

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Machine Learning in Wearable Biomedical Systems

*Muhammad E.H. Chowdhury, Amith Khandakar,  
Yazan Qiblawey, Mamun Bin Ibne Reaz,  
Mohammad Tariqul Islam and Farid Touati*

## Abstract

Wearable technology has added a whole new dimension in the healthcare system by real-time continuous monitoring of human body physiology. They are used in daily activities and fitness monitoring and have even penetrated in monitoring the health condition of patients suffering from chronic illnesses. There are a lot of research and development activities being pursued to develop more innovative and reliable wearable. This chapter will cover discussions on the design and implementation of wearable devices for different applications such as real-time detection of heart attack, abnormal heart sound, blood pressure monitoring, gait analysis for diabetic foot monitoring. This chapter will also cover how the signals acquired from these prototypes can be used for training machine learning (ML) algorithm to diagnose the condition of the person wearing the device. This chapter discusses the steps involved in (i) hardware design including sensors selection, characterization, signal acquisition, and communication to decision-making subsystem and (ii) the ML algorithm design including feature extraction, feature reduction, training, and testing. This chapter will use the case study of the design of smart insole for diabetic foot monitoring, wearable real-time heart attack detection, and smart-digital stethoscope system to show the steps involved in the development of wearable biomedical systems.

**Keywords:** machine learning, wearable medical devices, real-time detection, sensors for wearable devices

## 1. Introduction

Healthcare is one of the most crucial application areas that require faster development. In the wearable medical device domain, there is a lot of scope of research and development. With the recent technological development, wearable medical devices become feasible for real-time monitoring and diagnosis of medical problems. It has been accelerated with the availability of cheaper component and advancement in wireless communication technology, which has also enabled the application of the Internet of Things (IoT) in the wearable biomedical applications. According to a report by Grand View Research Inc. [1], the Internet of Things (IoT) penetration in the health sector is expected to reach a volume of approximately \$ 409.9 billion by 2022. Thus, there is a huge boom in the wearable industry. There is a wide range of wearable devices that transmit medical information to mobile and

web applications over the wireless network, enabling remote health monitoring. The wearable medical devices face the challenges of reliability, accuracy, precision, and robustness [2]. There are two important sub-sections for a complete wearable medical device—(i) hardware, which involves the sensor selection and characterization, noise removal from the acquired signals, and communication to the decision-making subsystem and (ii) software, which involves taking decisions based on the acquired signals. The second subsystem is where the machine learning comes into play, and the recent outburst of development in the field of machine learning has increased the possibility of remote monitoring and diagnosing with the acquired data from the wearable devices. The machine learning block of the wearable devices involves various steps such as preprocessing, feature selection, training on the labeled dataset, and testing to verify its accuracy and competency. The various parts of the machine learning block will be discussed in details later in this chapter. Thus, it is inevitable that machine learning methods that include algorithm training are required to enable the diagnosis of different diseases. Examples of success in machine learning lead researchers to the adoption of neural networks and other classifiers for medical diagnostic applications. However, this requires ground truth in the form of accurately curated data sources from carefully constructed subject trials with a properly distributed subject population of significant size. Limitations in ground truth data may lead to inaccurate results from the classifiers. This also includes a novel set of challenges for the design and execution of subject trials that are essential in the development and training of algorithms, as well as verification of system performance. Different open-source databases were used for hardware evaluation. The chapter is intended to provide a guidance to the researchers in computing for healthcare who wish to adopt the compelling mission of wearable device development for healthcare diagnostics.

