Visualization of Health Monitoring Data acquired from Distributed Sensors for Multiple Patients

Alex Page*, Tolga Soyata*, Jean-Philippe Couderc*†, Mehmet Aktas†, Burak Kantarci‡, Silvana Andreescu§

*Dept. of ECE
University of Rochester
Rochester, NY 14627
{apage4,soyata}@ece.rochester.edu

†UR Medical Center
University of Rochester
Rochester, NY 14627
{aktas,jip}@urmc.rochester.edu

‡Dept. of ECE
Clarkson University
Postdam, NY 13699
bkantarc@clarkson.edu

§Dept Chem & BioM Sci
Clarkson University
Postdam, NY 13699
eandrees@clarkson.edu

***Abstract*—As global healthcare systems transition into the digital era, remote patient health monitoring will be widespread through the use of inexpensive monitoring devices, such as ECG patches, glucose monitors, etc. Once a sensor-concentrator-cloudlet-cloud infrastructure is in place, it is not unrealistic to imagine a scenario where a physician monitors 20–30 patients remotely. Such an infrastructure will revolutionize clinical diagnostics and preventative medicine by allowing the doctors to access long-term and real-time information, which cannot be obtained from short-term in-hospital ECG recordings.**

While the large amount of sensor data available to a physician is incredibly valuable clinically, it is overwhelming in raw form. In this paper, the *data handling* aspect of such a long term health monitoring system is studied. Novel ways to record, aggregate, and visualize this flood of sensory data in an intuitive manner are introduced which allow a doctor to review days worth of data in a matter of seconds. This system is one of the first attempts to provide a tool that allows the visualization of long-term monitoring data acquired from multiple sensors.

I. Introduction

Physicians assess a patient's risk of cardiovascular diseases (CVD) by referring to his/her clinical history and performing highly observational and relatively insensitive physical examinations and electrocardiograms (ECG) [1], [2]. The pathology of CVD starts at earlier stages than it is observable by such conventional methodologies and there are no clinical tests that can detect the onset and progression of CVD. A highly-automated remote (in-home) health monitoring of clinically-relevant cardiac biomarkers could provide invaluable diagnostic information. Eliminating the need to administer such tests at the HCO could translate to substantial cost savings.

Currently, there are no suitable methods to assess and predict the risk of CVD and chronic heart failure in real time to enable effective therapeutic intervention [3]. Mechanisms that are involved in the development of CVD are complex and involve a variety of interrelated processes including changes in blood cholesterol, lipid metabolism, inflammation and oxidative stress [4], [5]. Therefore, a comprehensive monitoring system is required to effectively help clinical diagnostics.

The proposed system in Figure 1 will enable physicians to monitor patients and have automatic alarm providing feedback on patient long-term health status. This monitoring can be continuous in patients with high risk for life-threatening events, or periodic with a recording frequency depending on disease

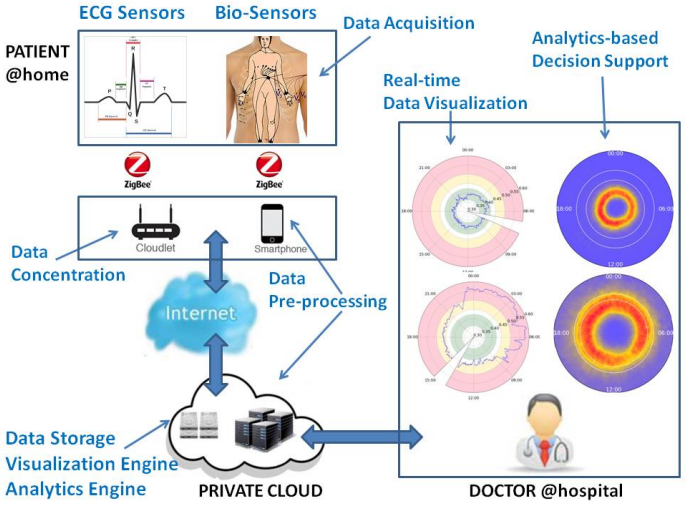

Fig. 1: Remote long term patient monitoring system including i) data acquisition, ii) data concentration, iii) data pre-processing, iv) cloud processing and analytics, v) visualization, and vi) analytics-based diagnostic assistance.

severity. This system is capable of monitoring ECG-related parameters using commercially available ECG patches [6], [7], as well as multiple other bio-markers of a patient via custom [8], [9] or off-the-shelf [10] bio-sensors in real-time. Sensory recordings of the patient will be transmitted from the patient's house (or any remote location) to the datacenter of the HCO in real-time in a secure fashion using well established encryption mechanisms [11]. Combining ECG monitoring parameters with such additional bio-markers improves the utility of the monitoring system to far beyond what is currently achievable with ECG-only monitoring [12] or single-biomarker monitoring (e.g., Glucose [10]). This technology will be disruptive because it has the potential to shift the paradigm of patient management in the US healthcare system.

