Abstract—Continuous monitoring of an individual’s health using wearable biomedical devices is becoming a norm these days with a large number of wearable kits becoming easily available. Modern wearable health monitoring devices have become easily available in the consumer market, however, real-time analyses and prediction along with alerts and alarms about a health hazard are not adequately addressed in such devices. Taking ECG monitoring as a case study the research paper focusses on signal processing, arrhythmia detection and classification and at the same time focusses on updating the electronic health records database in real-time such that the concerned medical practitioners become aware of an emergent situation the patient being monitored might face. Also, heart rate variability (HRV) analysis is usually considered as a basis for arrhythmia classification which largely depends on the morphology of the ECG waveforms and the sensitivity of the biopotential measurements of the ECG kits, so it may not yield accurate results. Initially, the ECG readings from the 3-Lead ECG analog front-end were de-noised, zero-offset corrected, filtered using recursive least square adaptive filter and smoothed using Savitzky-Golay filter and subsequently passed to the data analysis component with a unique feature extraction method to increase the accuracy of classification. The machine learning models trained on MITDB arrhythmia database (MIT-BIH Physionet) showed more than 97% accuracy using kNN classifiers. Neuralnet fitting models showed mean-squared error of as low as 0.0085 and regression value as high as 0.99. ECG abnormalities based on annotations in MITDB could be classified and these ECG observations could be logged to a server implementation based on FHIR standards. The instruments were networked using IoT (Internet of Things) devices and ECG event observations were coded according to SNOMED coding system and could be accessed in Electronic Health Record by the concerned medic to take appropriate and timely decisions. The system emphasizes on ‘preventive care rather than remedial cure’ as the next generation personalized health-care monitoring devices become available.

Keywords- Arrhythmia classification; Wearable IoT; Healthcare monitoring; ECG signal processing; Arrhythmia detection; Arrhythmia Neural-Net; MITDB Physionet; GP Connect; HL7; ECG FHIR; SNOMED-CT FHIR; HAPI FHIR

I. INTRODUCTION

Heart disease could be caused by a mix of conditions that are complex to measure and monitor. The Electrocardiogram (ECG) represents the heart condition as a waveform of bio-electric potential and is the most important tool used in cardiac monitoring. The research activity demonstrated in this paper, at an outset, focused on the ECG data acquisition, signal conditioning and filtering followed by classification and predictive analysis whereby an arrhythmia could be electronically detected ‘before’ the heart condition would start to deteriorate. Four types of abnormal patterns of heart rhythms have been considered in this paper for detection and classification of arrhythmia. MITDB (MIT-BIH Database) records were used to generate learning and prediction models using MATLAB WFDB (Waveform Database) software package [5] [6] [7] [8]. The research paper first presents a review on modern wearable ECG kits and current state of the research on ECG data acquisition, followed by data analysis and real time updates to electronic health records. Currently available wearable ECG monitoring systems enable 24 to 48 hours of continuous monitoring, or even longer; these systems however, do not have the predictive analysis component to detect or alert individuals of ECG abnormalities and cardiac risks ahead of time. Many of these devices have been implemented on hardware that restrict mobility, so cannot be used by patients whilst engaged in their day to day activities, which is when they are most likely to suffer a cardiac arrest or a heart attack. The paper also describes the methods used to filter raw ECG waveform and convert it to MITDB format to train ECG classifiers for predictive analysis. A real-time smart IoT (Internet of Things) system has been demonstrated, which is based on a widely accepted HL7 FHIR standard [20]. FHIR standard and its underlying infrastructure is now being adopted by many device manufacturers to enable instruments to update electronic health records in real time.