The chapter draws from the experience in the development of a series of biomedical devices from primary research done in the field of wearable devices for monitoring heart attack, abnormal heart sounds, gait analysis in diabetic foot monitoring, and blood pressure estimation using photoplethysmogram (PPG) signal. Therefore, in this chapter, four major health issues were discussed and how these issues can be well monitored without visiting hospital facilities were discussed. These experiments from various previously performed case studies by the team can help in providing a step-by-step roadmap on how to develop intelligent wearable biomedical system. Section 2 discusses the sensor selection and characterization process for a smart insole for gait analysis in diabetic foot monitoring. Section 3 discusses a case study, how continuous blood pressure can be estimated very accurately using PPG signal with the help of machine learning algorithm. Section 4 shows a wearable device for heart attack detection for the driver in a vehicle environment using classical machine learning approach, while Section 5 describes a wearable monitoring device, which is developed for heart sound monitoring to identify cardiovascular abnormality. Section 6 discusses the key contributions of the chapter along with the major results and their applications in real world. Finally, Section 7 concludes the chapter.

## **2. Smart insole for gait analysis**

Gait analysis is an important feature for diabetic foot monitoring. A force between the foot and ground, called the vertical ground reaction force (vGRF), can be acquired using different wearable sensors such as force sensors, strain gauges, magneto-resistive sensors, accelerometers, gyroscopes, and inclinometers or using different type of instruments such as force plate system and

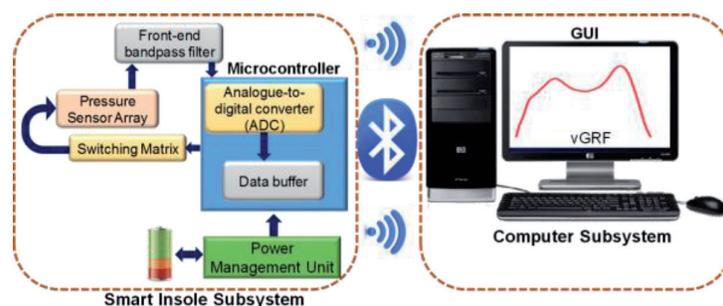
multi-camera-based system. Despite the improvements in technology, the cost of the multi-camera setup is high and it requires a long post-processing time. Furthermore, the number of strides that can be captured is limited, making it not applicable for personal use. Some efforts from different research groups have been made to overcome the limitations of traditional force plates by equipping treadmills with few force plates [3]. However, treadmills impose restrictions as subjects cannot change their direction of walking. Therefore, an increase in research interest was observed toward the development of smart insoles, where wearable sensors can be employed to detect vGRF, joint movements, acceleration of lower extremities, and other gait variables [4]. Smart insole provides the researchers with a flexible, portable, and comfortable solution for vGRF measurement. It includes different features including monitoring, processing, and displaying plantar pressure using pressure sensors embedded in the insole.

**Figure 1** shows the complete block diagram of such a wearable system, where the pressure sensor array is placed in a customized shoe above the control circuit. Pressure data have to be digitized in the insole subsystem before it is sent via a Bluetooth to a host computer for post-processing and analysis. This subsystem is powered by a battery alongside a power management unit. Pressure data can then be analyzed to extract various gait characteristics for different gait applications.

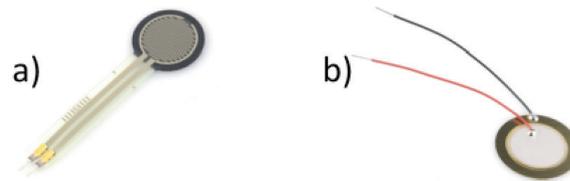
In this section, the design of smart soles to detect vGRF during gait, with two different types of low-cost commercial force sensor: FSRs and flexible piezoelectric sensors (**Figure 2**), is discussed. All the sensors can be calibrated, details of which can be found in [5], where a simple low-cost calibration method based on load cells has been shown to mitigate the need of expensive calibration devices or Motion Analysis Lab as a calibration reference.

As shown in **Figure 3**, the most common places of foot plane where most of the pressure is exerted during gait are heel, metatarsal heads, hallux, and toe. These are the places where the calibrated sensors are to be placed on the smart insole.