Despite its potential to improve diagnostics substantially, the proposed system introduces challenges in handling a massive volume of data. Visualization of such multi-dimensional data, encompassing ECG parameters and multiple bio-markers is not straightforward. Well known century-old ECG-based visualization of a patient's cardiac operation [13] provides limited information for a short operational interval. In this paper, a

new visualization mechanism will be developed that allows the doctor to visualize ECG over $\geq$24 hours. Furthermore, novel ways to aggregate the information from multi-modal bio-sensors will be studied to improve the dimensionality of the visualization. Our contributions are as follows: 1) A system is described that automates and standardizes the pre-processing of the sensed data, which will significantly improve the diagnostic quality of the acquired data, 2) Each building block of this system is studied in detail and the challenges in constructing these blocks are identified, 3) A pilot cardio-vascular monitoring application is selected and the analysis is performed within the context of cardiovascular disease monitoring, 4) The final step of visualization is studied in great detail and a novel methodology is proposed for visualizing uni-modal or multi-modal interrelated cardiac biomarkers in real-time on a doctor's smartphone.

The remainder of this paper is organized as follows: In Section II, we provide a detailed description of the components of the system. In Section III, details about the acquired long term sensor data are provided. Based on this pre-processed and summarized data, we provide a characterization approach for well-described disease states in Section IV. In Section V, we demonstrate novel visualization approaches to allow the doctor to monitor multiple patients in real-time. We provide conclusions and pointers for future work in Section VI.

II. The System

The system depicted in Figure 1 has four major components: (1) the sensors on the patient, (2) gateway device(s) for data aggregation and transmission, (3) the datacenter, and (4) the doctor's computer running frontend application(s). In this section, we will discuss some of the important tasks that these components must perform.

A. Data Acquisition and Data Privacy

Data Acquisition: The patient wears sensors that operate from a small battery for up to a few months. The only task of these sensors is to upload their data to the concentrator through a low-power communication protocol (e.g., Zigbee), thereby eliminating the necessity of a power-hungry processor. They may cache data for a limited time if the concentrator is out of range. Connection of sensors to microprocessors, instrumentation systems and control networks has been standardized by the IEEE 1451 family, including common network functions [14], wireless communication protocols [15], transducers to Radio Frequency Identification (RFID) Systems Communication Protocols [16] and several other functions. Additionally, ISO IEEE 11073 personal health data standards family is also available to use for medical device communications, including specifications for communications through ECG monitoring [17], continuous glucose monitoring [18], cardiovascular activity monitoring [19] and several other applications. Adopting these two standards [20] can ensure interoperability through heterogeneous body sensors.

Privacy of the medical data (Layers-1-2-3): Although personally-identifiable information can be removed before communicating sensed data, *aggregate disclosure attacks* aim at deducing information through pattern recognition methods and context awareness [21], [22]. Random linear network coding along with *lightweight homomorphic encryption* is shown to be efficient to overcome malicious adversities via network analysis in multi-hop wireless networks [23], but *fully homomorphic encryption* is impractical [24], [25].

B. Data Aggregation

In the emerging Internet of Things (IoT) architectures [26], a concentrator acts as a communication gateway for the sensors and connects each sensor to the Internet [27]. These steps can be achieved in a cost efficient and scalable manner if cloud computing is integrated into the IoT architecture [28]. Remote healthcare monitoring is reported to be an application domain that can benefit from cloud-IoT integration [29]. The sensory network infrastructure in this paper departs from this vision as shown in Figure 1 by treating the bio-sensor array as a form of an IoT infrastructure, where the HCO datacenter is a *private cloud*, and the cloudlet in the patient's house is a concentrator (either the patient's smartphone, or a dedicated cloudlet [30]).

Trustworthiness of Aggregated Data (Layer-1): In our proposed system, multiple sensors are deployed in the same region and mostly in the same transmission range. This introduces resiliency issues to the sensory system where the entire sensor network can fail requiring prompt intervention. As the collected data from the sensory system is expected to be correlated with any other indicator of cardiac status, off-the-shelf heart monitoring systems [31], [32] can be integrated into the proposed sensory system, and detect anomalies in the biosensor signals through correlation analysis.

Context-aware concentration via smart devices (Layer-1): Smartphones of the patient and/or the attendants can offer ideal platforms to replace the concentrators in the Internet of Things (IoT) infrastructure as current smart phones can use both LTE and WiFi as the backhaul network. Aggregation tasks can be handled either in a local cloudlet or in the HCO's datacenter. We propose context-aware concentration of the data in the cloudlet (i.e., via WiFi connectivity) or in the HCO datacenter (i.e., via LTE connectivity). The former leads to one tenth of the latter's access delay, half the power of the latter's power consumption and ten times the latter's throughput [33], [34]. The tasks on the aggregated data can be partitioned between the cloudlet and the datacenter, however this we propose context-aware partitioning of the data between these two entities. Context will be defined as a function of the current and expected status of the patient. In order to ensure fast convergence and efficiency, the concentrator should adopt the estimator algorithms applied to learning automata [35].

C. Data Preprocessing

Data arriving at this stage has already been consolidated across sensors, and is parsable. There are a few steps to complete, which are discussed in more detail in Section III. Depending on the work required for each preprocessing stage, this task may be split up across many devices.

D. Visualization Engine

This is the component responsible for producing human-readable output from the preprocessed sensor information. The doctor may be accessing this via a frontend on a desktop, phone, or tablet, and the task distribution — to generate visulization outputs after heavy computation — may need to be adjusted accordingly. There are many challenges involved in deciding what type of image is most useful for evaluation or diagnosis of a specific patient; our approach to these challenges is extensively discussed in Section V-A.