II. ECG OVERVIEW, DATA ANALYSIS AND HEALTHCARE INTEROPERABILITY PROBLEMS AND CHALLENGES

A. ECG Overview And MIT-BIH Arrhythmia Database

The heart comprises of a muscle called myocardium that is rhythmically driven to contract and drives the circulation
of blood throughout the body. Before every normal heartbeat, a wave of electrical current passes through the entire heart and the pattern of this electrical current and its propagation spreads over the entire structure of the heart and leads to an effective flow of blood in and out of the heart. The measurable change in bio-potential difference (voltage) on the human body surface due to this activity results in a waveform signal which is known as an electrocardiogram (ECG, or EKG). [1] A broad number of factors affect the ECG, which are, but not limited to: abnormalities in cardiac muscles, metabolic abnormalities of the myocardium, posture and state of human body, noise due to other bio-electric activities inside the human body and over the surface of human skin. ECG analysis is a non-invasive procedure and is quite essential in patient health monitoring. A heart rhythm is tachycardic if it is greater than 100 beats/min and bradycardic if the rate is less than 60 beats/min. [1][2] The P, QRS and the T sub-waves are a result of the electrical activity of the heart cells which surface as potential differences on the skin and which can be measured using suitable equipment, which is the ECG kit. The graphical recording of these body surface potentials as a function of time produces the electrocardiogram [3][4]. The 12 lead ECG kit which is extensively used in clinical environment, consists of augmented limb leads (aVR, aVL and aVF) in addition to limb leads (I, II & III) and six chest leads (V1, V2, V3, V4, V5 & V6) forming the basis of the 12 lead ECG analyses. The 3 and 5 lead variations of ECG monitoring does not provide detailed views like a 12 lead ECG, but is most often sufficient for monitoring purposes. [2][3][4] The results obtained in this paper are based on a portable 3 lead ECG kit. For the analyses presented in this paper, MITDB i.e. MIT BIH ECG database maintained and annotated at MIT, was used for data analysis and it is widely used in industry and academia for ECG related research. Several researchers have used MITDB for feature extraction based on ECG morphology and have developed machine learning algorithms for detection and classification of arrhythmia [5]. The database consists of ECG recordings that has a wide range of waveforms that could possibly cover most of the abnormal beat waveforms and which can be used to build a machine learning model and test it. [6] The MIT-BIH Arrhythmia Database contains 48 half-hour two-channel ambulatory ECG recordings, obtained from 47 subjects studied by the BIH Arrhythmia Laboratory. The recordings were digitized at 360 samples per second per channel with 11-bit resolution over a 10 mV range. [6][7] Two or more cardiologists have independently annotated each record beat by beat by identifying abnormal beats in the waveform. Machine learning classification models have been successfully developed in the past with high degree of accuracy [9][14]. The drawback however is that these models tend to accept test data in large samples and perform analysis on entire dataset in a single execution cycle and can’t be used in real-time monitoring in order to generate alarms and alerts related to arrhythmia.[27] To overcome the drawback, the signal processing and analysis section demonstrates a pipelined operation of data acquisition in burst mode followed by ECG classification in real time. There also exists a high degree of correlation between Heart Rate Variability (HRV) and arrhythmia. The problem however is that HRV analysis depends on the morphology of the ECG waveform and QRS detection [10][15], which depends on the accuracy of the ECG equipment and accurate 12-lead ECG equipment are not portable and certainly not wearable. Furthermore, the QRS waveforms and ECG recordings are relative to an individual and the HRV (Heart Rate Variability) analyses and the risks inferred from the readings are influenced by gender and age information [37]. In order to increase the accuracy, several signal processing techniques were used along with a unique feature extraction method as explained in signal processing and analysis section of this paper.

B. Wearable ECG Kits and Healthcare Interoperability Infrastructure

There are commercially available wearable 3 Lead ECG kits which can take ECG sample readings while the person under observation is engaged in day to day activities. A Holter monitor is a wearable ECG monitoring machine, that can be carried in a pocket or pouch worn around the neck or waist and it continuously records the heart's rhythms for 24 to 48 hours during normal human activity [26]. The Holter monitor only captures ECG signals and the readings have to be carried to the concerned medic for further analysis and interpretation. Holter monitor and similar kits focus only on monitoring and data-acquisition and the intelligent aspect of prediction and raising appropriate alarms is never provided or is limited in functionality. The system proposed in this paper consists of a software implemented on the wearable device which encapsulates the sample readings in a standard JSON format [11] and transmits the samples to the analytical server that implements the classification and prediction algorithms using the utilities provided by MITDB WFDB database. Attempts have been made on previous occasions to make real time updates to electronic health records using internet client server technologies. [32] To develop such an infrastructure is however, a government policy decision rather than an implementation exercise. The infrastructure has to be widely accepted by masses in a particular country and also by the government policy decision makers.