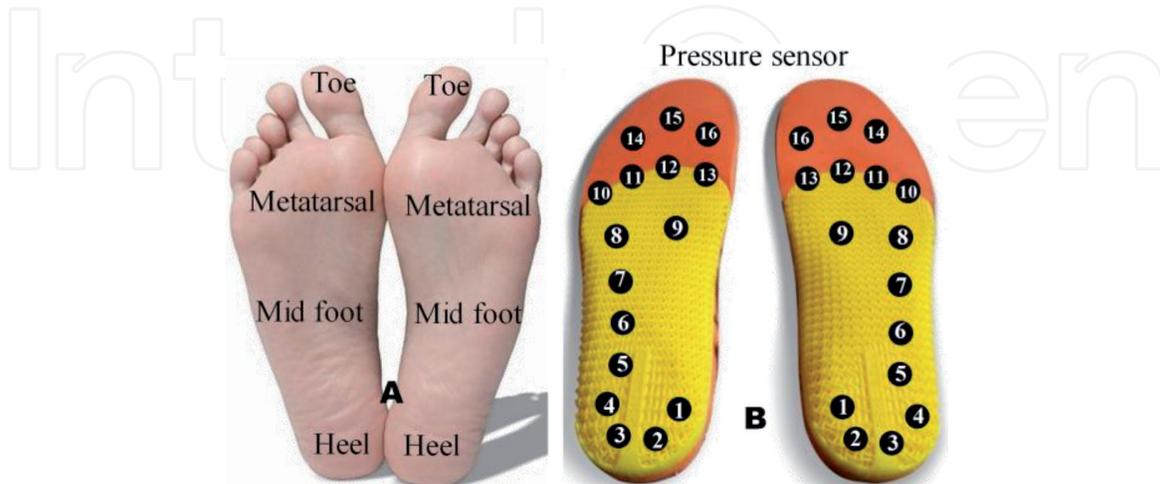
In each insole, 16 sensors were positioned to record pressure values in different locations of the foot sole as shown in **Figure 3b**. Since people typically do not exert pressure on the Medial arch area, there is no sensor placed in this area [8]. Sixteen FSRs/piezoelectric sensors were used to collect data from smart insole, as can be seen in **Figures 4** and **5**. Sixteen inputs were multiplexed to one output through a 16-to-1 multiplexer, then applied to an ADC input of the microcontroller, and sent to host computer. The acquisition circuit was connected to the microcontroller module using an insulated ribbon cable to minimize wire path and electrical hazard. The instrumented insole was placed on the top of the insole of the user's shoe to acquire pressure data while the other end of the ribbon cable is connected to a small box with an adhesive strap to wrap around the leg. Real-time pressure data from different sensors were collected in the microcontroller and sent to a PC using



**Figure 1.**  
*Block diagram for a diabetic foot monitoring smart foot sole.*



**Figure 2.**  
(a) FSR sensor (Interlink Electronics) [6] and (b) piezoelectric sensor (Murata Manufacturing Co.) [7].



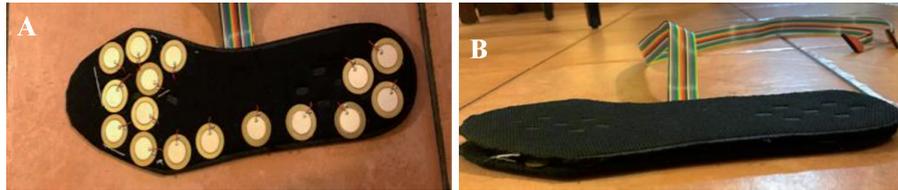
**Figure 3.**  
Selected foot area for pressure acquiring (a) and array of pressure sensor (b).



**Figure 4.**  
FSR-based smart insole: top (a) and bottom (b).

Bluetooth low energy interface, and in the PC, a graphical interface was used to display and analyze the signal.