E. Analytics Engine

While a primary goal of this system is to provide a picture of an individual's health status to the doctor, this assistance can be greatly improved by allowing comparison of the individual to various groups (such as other people of the same age or gender, taking the same medications). In order to perform these comparisons, we need to develop "norms" for different populations by utilizing existing databases (such as the THEW [36] or PhysioNet [37] databases). This process will be discussed in Section V-B.

III. Preprocessing of Sensor Data

ECG data — amplitude, in mV — is typically sampled at 200Hz or 1000Hz, at 16 bits per sample. This amounts to roughly 100MB of data per patient per day [38]. However, an ECG recorder uses more than one sensor (i.e. lead), so we must multiply by the number of leads. 3-lead ECGs are common, meaning that a hospital monitoring 1000 patients will generate 100+ GB of raw data per day. (Though, this is usually compressible down to less than 40% of its original size using a utility like `zip`. More specialized compression algorithms may be used to take advantage of the specific data structure of ECG to achieve higher compression rates).

A typical ECG waveform for one heartbeat is shown in Figure 2. Each lead of the ECG monitor will capture between about 100 and 1000 amplitude samples during a single beat. What the doctor is interested in, though, are values like the heart rate or the QT interval. There are roughly 20 such measurements we may want to store for every heart beat, such as the location of certain features in the QRS complex. These annotations are stored in supplementary files or databases that may very well occupy more space than the original recording. Accordingly, the hospital may need to plan for a couple of terabytes per day to cover their 1000 patients.

Now that we understand the form of the raw data, we will summarize the preprocessing stages that convert it to clinically-relevant information.

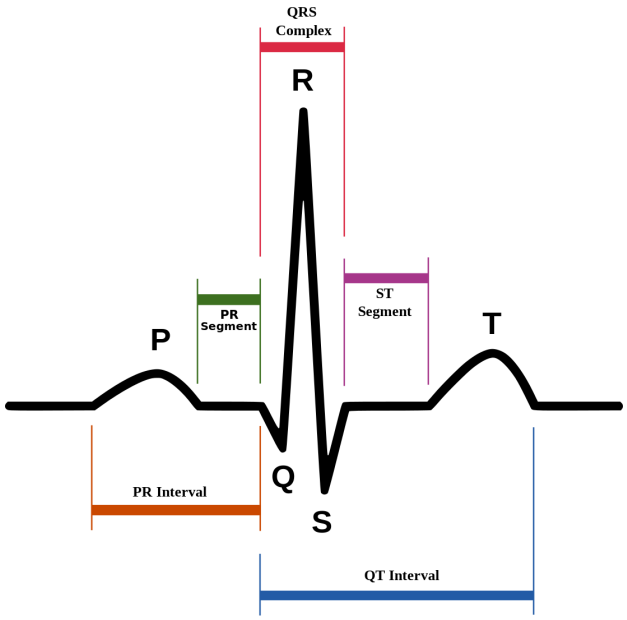

Fig. 2: ECG waveform for one heart beat.

time	lead	feature	amplitude
15:00:05.135	1	(P	0.15
15:00:05.150	2	(P	0.36
15:00:05.165	1	P	0.19
15:00:05.175	2	P	0.34
15:00:05.190	1	P)	0.15
15:00:05.195	2	P)	0.37
15:00:05.230	1	(Q	0.12
15:00:05.245	2	(Q	0.43
15:00:05.270	2	R	0.96
15:00:05.280	1	R	1.68
15:00:05.335	2	S)	0.32
15:00:05.340	1	S)	0.12
15:00:05.445	2	(T	0.57
15:00:05.540	2	T	0.69
15:00:05.560	1	(T	0.24
15:00:05.630	2	T)	0.55
15:00:05.640	1	T	0.42
15:00:05.715	1	T)	0.26
15:00:06.145	1	(P	0.04
15:00:06.170	1	P	0.12
15:00:06.170	2	(P	0.32
15:00:06.195	2	P	0.31
15:00:06.210	2	P)	0.31
15:00:06.215	1	P)	0.07
15:00:06.250	1	(Q	0.09
15:00:06.265	2	(Q	0.38
15:00:06.285	2	R	0.96
15:00:06.300	1	R	1.59
15:00:06.355	1	S)	0.06
15:00:06.355	2	S)	0.36
15:00:06.460	2	(T	0.55
15:00:06.555	2	T	0.66
15:00:06.575	1	(T	0.24
15:00:06.650	2	T)	0.53
15:00:06.660	1	T	0.33
15:00:06.735	1	T)	0.14

Fig. 3: Approximately 2 seconds of annotation data from a 2-lead ECG.

A. Stage 1: Raw Data → Primitives

The primitives will be extracted from the ECG recording, resulting in a table like that of Figure 3. This process is performed by a DSP algorithm, which, in the case of ECG analysis, is typically based around wavelet transforms. The computational complexity of this stage is an important factor in determining where it should run; as alluded to Section II-C, there is a tradeoff between running it in the datacenter vs. on a device nearer to the patient. The primitives for most interesting ECG-based values are the locations and amplitudes at the start, beginning, and end of the labeled features of Figure 2 (P, Q, R, S, T).