![Figure 1. Filtering the ECG signal obtained from 3-lead ECG kit using MATLAB FD/a tool adaptive filter to cancel noise.](image-url)
HL7 (Health Level-7) is a standards agency that develops standards for electronic health resources and FHIR is a standard that enables electronic health records (EHR) interoperability through web-service interfaces. [20] In the next section one such implementation is described that demonstrates real-time updates to an EHR sandbox using FHIR web-services and SNOMED CT coding system, which is a coding system to describe clinical concepts, vocabulary, descriptions and inter-relationships between clinical codes.

III. SIGNAL PROCESSING, DATA ANALYSIS AND HEALTHCARE INTEROPERABILITY RESOURCES

A. Signal Processing, Dataset Preparation and Analysis

In this section, methods to combine real time ECG samples acquisition with digitization according to MITDB have been demonstrated followed by preparation of dataset for data analysis. The 3-lead ECG sensor was used for data acquisition which is based on Analog Devices ADI 8233 [31] analog front-end. The AD8233 is an integrated signal conditioning block for ECG and other biopotential measurement applications. It has been designed to extract, amplify, and filter small biopotential signals in the presence of noisy conditions, such as those created by motion or remote electrode placement. The design allows an ultralow power analog-to-digital converter (ADC) or an embedded microcontroller to easily acquire the output signal. The raw samples were captured at 1 KHz 16 bit resolution as dictated by MITDB database as the samples had to be digitized according to a format acceptable by MITDB database such that the WFDB routines in MATLAB could be used on these sampled waveforms. The 3-lead ECG kits samples are noisy due to bio-electric interference of the external environmental conditions, the body posture and motion. The signal was de-noised using MATLAB Filter Design and Analysis Tool for implementing recursive least square (RLS) adaptive filter for noise cancellation (Figure 1) and by implementing a 2nd order High Pass Butterworth filter to attenuate signals at 1 KHz, followed by zero-offset signal conditioning and removing the baseline wandering (Figure 2). [28][29] For further baseline wandering correction and for signal smoothing Savitzky-Golay filter with order N=4 and frame length of 15 was applied to smooth the signal without destroying the original signal properties in digital format. [30] In order to prepare datasets to train classifiers, this digitized signal was sampled using the ‘WRSAMP’ utility at 128 Hz and the signal was amplified in such a way that a step of 1 millivolt would appear as sample values that differ by 102.4 units [6] [7]. The utility generated a binary signal file (.dat extension file which contained the actual signal) and a signal header file (.hea extension which contained the signal processing information). Before beginning the study of RR intervals (inter-beat intervals) the exact locations (times of occurrence) of the QRS complexes in the ECG had to be determined. In MITDB database this information is present in an annotation file (with .ann or .qrs extension). To create the beat annotation file to determine QRS complexes, ‘GQRS’ utility was used, which created the QRS peak detection annotation file. To determine RR intervals the ‘ANN2RR’ utility was applied across the acquired digital binary signal (.dat file) and by using the QRS annotation file (.qrs extension). The acquired signal was now ready for data analysis. To extract additional features ‘ECGPUWAVE’ could be used to locate ‘P’ and ‘T’ waves and their boundaries. The acquired signal was going to be used as a test signal, so a classifier model had to be developed and trained and validated on MITDB database.