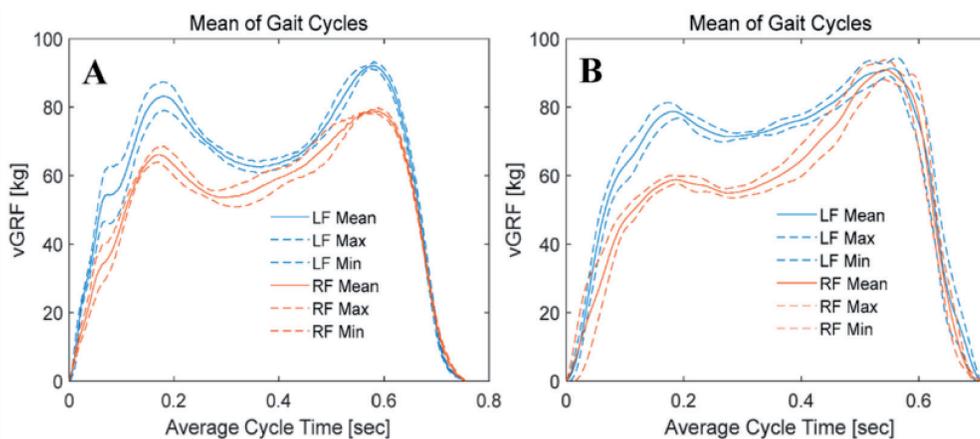
A commercial smart insole system called F-scan, which is one of the best insoles currently available in the market, was used to validate the performance of the designed smart insole (**Figure 6a**). The F-scan system is designed with very thin flexible printed circuit board with 960 sensing points, and each point can measure pressure with eight-bit resolution with a sampling frequency of maximum 750 Hz. However, the F-scan system is very expensive; the wired system costs around \$13,000, while the wireless systems costs around \$17,000. On the other hand, the proposed instrumented insole costs only ~500\$. Typically the vGRF peak is only 10% of the subject's weight. So, the data were calibrated based on the subject's weight. The subject is supposed to stand on one foot for few seconds while wearing the smart insole, and the pressure is calculated for the purpose of calibration. If the subject's weight is higher than the pressure measured, the calibration was done by a calibration factor greater than 1 while the calibration factor is less than 1 for



**Figure 5.** (a) Piezo insole with 16 piezo sensors and (b) additional insole layer placed on the top of piezo insole to ensure comfortability.



**Figure 6.** Commercial F-scan smart insole (a), subject 01 is wearing the F-scan system (b), and FSR-based prototype system (c).



**Figure 7.** Comparison between the mean and standard deviation of vertical ground reaction forces (vGRFs) from left (blue) and right (orange) foot using F-scan system (a) and FSR-based system (b).

patient's weight less than the applied pressure. Calibration factor was calculated in the similar way for the FSR insole. **Figure 6b** and **c** show the application of F-scan and FSR-based system on the same subject while acquiring the vGRF signal.

The FSR-based smart insole can acquire high-quality vGRF for different gait cycles, as seen in **Figure 7**, and it was found that the flexible piezoelectric sensors were performing poor in calibration due to their sensitivity to 3D forces requiring special force calibration machines to control the applied force in x, y, or z directions. Therefore, piezoelectric sensors cannot be utilized as a substitute for FSR in designing smart insole. However, the piezoelectric sensors if calibrated properly can be used to detect the starting and ending of each gait cycle.

### 3. Estimating blood pressure from the photoplethysmogram signal

It is very important to monitor the blood pressure as high blood pressure puts a huge amount of strain on the arteries and the heart, resulting in increase in the

probability of clogging of lumens. This clot (clogging of lumens) may cause several health problems such as heart attack, stroke, kidney diseases, and dementia. Thus, continuous blood pressure monitoring, at least four times daily at regular intervals, is preferable. Photoplethysmography (PPG) can be used for measuring the amount of light absorbed or reflected by blood vessels in the living tissue which can be extended to different aspects of cardiovascular surveillance including identification of blood oxygen saturation, heart rate, BP estimation, cardiac output, respiration, arterial aging, endothelial control, microvascular blood flow, and autonomic function. PPG waveforms generally have three distinct features—systolic peak, diastolic peak, and a notch in between—as shown in **Figure 8**.

This section will discuss how the less cumbersome method of PPG signal acquisition can be used in estimating blood pressure, details of which can be found in [9]. This section will also discuss how a trained network using reliable labeled dataset can be used in such a study.