B. Stage 2: Primitives → Clinical Markers

From the table in Figure 3, the distance between different features can be computed. Many of these values ("segments" or "intervals" in Figure 2) are good markers of potential cardiac issues. We may also be interested in values *computed from* intervals; heart rate in beats per minute, for example, is $\frac{60}{\text{RR (sec)}}$. RR is measured from R to R in consecutive beats.

IV. Characterization of Known Disease States

Though data is being collected from many sensors, only a subset of the sensors (and measured parameters) will be relevant to a particular disease. Since every added bio-marker increases the cost and complexity of monitoring, we envision a standardized system that allows the physicin to monitor only a set of *prescribed parameters*.

A. Unimodal Monitoring

In some cases, a disease may be mainly characterized by a single parameter. In those cases, we only need to provide a single picture to the doctor, e.g. a plot of that parameter vs. time, or some statistics about it. For example, to evaluate whether a patient has Long QT Syndrome (LQTS) — explained in Section V — only two values need to be computed from ECG sensors, QT and RR, as depicted in Figure 2. If we have other sensor data for the same patient, e.g. from a blood glucose monitor, this data is not relevant to the LQTS evaluation. In order to use sensors for diagnostic purposes, then, we need to know which sensors (and which parameters from them) should be used for each specific test that the doctor is trying to perform. In the case of LQTS, plotting the QTc value (QT, corrected for heart rate) will be sufficient, since there is a fairly clear relationship for a *normal* vs. *long* QTc value. The corrected QT value is normally calculated from the QT and RR values as follows [13], where 500ms is a reasonable threshold which also depends on the patient:

$$QT_c = \frac{QT}{\sqrt[3]{RR/\text{sec}}} \implies \begin{cases} \text{Normal}: & QT_c \leq 500\,\text{ms} \\ \text{LQTS}: & QT_c > 500\,\text{ms} \end{cases} \quad (1)$$

QTc monitoring is a unimodal (single bio-marker) type of monitoring, since, although the QT and RR parameters were used in the calculation of QTc in Equation 1, the system should only display the QTc information to the doctor as QTc is the only clinically relevant parameter in this case. For visualization purposes, this is also the only parameter that matters.

B. Bimodal Monitoring

Certain diseases cannot be monitored accurately by visualizing only a single parameter. For example, when a patient is on a drug named Tikosyn [39], (s)he has to be monitored for potentially going into heart failure. For this, in addition to the QTc value described previously, the heart rate (or equivalently, RR) value must be visualized for irregularity. A patient that is showing irregular heartbeats should be red-flagged by the visualizer as a potential cardiac risk. Although this process involves the plotting of two separate bio-markers (QTc and RR), the heartrate itself (RR) is not really the parameter of concern as much as its irregularity. It might suffice to plot the *irregularity* as a single-bit Boolean value, such as a red dot that is associated with each QTc value. From the standpoint of data compression, this has the highest information ratio, but requires a better algorithm to detect RR-irregularity.

C. Multi-modal Monitoring

To extend our discussion from the bimodal monitoring, where a doctor monitors two bio-markers 1) RR-irregularity and 2) QTc, now, let's add a third crucial bio-marker: 3) BNP levels (B-type Natriuretic Peptide). Since BNP level in the blood is a clear indicator of a heart failure, this could serve as a third parameter in the visualization. We assume that this type of monitoring will involve some advanced bio-sensors in addition to ECG sensors, as shown in Figure 1. BNP levels of 100–900 pg/mL indicate degrees ranging anywhere from no heart failure (100 pg/mL) to severe heart failure (900 pg/mL).

Strictly from the standpoint of data handling, we deduce that BNP and QTc are parameters whose values must be shown in a range, whereas RR can either be shown in a range, or simply visualized as a Boolean value such as *regular heartbeat* vs. *irregular heartbeat* for this specific monitoring. For BNP, although the parameter range spans from 100–900, what really matters is whether this value is above 300, denoting a cardiac failure. In the case of QTc, a similar argument can be made, where QTc values over 450 or 500 ms denote *abnormal*.

Since our goal is to provide this information to the doctor in a highly-summarized, yet fully-informational visualization mechanism, added parameters increase the complexity of the visualization task, however, do not necessarily make the visualization more difficult. The increased utility in providing the actual value of BNP is not much more than providing the BNP information in 4 different *levels*, e.g., by using 4 different colors: ≤ 100 as *green*, the range 100–200 as *yellow*, 200-500 as *orange* and 500-900 as *red*. While this reduces the "data burden" on the doctor expotentially, by practically reducing the information to 2 bits (i.e., 4 levels), it barely reduces the *information content*. In the following section, we will develop novel methodologies for the visualization of multi-modal monitoring results based on these concepts.

V. Visualization and Decision Support

Using the data in Figure 3, we can compute the clinical markers such as RR. This allows us to look for problems that the patient may be having. For example, if we want to know how often the patient's heart rate (HR) exceeded 90BPM, we could query for cases where HR is > 90 and get an answer almost instantly. This is an improvement on the current system which may only provide min/average/max heart rates from an entire ECG recording. However, we can improve this even more by providing a view of the *entire* period (e.g. 24 hours) without discarding information like a simple average does.