In order to develop the classifier model, initially, from each of the 47 records in the MITDB database, normal beats were removed from the samples and only the abnormal beats that were annotated according to the WFDB annotation files, were read by the ‘RDANN’ utility. The utility uses MIT BIH database's own annotation file, and since this file was manually annotated by ECG experts, correct results could be obtained. [6][7] For all the records, the feature vectors that attributed to abnormal heart rhythm were identified. The important features were: Age, Gender, the ECG signal amplitude for each sample (millivolts), RR interval (inter-beat interval in milliseconds), WABP (arterial blood pressure in millivolts) and instantaneous heart rate measured at the ‘instance’ when the abnormal heart beat annotation occurred in ECG recording.[27] ‘WABP’ utility was used to generate Arterial Blood Pressure (ABP) annotations. The ‘WRANN’ utility in WFDB toolkit annotated the waveform as required by the MITDB database for further analysis. The ‘TACH’ utility was used to obtain the instantaneous heart rate from the samples. These feature vectors were derived for 4 annotation types: V (Premature Ventricular Contraction: PVC), A (Atrial Premature Beat: APB), L (Left bundle branch block beat), R (Right bundle branch block beat) [6] [7] [27]. It has been observed that these 4 annotation types do occur in ECG recordings of healthy subjects as well and can go unnoticed without showing any symptoms [16]. It may take as many as three consecutive PVCs before a ventricular tachycardia is detected or confirmed [17]. Similar argument could be made for ‘A’ type annotation which refers to Premature Atrial
Complexes (PAC) and though it can go undetected without showing any symptoms, it may lead to atrial fibrillation [18]. An impact factor was calculated from age and frequency of distribution information so as to associate weights to the recordings and to determine how much an individual’s ECG recording impacted the overall classification process. [27]

Classification algorithms were initially trained on the feature vectors with 70% training set, 15% validation set and 15% test set, however, other combinations of percentage values have also been experimented with results shown in Table 1. From each of the 47 records in MITDB database about 650,000 samples per record were used to train the classifiers in order to classify a heartbeat sample as belonging to a category (or label) of an abnormal beat type. From the experiments performed using several classifiers, k-NN classifiers produced more than 97% accuracy as shown in Figure 3. The figure shows 4 different annotation types representing 4 different types of arrhythmia denoted by V, A, L, R annotation types. In addition, neural network fitting and pattern recognition methods were also experimented with. Table 1 shows pattern recognition accuracy over several combinations of percentages of training-validation-test data showing no bias or over-fit. [27] The MATLAB Neural Net Fitting models with 10 hidden layers and Levenberg-Marquardt training algorithm yielded the values as presented in Table 1.

### Table 1. Neuronal Pattern Recognition Results for Various Percentage of Training Validation and Test Data

<table>
<thead>
<tr>
<th>Type of Neural Network</th>
<th>Percentage (%) Training- Validation- Test data</th>
<th>Mean-squared error (MSE) and Regression R for test data.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuronal Fitting</td>
<td>70-15-15</td>
<td>MSE= 0.0085 and R= 0.99</td>
</tr>
<tr>
<td></td>
<td>60-20-20</td>
<td>MSE=0.0017 and R=0.99</td>
</tr>
<tr>
<td>Neuronal Pattern</td>
<td>70-15-15</td>
<td>Cross-Entropy Error for test data: 7.6 Misclassification Error: 1.2%</td>
</tr>
<tr>
<td>Recognition</td>
<td>60-20-20</td>
<td>Cross-Entropy Error for test data: 9.9 and Misclassification Error: 2.1%</td>
</tr>
<tr>
<td>MATLAB Neural Net</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Figure 3](image.png)

Figure 3. k-NN classifier with Number of neighbors: 10; Distance metric: Euclidean; Distance weight: Squared inverse; Principal Components:4 out of 6 yielding Accuracy of 97.1% for classifying 4 predictor annotation classes of Arrhythmia in MITDB database. Ensemble Bagged Tree classifier produced classification accuracy of 97.4% without PCA (Principal Component Analysis) and 95.4% with 3 Principal Components. The MATLAB classification learner was set to cross-validate at 10- fold validation and 25% hold-out validation. [27]

![Figure 4](image.png)

Figure 4. Confusion matrix for Ensemble Bagged Trees method for classifying from a total of 24190 samples across all 47 MITDB records. Feature vector (age, gender, signal strength, RR interval, heart rate, impact factor) was used for classifying (1,2,3,4) corresponding to (V, A, L, R) type of annotations. [27]