### **3.1 Dataset**

The dataset used in this study was taken from Liang et al. [10], which is publicly available. The dataset contained 657 PPG signal samples from 219 subjects. PPG signals were first assessed to check the signal quality and then randomly divided into two sets. Then, 85% of the data were used for training and validation, and 15% of the data were used for testing the performance of the model. The PPG signals were preprocessed before they were sent for feature extraction. After extracting meaningful features, feature selection techniques were used to reduce computational complexity and the chance of over-fitting the algorithm. The features were then used to train machine learning algorithms. The best regression model was selected for SBP and DBP estimation individually. In the quality assurance process, 222 signals from 126 subjects were finally kept for this study. **Figure 9** shows the sample PPG signal which was divided into fit and unfit for the study.

### **3.2 Normalization and filtration**

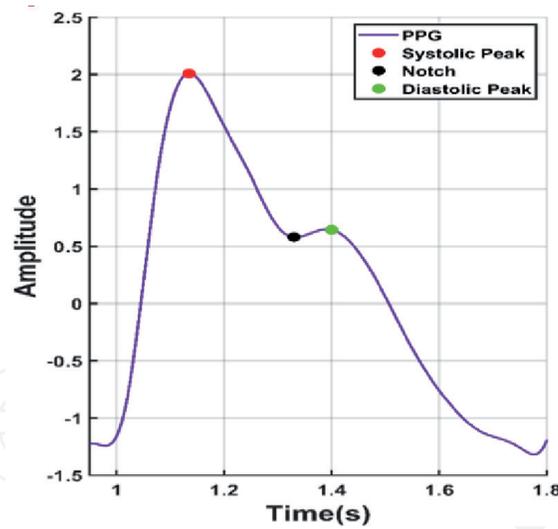
To extract meaningful information from the signals, it was necessary to normalize all the signals. The Z-score technique was used to normalize the signals in this study to get amplitude-limited data. The normalized signal is passed into a low-pass Butterworth filter to remove the high frequency that is acquired by the signal.

### **3.3 Baseline correction**

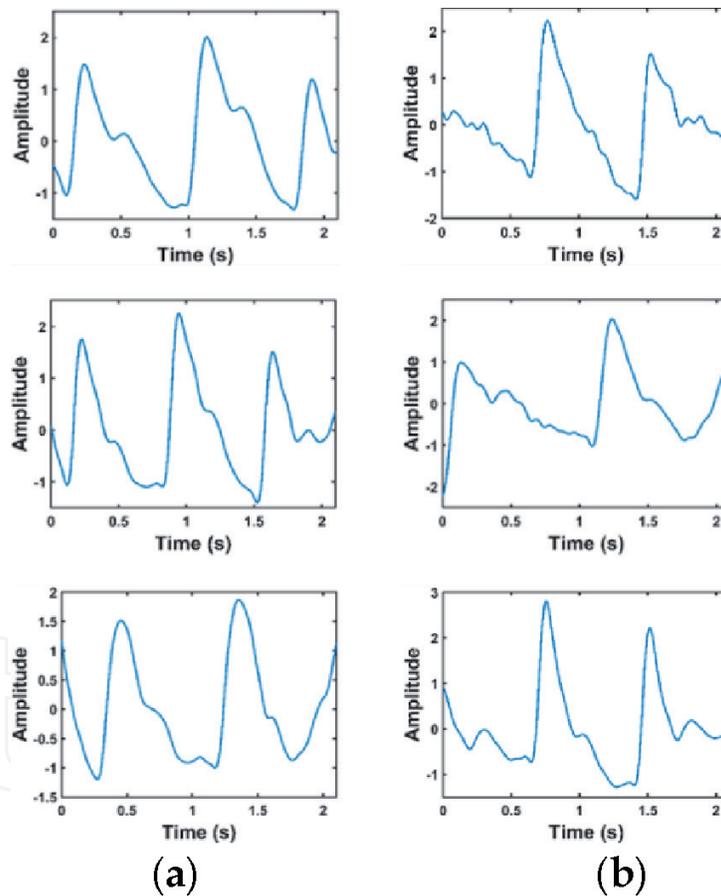
The PPG waveform is commonly contaminated with a baseline wandering due to respiration at frequencies ranging from 0.15 to 0.5 Hz. Therefore, it is very important that the signal is properly filtered to remove the baseline wandering but that important information is preserved as far as possible. A polynomial fit was used to find the trend in the signal and subtracted to get the baseline corrected signal, as shown in **Figure 10**. The next step is to extract features from the preprocessed signal.