A. Visualizing long-term monitoring data

To illustrate a case where current visualization methods are not ready for the oncoming flood of sensor data, we first focus

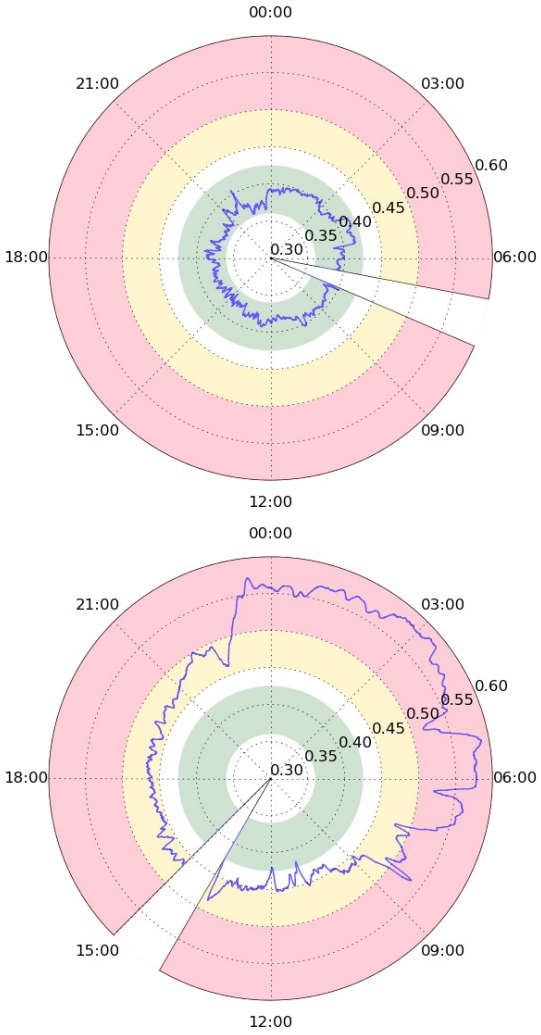

Fig. 4: QTc (in seconds) over 24 hours. Top: healthy patient. Bottom: LQT2 patient. In this plot, colors are used to demarcate healthy (green), borderline (yellow), and abnormal (red) QTc regions.

our attention on a specific example: detecting prolonged QT. The QT interval is a duration that can be measured from an ECG, and it indicates how long the heart's ventricular repolarization cycle takes. Prolonged QT is an important marker for potentially fatal events [40], and subjects with prolonged QT are said to have *Long QT Syndrome* (LQTS). QT varies with heart rate, so it is usually corrected based on the current heart rate [13]. The corrected QT (called QTc) is a more stable value and is around 400ms in a healthy person, and it may go up to 500ms or even higher with LQTS [41]. Long-term monitoring of this (and other ECG features) is typically done using a *Holter monitor*, a portable ECG recorder.

Assume that a cardiologist has 20 patients. Through continuous monitoring and automated data analysis, then, (s)he has access to a table containing yesterday's *two million* QTc values. Obviously, information in this form can not be parsed by a human. Currently, to condense the data into a meaningful snapshot of a patient's day, the doctor will manually spot-check about 10 seconds of the patient's ECG, and review the computed average values of a full 24-hour recording. This system throws away lots of key information; in the case of LQTS, QTc could be prolonged for several minutes or even hours without the doctor noticing the problem. There is a clinical need for a better way to visualize the full data set.

We investigate plotting techniques than can display an entire 24-hour QTc data set in a single picture. Errors in detection due to noise or other sensor issues are common, making naive plots of this type incomprehensible. We will be using three steps to generate plots containing all relevant information: (1) Clean the data by using multiple sources, (2) Remove remaining noise, and (3) Plot the data in a useful and intuitive way. An ECG typically has between 2 and 12 leads. In Step (1), we use the information from all available leads to choose the "best" QTc value at each heart beat (operating on the output of Stage 2 in Section III). In Step (2), the "bad beats" due to sensor noise and hardware/software errors are removed using a median filter. Finally, we observe that a conventional 2D plot of QTc vs. time would "break" at the ends of the plot. Because we expect a patient's QTc to be roughly the same at the start and end of a 24-hour recording, it is convenient in Step (3) to plot QTc on *polar* axes, using radius to indicate QTc and angle to indicate time of day. This form also makes different times of day easily distinguishable. Figure 4 demonstrates the result for two 24-hour data sets. If a cardiologist had only looked at the second patient's *average* QTc (about 500ms), it would have obscured the the fact that their QTc is around 570ms for several hours. The plot takes minimal time to process, and is infinitely more useful than a simple average. With a quick glance, the doctor sees not only if a person is healthy, or if they have prolonged QT, but *when* they have prolonged QT. While we have focused on this QTc example and 24-hour observation periods, the process and framework will be similar to monitor other medical markers such as O_2 saturation or glucose levels, and over different intervals.