### B. HL7 FHIR Interoperability

A 3-lead ECG kit was interfaced to an Arduino microcontroller which transmitted the samples to the Linux system running WFDB/MATLAB through a serial port. The WFDB analytical server software on Linux system performed arrhythmia detection and classification tasks. A JSON (JavaScript Object Notation) structure [12] that encapsulated the sample readings was adopted which passed the samples to the HAPI (HL7 API) FHIR test server [33][34] to be logged to EHR if anomalies were detected (Figure 5). The ECG samples encapsulated with JSON data structures were transmitted over 10 seconds intervals as ECG rhythm strips are generally for a 10 seconds duration.
[2][3] and ECG readings conformed to HL7 and FHIR standards specification [13]. HAPI FHIR is Java implementation of FHIR and enables development of RESTful Webservices. As listed in Table 2, an Observation object encapsulates an instance of observation and was coded according to SNOMED-CT coding system (code: 428803005 for 3-lead ECG monitoring) [21]. An observation object could be an event of abnormal signal in ECG waveform. The HAPI FHIR data store, (Figure 5) which could be a FHIR based EHR like GP-Connect/GP-SoC from NHS UK [19][23]. Once abnormal beats or waveforms were detected appropriate alarms could be raised and potentially be passed on to the health-care agency entrusted with the patient care. The FHIR specification on security also supports OAuth authentication and authorization service which the consumer services can embed in their FHIR servers. [24][25]

![Image of a diagram related to real-time ECG monitoring system using FHIR client implemented on portable IoT device. ECG FHIR specific Observations were coded according to SNOMED CT coding system.](image)

**TABLE 2.: HAPI FHIR CODE TO UPLOAD ECG OBSERVATIONS TO FHIR SERVERS**

```java
// Create a FHIR Observation object.
Observation observ = new Observation();
// Assign a randomly generated Universal ID (UUID).
observ.setId(uuid)
// Set the Observation code according to a Coding System
// Coding System refers to ECG code in SNOMED CT
observ.setValue(new QuantityDt() .setValue(1200) .setUnit("mV")
// Set the Date and Time stamp for the observation
observ.setIssued(new InstantDt("2017-05-05T15:30:10+01:00"))
```

---

**IV. CONCLUSIONS AND DISCUSSION**

The motivation behind the research was to combine a wearable ECG monitoring kit with real-time arrhythmia classification and prediction server, raise appropriate alarms and at the same time upload and log the events to electronics health records database using HL7 and FHIR standards. In addition to data analysis, the major challenge was signal conditioning as 3 lead ECG kits were used on individuals engaged in day to day activities. The de-noising and filtering methods based on adaptive filters, smoothing using Savitzky Golay filters and elimination of baseline wandering of ECG signals helped gathering and digitizing the signal for further analysis. The learning models trained using MITDB arrhythmia database showed 97% and higher percentage of accurate results in classification and prediction. Instead of unsupervised learning algorithms, supervised learning models based on 6 features (Figure 3) yielded more accurate results in classification and prediction of 4 types of arrhythmia (V, A, L, R annotations in MITDB). Pattern recognition using neural networks was also experimented with and Table 1 shows pattern recognition accuracy over several combinations of percentages of training-validation-test data showing no bias or overfit. Despite of the availability of accurate classification [9] algorithms, ECG equipment could not be used in the past for real time ECG classification due to non-real time batch processing nature of the algorithms, where the analysis was done after data acquisition stage. Traditionally, ECG arrhythmia classification relied on HRV analysis which produced accurate results [15][22] though it relied on accurate ECG equipment which was not portable or wearable. The signal conditioning algorithms and feature extraction methods illustrated in this paper produced almost as accurate results as produced by machine learning and feature extraction models that relied on HRV analysis and ECG morphology. By combining the real-time data acquisition, filtering and signal processing mechanisms with asynchronous JSON based transmission in burst intervals of 10 seconds each, the classification server could analyse and classify arrhythmia continuously in real time and could raise appropriate alarms. An implementation of a HAPI-FHIR test
server demonstrated the real-time logging of ECG observations, according to a standard SNOMED coding system, to EHR for further analysis by the general practitioners and medics. This should assist incident response medical health-care teams to prepare for an emergency ahead of time and may prevent or reduce hazardous situations. With further consolidation and standardization of FHIR based EHR and other ubiquitous platforms the same model could be extended to monitor and respond to emergencies related to other health conditions of individuals whilst they remain engaged in day to day activities.

REFERENCES