### **3.4 Feature extraction**

The block diagram summarizing the feature extraction details adopted in the study is shown in **Figure 11**. A PPG waveform contains many informative information such as systole, diastole, notch, pulse width, and peak-to-peak interval. Some of the



**Figure 8.**  
*A typical photoplethysmography (PPG) waveform with notch, systolic peak, and diastolic peak.*

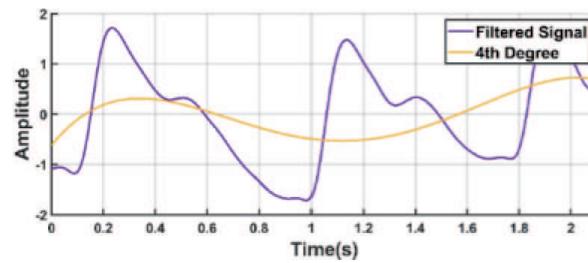


**Figure 9.**  
*Comparison of waveforms that are fit and unfit for the study. (a) Fit data and (b) unfit data.*

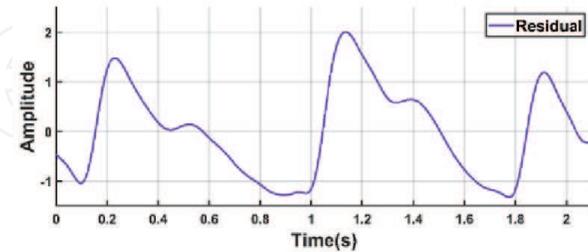
distinctive features of the PPG waveform might not be dominant in some patients, such as the notch prevalence changing with age.

### 3.5 Feature reduction

Feature selection or reduction is important to reduce the risk of over-fitting the algorithms. In this section, three feature selection methods: correlation-based



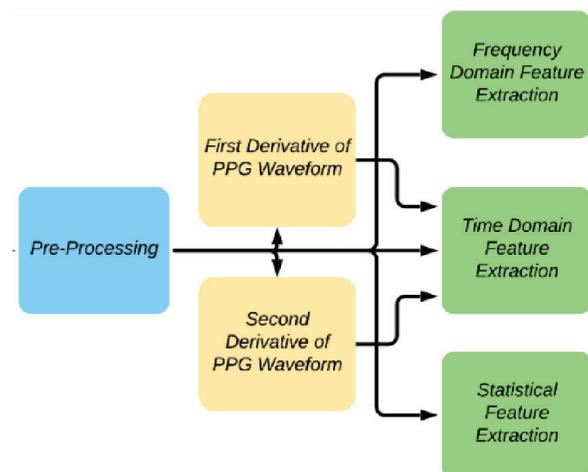
(a)



(b)

**Figure 10.**

Baseline correction of PPG waveform. (a) PPG waveform with the baseline wandering and fourth degree polynomial trend and (b) PPG waveform after detrending.

**Figure 11.**

Overview of feature extraction.

feature selection (CFS), ReliefF features selection, and features for classification using the minimum redundancy maximum relevance (fscmr) algorithm is discussed. ReliefF is a feature selection algorithm, which randomly selects instances and adjusts the weights of the respective element depending on the nearest neighbor. Correlation is a test used to evaluate whether or not a feature is highly correlated with the class or not highly correlated with any of the other features. On the other hand, the fscmr algorithm finds an optimal set of features that are mutually as dissimilar as possible and can effectively represent the response variable. The algorithm minimizes a feature set's inconsistency and maximizes the relevance of a feature set to the answer variable. MATLAB built-in functions can be used for CFS, ReliefF, and fscmr feature selection algorithms.