B. Analytics-based decision support

The plots presented in Figure 4 have standard clinical "warning" regions marked in yellow; 450ms is considered prolonged QTc in men, 470ms is prolonged in women, and in either case we indicate "danger" in red above 500ms. Green indicates a typical range for healthy people. However, these thresholds do not account for a patient's age, prescriptions, congenital disorders, for the time of day, or many other factors. It is currently possible to build a database from drug trials and other clinical data sets (such as the THEW [36] or PhysioNet [37] databases), constructing a picture of the "norm" for features like QTc within a particular group. As more sensor data (with demographic information) becomes available due to continuous monitoring, these reference ranges will become very well defined for all populations. Figure 5 demonstrates the use of these dynamic reference ranges based on QTc data extracted from the THEW database [36], which can be used to eliminate fixed and overly-general thresholds

such as the *static bands* shown in Figure 4. We can compare them to the expected values *for their specific population* by dynamically adjusting the normal bands. Figure 5 shows how a "normal band" around a 24-hour clock (one σ about the mean) can be created from the histogram data for a specific group (e.g., *healthy male patients*). The plot demonstrates that not all medical markers will be constant with time, e.g., we cannot say that a patient's QTc is *normal* without taking the time of day into consideration.

Figure 5 (top) shows an LQT2 patient with slightly elevated QTc at night, which is not too unusual compared with other LQT2 patients. In Figure 5 (bottom), we see another LQT2 patient with elevated nighttime QTc, which is a very extreme case when compared with their peers, indicating a clear health issue with this specific patient. We could augment these plots further by highlighting areas of the clock where fatal events tend to occur (again, based on hundred or thousands of records for similar patients, not simply on a single static threshold). Such customizations offer much more relevant information to the doctor than any current tool.

The 24-hour plots can also be useful in diagnostics. If a doctor suspects LQTS based on an ECG during a clinical visit, one Holter recording plotted in this form can be used to confirm it. There are many genes that can cause LQTS. Extra QTc prolongation at night is characteristic of LQT2 or LQT3, but not LQT1 [42]. From the plots alone, a cardiologist therefore already has a good idea of the diagnosis for the patients in Figure 5, possibly avoiding an expensive DNA test. The same type of technique can be used in a decision support system: We could attempt to classify/diagnose a patient by matching their Holter recording with the normal ranges of various groups (LQT1, LQT2, etc.), and the doctor can use the picture to help confirm or reject the recommendation.

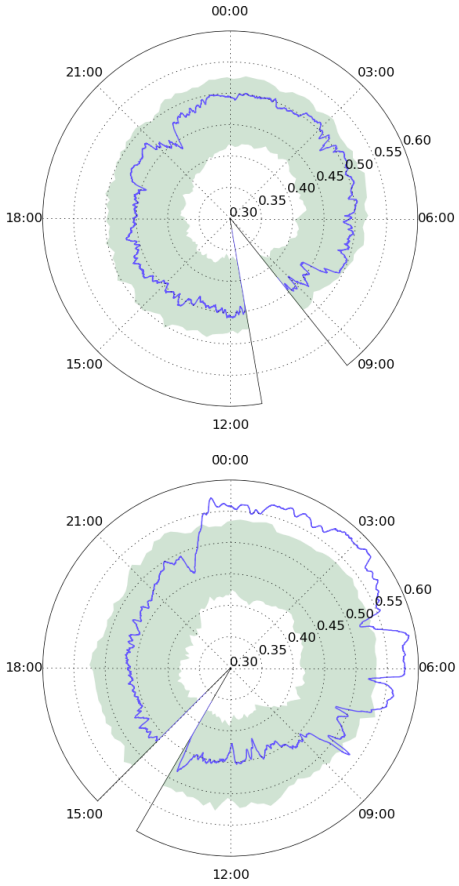

Fig. 5: QTc for two male LQT2 subjects, superimposed on the normal ranges for their peers. The bottom plot is the same recording as the bottom plot in Figure 4. Note that the normal ranges in these two QTc plots are slightly different because one of the patients is on beta blockers, which changes our expectation for QTc (i.e., the dynamically drawn bands).

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we described the building blocks of a system that allows long-term patient monitoring. This system is works with standardized protocols for transmitting the acquired sensor data, such as IEEE 1451 and ISO IEEE 11073. The components that make up this system are data acquisition, data concentration and aggregation, data pre-processing, and visualization. The primary contribution of this paper is the introduction of a novel visualization mechanism that allows a doctor to monitor multiple (e.g., 20–30) patients in real time. Currently, physicians in even the leading US medical institutions use inefficient ECG-based visualization mechanisms: a patient is either monitored for less than a minute through ECG at the hospital, or monitored for 24 hours through Holter devices, for which the summarized results are presented to the physician. Either method has a high probability to miss clinically relevant information related to patient health status.

We have demonstrated a simple way to visualize one factor, QTc. Research is required to determine the best ways to aggregate this heterogeneous sensor data into a clinically-relevant summary for a specific patient and for specific illnesses. This may involve multiple plots/tables, nonlinear axes, and/or methods to combine the information onto the same plot (e.g. 3D views, or interactive animations). The result must take only a few seconds for the doctor to review, without excluding any key information. Another visualization challenge is finding the best way to detect and highlight very short-duration events; 24-hr plots only allow us to see events that last for at least a few minutes, which has insufficient accuracy in some cases.

