real data collected from a pilot clinical trial, we have demonstrated the strength of WANDA's analytics engine by accurately predicting the worsening of HF symptoms in patients. Compared to the commonly accepted predictor of daily weight changes, the algorithms of WANDA analytics engine have identified the 3-day and 7-day fluctuation of blood pressure readings to be highly correlated to the worsening of HF symptoms. As a result, the WANDA analytics engine is able to build prediction models that are up to 74% accurate, which is
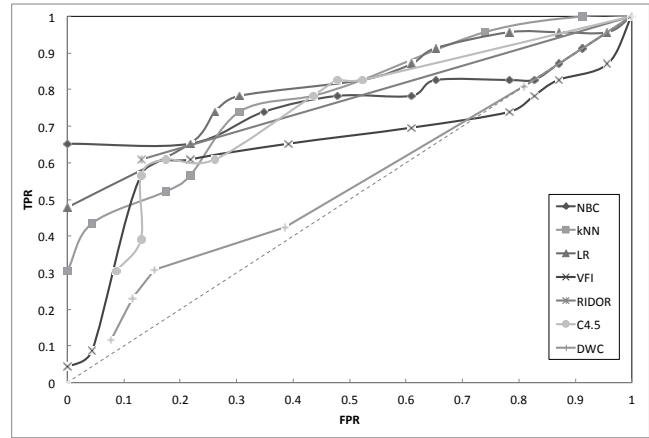
more than 20% higher than using daily weight change alone. Furthermore, the sensitivity is improved by at least 45% when using these models. Finally, we have shown that the accuracy of the WANDA analytics engine is only 9% lower than the estimated upper bound.

The prediction approach presented in this paper is only an example of intelligent analytics that our remote monitoring system offers. In future work, we plan to expand the WANDA analytics engine into several areas beyond predictions, including task optimization, unsupervised learning, individualized health guidance for patients, context-aware prediction, and long-term pattern discovery.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] D. Lloyd-Jones, R. J. Adams, T. M. Brown, M. Carnethon, S. Dai, G. De Simone, T. B. Ferguson, E. Ford, K. Furie, and C. Gillespie, "Executive summary: heart disease and stroke statistics--2010 update: a report from the American Heart Association," *Circulation,* vol. 121, p. 948, 2010.

[2] J. S. Ross, J. Chen, Z. Lin, H. Bueno, J. P. Curtis, P. S. Keenan, S. L. T. Normand, G. Schreiner, J. A. Spertus, and M. T. Vidán, "Recent national trends in readmission rates after heart failure hospitalization," *Circulation: Heart Failure,* vol. 3, pp. 97-103, 2010.

[3] S. F. Jencks, M. V. Williams, and E. A. Coleman, "Rehospitalizations among patients in the Medicare fee-for-service program," *New England Journal of Medicine,* vol. 360, pp. 1418-1428, 2009.

[4] S. I. Chaudhry, B. Barton, J. Mattera, J. Spertus, and H. M. Krumholz, "Randomized trial of telemonitoring to improve heart failure outcomes (Tele-HF): study design," *Journal of cardiac failure,* vol. 13, pp. 709-714, 2007.

[5] (2011).*Pharos-Innovations*.Available: http://www.pharosinnovations.com

[6] O. Z. Soran, A. M. Feldman, I. L. Piña, G. A. Lamas, S. F. Kelsey, F. Selzer, J. Pilotte, and J. R. Lave, "Cost of medical services in older patients with heart failure: those receiving enhanced monitoring using a computer-based telephonic monitoring system compared with those in usual care: the heart failure home care trial," *Journal of cardiac failure,* vol. 16, pp. 859-866, 2010.

[7] O. Z. Soran, I. L. Piña, G. A. Lamas, S. F. Kelsey, F. Selzer, J. Pilotte, J. R. Lave, and A. M. Feldman, "A randomized clinical trial of the clinical effects of enhanced heart failure monitoring using a computer-based telephonic monitoring system in older minorities and women," *Journal of cardiac failure,* vol. 14, pp. 711-717, 2008.

[8] (2011). *Alere.* Available: http://www.alere.com

[9] A. Mortara, G. D. Pinna, P. Johnson, R. Maestri, S. Capomolla, M. T. La Rovere, P. Ponikowski, L. Tavazzi, and P. Sleight, "Home telemonitoring in heart failure patients: the HHH study (Home or Hospital in Heart Failure)," *European journal of heart failure,* vol. 11, pp. 312-318, 2009.

[10] A. S. Desai and L. W. Stevenson, "Connecting the circle from home to heart-failure disease management," *New England Journal of Medicine,* vol. 363, pp. 2364-2367, 2010.