We currently manage and process hundreds of Holter recordings on a desktop computer. For simplicity, the results are stored in SQLite databases. One research objective will be to scale this to a proper cloud-based database backend capable of handling larger data sets, and sources other than ECG recordings. In parallel with this, we will create a means to automatically update the group-specific statistics when new records are added to the database.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation grant CNS-1239423 and a gift from Nvidia Corp.

REFERENCES

[1] E. J. Petr, C. R. Ayers, A. Pandey, J. A. Lemos, T. Powell-Wiley, A. Khera, D. M. Lloyd-Jones, and J. D. Berry, "Perceived lifetime risk

for cardiovascular disease (from the dallas heart study)," *The American Journal of Cardiology*, vol. 114, no. 1, pp. 53 – 58, 2014.

[2] J. Saul, P. J. Schwartz, M. J. Ackerman, and J. K. Triedman, "Rationale and objectives for ecg screening in infancy," *Heart rhythm : the official journal of the Heart Rhythm Society*, vol. 11, no. 12, pp. 2316 – 2321, 2014.

[3] W.-H. Lin, H. Zhang, and Y.-T. Zhang, "Investigation on cardiovascular risk prediction using physiological parameters," *Computational and Mathematical Methods in Medicine*, vol. 2013, no. 1, pp. 1–21, 2013.

[4] H. Nojiri, T. Shimizu, M. Funakoshi, O. Yamaguchi, H. Zhou, S. Kawakami, Y. Ohta, M. Sami, T. Tachibana, H. Ishikawa *et al.*, "Oxidative stress causes heart failure with impaired mitochondrial respiration," *Journal of Biological Chemistry*, vol. 281, no. 44, pp. 33 789–33 801, 2006.

[5] H. Otani, "Reactive oxygen species as mediators of signal transduction in ischemic preconditioning," *Antioxidants and Redox Signaling*, vol. 6, no. 2, pp. 449–469, 2004.

[6] Alivecor, "ECG screening made easy," http://www.alivecor.com/, 2013.

[7] C. Leaf, "World's Thinnest 3-Lead ECG Patch," http://www.clearbridgevitalsigns.com/brochures/CardioLeaf_ULTRA_Brochure.pdf.

[8] M. Ganesana, J. S. Erlichman, and S. Andreescu, "Real-time monitoring of superoxide accumulation and antioxidant activity in a brain slice model using an electrochemical cytochrome¡ i¿ c¡/i¿ biosensor," *Free Radical Biology and Medicine*, vol. 53, no. 12, pp. 2240–2249, 2012.

[9] J. Njagi, M. M. Chernov, J. Leiter, and S. Andreescu, "Amperometric detection of dopamine in vivo with an enzyme based carbon fiber microbiosensor," *Analytical chemistry*, vol. 82, no. 3, pp. 989–996, 2010.

[10] Sensys Medical, Inc., "Near-Infrared Spectroscopy," http://www.diabetesnet.com/diabetes-technology/meters-monitors/future-meters-monitors/sensys-medical.

[11] National Institute of Standards and Technology, "Advanced encryption standard (AES)," Nov. 2001, fIPS-197.

[12] Mega Electronics Ltd. (2014) eMotion ECG. [Online]. Available: http://www.megaemg.com/products/emotion-ecg/

[13] L. S. Fridericia, "Die Systolendauer im Elektrokardiogramm bei normalen Menschen und bei Herzkranken," *Acta Medica Scandinavica*, vol. 53, pp. 469–486, 1920.

[14] "Ieee standard for a smart transducer interface for sensors and actuators - common functions, communication protocols, and transducer electronic data sheet (teds) formats," *IEEE Std 1451.0-2007*, pp. 1–335, Sept 2007.

[15] "Ieee standard for a smart transducer interface for sensors and actuators wireless communication protocols and transducer electronic data sheet (teds) formats," *IEEE Std 1451.5-2007*, pp. C1–236, Oct 2007.

[16] "Ieee standard for smart transducer interface for sensors and actuators–transducers to radio frequency identification (rfid) systems communication protocols and transducer electronic data sheet formats," *IEEE Std 1451.7-2010*, pp. 1–99, June 2010.

[17] "Health informatics–personal health device communication part 10406: Device specialization–basic electrocardiograph (ecg) (1- to 3-lead ecg)," *IEEE Std 11073-10406-2011*, pp. 1–73, Nov 2011.

[18] "Health informatics–personal health device communication - part 10425: Device specialization–continuous glucose monitor (cgm)," *IEEE Std 11073-10425-2014*, pp. 1–0, Oct 2014.

[19] "Health informatic–personal health device communication part 10441: Device specialization–cardiovascular fitness and activity monitor," *IEEE Std 11073-10441-2013 (Revision of IEEE Std 11073-10441-2008)*, pp. 1–108, March 2013.

[20] B. Selim, T. Iraqi, and H.-J. Choi, "An interoperable mult-sensor system for healthcare," in *IEEE GCC Conference and Exhibition*, Nov 2013, pp. 22–27.

[21] A. Abbas and S. Khan, "A review on the state-of-the-art privacy-preserving approaches in the e-health clouds," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 4, pp. 1431–1441, July 2014.

[22] A. Gkoulalas-Divanis, G. Loukides, and J. Sun, "Toward smarter healthcare: Anonymizing medical data to support research studies," *IBM Journal of Research and Development*, vol. 58, no. 1, pp. 9:1–9:11, Jan 2014.

[23] Y. Fan, Y. Jiang, H. Zhu, J. Chen, and X. Shen, "Network coding based privacy preservation against traffic analysis in multi-hop wireless networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 3, pp. 834–843, March 2011.

[24] O. Kocabas, T. Soyata, J.-P. Couderc, M. Aktas, J. Xia, and M. Huang, "Assessment of cloud-based health monitoring using homomorphic encryption," in *Proceedings of the 31st IEEE International Conference on Computer Design (ICCD)*, Ashville, VA, USA, Oct 2013, pp. 443–446.

[25] A. Page, O. Kocabas, S. Ames, M. Venkitasubramaniam, and T. Soyata, "Cloud-based secure health monitoring: Optimizing fully-homomorphic encryption for streaming algorithms," in *IEEE Globecom 2014 Workshop on Cloud Computing Systems, Networks, and Applications (CCSNA)*, Austin, TX, Dec 2014, p. accepted.

[26] C. Aggarwal, N. Ashish, and A. Sheth, "The internet of things: A survey from the data-centric perspective," in *Managing and Mining Sensor Data*, C. C. Aggarwal, Ed., 2013, pp. 383–428.

[27] J. Vazquez and D. Lopez-de Ipina, "Social devices: Autonomous artifacts that communicate on the internet," in *The Internet of Things*, C. Floerkemeier, M. Langheinrich, E. Fleisch, F. Mattern, and S. Sarma, Eds. Springer Berlin Heidelberg, 2008, vol. 4952, pp. 308–324.

[28] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (iot): A vision, architectural elements, and future directions," *Future Gener. Comput. Syst.*, vol. 29, no. 7, pp. 1645–1660, Sep. 2013.

[29] C. Doukas and I. Maglogiannis, "Bringing iot and cloud computing towards pervasive healthcare," in *Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2012 Sixth International Conference on*, July 2012, pp. 922–926.

[30] T. Soyata, R. Muraleedharan, C. Funai, M. Kwon, and W. Heinzelman, "Cloud-Vision: Real-Time face recognition using a Mobile-Cloudlet-Cloud acceleration architecture," in *Proceedings of the 17th IEEE Symposium on Computers and Communications (IEEE ISCC 2012)*, Cappadocia, Turkey, Jul 2012, pp. 59–66.

[31] E. Agu, P. Pedersen, D. Strong, B. Tulu, Q. He, L. Wang, and Y. Li, "The smartphone as a medical device: Assessing enablers, benefits and challenges," in *10th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, June 2013, pp. 76–80.

[32] iPhone Cardio Monitoring Application, http://www.cardioo.com/.

[33] Y. Jararweh, L. Tawalbeh, F. Ababneh, and F. Dosari, "Resource efficient mobile computing using cloudlet infrastructure," in *IEEE Ninth International Conference on Mobile Ad-hoc and Sensor Networks (MSN)*, Dec 2013, pp. 373–377.

[34] T. Soyata, H. Ba, W. Heinzelman, M. Kwon, and J. Shi, "Accelerating mobile cloud computing: A survey," in *Communication Infrastructures for Cloud Computing*, H. T. Mouftah and B. Kantarci, Eds. Hershey, PA, USA: IGI Global, Sep 2013, ch. 8, pp. 175–197.

[35] B. Oommen, "Recent advances in learning automata systems," in *2nd International Conference on Computer Engineering and Technology (ICCET)*, vol. 1, April 2010, pp. V1–724–V1–735.

[36] J. Couderc, "The telemetric and holter ECG warehouse initiative (THEW): A data repository for the design, implementation and validation of ECG-related technologies," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*. IEEE, 2010, pp. 6252–6255. [Online]. Available: http://www.thew-project.org/

[37] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000 (June 13), circulation Electronic Pages: http://circ.ahajournals.org/cgi/content/full/101/23/e215 PMID:1085218; doi: 10.1161/01.CIR.101.23.e215.

[38] F. Badilini, "The ISHNE holter standard output file format," *Annals of noninvasive electrocardiology*, vol. 3, no. 3, pp. 263–266, 1998.

[39] FDA, "Tikosyn (dofetilide), nda 20-931, risk evaluation and mitigation strategy document," Tech. Rep., August 2011. [Online]. Available: http://www.fda.gov/downloads/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/UCM266277.pdf

[40] H. Morita, J. Wu, and D. P. Zipes, "The QT syndromes: long and short," *The Lancet*, vol. 372, no. 9640, pp. 750 – 763, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0140673608613070

[41] J. N. Johnson and M. J. Ackerman, "QTc: how long is too long?" *Br J Sports Med*, vol. 43, no. 9, pp. 657–662, Sep 2009.

[42] M. Stramba-Badiale, S. G. Priori, C. Napolitano, E. H. Locati, X. Vinolas, W. Haverkamp, E. Schulze-Bahr, K. Goulene, and P. J. Schwartz, "Gene-specific differences in the circadian variation of ventricular repolarization in the long QT syndrome: a key to sudden death during sleep?" *Ital Heart J*, vol. 1, no. 5, pp. 323–328, May 2000.