[11] A. S. Desai, "Home Monitoring Heart Failure Care Does Not Improve Patient Outcomes: Looking Beyond Telephone-Based Disease Management," *Circulation,* vol. 125, pp. 828-836, 2012.

[12] (2011).*Corventis-AVIVO.*Available: http://www.corventis.com/us/avivo.asp

[13] (2012). *Medtronic. CareLink.* Available: http://www.medtronic.com

[14] (2011). *IdealLife.* Available: http://www.ideallifeonline.com

[15] R. Fielding, "Representational state transfer (REST)," *Architectural Styles and the Design of Network-based Software Architectures. University of California, Irvine,* p. 120, 2000.

[16] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD Explorations Newsletter,* vol. 11, pp. 10-18, 2009.

[17] M. Suh, C. A. Chen, J. Woodbridge, M. K. Tu, J. I. Kim, A. Nahapetian, L. S. Evangelista, and M. Sarrafzadeh, "A Remote Patient Monitoring System for Congestive Heart Failure," *Journal of medical systems,* pp. 1-15, 2011.

[18] S. A. Hunt, D. W. Baker, M. H. Chin, M. P. Cinquegrani, A. M. Feldman, G. S. Francis, T. G. Ganiats, S. Goldstein, G. Gregoratos, and M. L. Jessup, "ACC/AHA guidelines for the evaluation and management of chronic heart failure in the adult: executive summary: a report of the American College of Cardiology/American Heart Association task force on practice guidelines," *Journal of the American College of Cardiology,* vol. 38, p. 2101, 2001.

[19] 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, pp. 1549-1554, 2007.

[20] J. Lewin, M. Ledwidge, C. O'Loughlin, C. McNally, and K. McDonald, "Clinical deterioration in established heart failure: What is the value of BNP and weight gain in aiding diagnosis?," *European journal of heart failure,* vol. 7, pp. 953-957, 2005.

[21] J. Zhang, K. M. Goode, P. E. Cuddihy, and J. G. F. Cleland, "Predicting hospitalization due to worsening heart failure using daily weight measurement: analysis of the Trans-European Network-Home-Care Management System (TEN-HMS) study," *European journal of heart failure,* vol. 11, pp. 420-427, 2009.

[22] T. Cover and P. Hart, "Nearest neighbor pattern classification," *Information Theory, IEEE Transactions on,* vol. 13, pp. 21-27, 1967.

Table IV. Classification Algorithms Used in Experiment

Algorithm	Description	Parameters
Naïve Bayes Classifier (NBC)	NBC uses Bayes' theorem and assumes each feature is conditionally independent. The posterior probability of a class C given a set of features F1, F2, … Fn is therefore $$p(C|F_1, F_2, ..., F_n) \propto p(C)\prod_{i=1}^{n} p(F_i|C)$$ The prior and likelihood can be calculated from the training set directly and applies to the test set.	
Nearest Neighbor (kNN)	An unlabeled instance is classified based on its nearest k neighbors from the training set based on majority vote. Euclidean distance of the features is often used to determine the closeness of two instances $$d = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$ where x_i and y_i are corresponding features from the two instances.	k = 5
Logistic Regression (LR)	Logistic regression is a type of regression that predicts the outcome of a binary depend variable (class) based on a set of independent variables (features) X_i using the logistic function $$p(x) = \frac{1}{1+e^{-f(x)}} \ , \ f(x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_i X_i$$ where $p(x)$ is the probability of x being class 1, given regression coefficients $\beta_1 … \beta_i$ and intercept β_0.	
Voting Feature Interval (VFI)	VFI classifies builds upper and lower bounds around each class for each feature. Classification is based on majority voting, where a vote for class C based from feature *a* is computed as $$v(a,C) = \frac{\text{interval\_class\_count}(a,i,C)}{\text{class\_count(C)}}\left(\frac{H(C|a)}{\text{max uncertainly}}\right)^{\text{bias}}$$	bias = 0.6
Ripple-Down Rule Learner (RIDOR)	RIDOR is a version of Ripple Down Rule using the Indcut algorithm where the default rules are first generated based on least error rate. A set of exception rules are generated to predict classes other than those covered by the default rules.	
C4.5 Decision Tree (C4.5)	C4.5 constructs a decision tree based on information entropy. Each feature in the training set is evaluated for its information gain $$IG(T,a) = H(T) - H(T|a) \qquad H(T) = -\sum_{i=1}^{n} p(x_i)\ln p(x_i)$$ Feature with the largest *IG* and have yet been used is chosen to split the decision tree at each node until the confidence falls below certain threshold.	confidence = 0.25