After the features were extracted, the feature matrix was trained with machine learning algorithms. The Regression Learner App of MATLAB 2019b was used to estimate the BP. Five different algorithms [linear regression, regression trees, support vector regression (SVR), Gaussian process regression (GPR), and ensemble trees] with their variations to a total of 19 algorithms were trained using the 10-fold

Selection criteria	Performance criteria	Systolic blood pressure		Diastolic blood pressure	
		Optimized GPR	Optimized ensemble trees	Optimized GPR	Optimized ensemble trees
Features from the literature	MAE	6.79	12.43	4.49	8.17
	MSE	180.99	231.15	70.06	104.45
	RMSE	13.45	15.20	8.37	10.27
	R	0.79	0.73	0.74	0.57
All features (newly designed and from the literature)	MAE	3.30	10.886	2.81	7.96
	MSE	72.95	264.24	30.70	111.97
	RMSE	8.54	16.25	5.54	10.58
	R	0.92	0.67	0.90	0.56
ReliefF	MAE	<b>3.02</b>	11.32	<b>1.74</b>	5.99
	MSE	<b>45.49</b>	284.69	<b>12.89</b>	62.04
	RMSE	<b>6.74</b>	16.84	<b>3.59</b>	7.88
	R	<b>0.95</b>	0.65	<b>0.96</b>	0.78
FSCMRMR	MAE	6.11	14.65	6.80	8.22
	MSE	108.96	321.63	77.26	110.84
	RMSE	10.44	17.93	8.78	10.53
	R	0.88	0.58	0.72	0.56
CFS	MAE	12.95	16.27	7.59	7.89
	MSE	361.96	448.25	108.43	106.72
	RMSE	19.02	21.17	10.41	10.33
	R	0.50	0.28	0.57	0.58

**Table 1.** Evaluation of the outperforming algorithms for estimating systolic blood pressure (SBP) and diastolic blood pressure (DBP) after optimization.

cross-validation. Out of all these algorithms, two best performing algorithms, Gaussian process regression and ensemble trees, were tested.

### 3.6 Results

To evaluate the performance of the ML algorithms for estimating BP, it is important to select proper metrics, i.e., four in this case and are stated below. Here,  $X_p$  is the predicted data while the ground truth data is  $X$  and  $n$  is the number of samples.

$$\text{Mean absolute error, MAE} = \frac{1}{n} \sum_n |X_p - X| \quad (1)$$

$$\text{Mean squared error, MSE} = \frac{\sum |X_p - X|^2}{n} \quad (2)$$

$$\text{Root mean squared error, RMSE} = \sqrt{\frac{\sum |X_p - X|^2}{n}} = \sqrt{\text{MSE}} \quad (3)$$

$$\text{Correlation coefficient, } R = \sqrt{1 - \frac{\text{MSE}(\text{Model})}{\text{MSE}(\text{Baseline})}} \quad (4)$$

$$\text{where } \text{MSE}(\text{baseline}) = \frac{\sum |X - \text{mean}(X)|^2}{n}$$

When using the Regression Learner App in MATLAB, the above criteria are automatically calculated by MATLAB and these values can be used to evaluate the performance of the algorithms.

In **Table 1**, it can be noticed that the ReliefF feature selection algorithm produced the best result when combined with GPR. The feature selected using a combination of ReliefF and GPR performed the best in estimating SBP, while CFS and GPR performed best for DBP. Moreover, R scored 0.74 and 0.68 for SBP and DBP, respectively, which means that there is a strong correlation with the predictors and the ground truth.

#### 4. Wearable real-time heart attack detection

Myocardial infarction (MI), commonly known as heart attack, is considered one among the reasons of human death and disability worldwide. Although a heart attack is life threatening, it has early symptoms and signs that could be measured to help save many lives and to avoid consequences. There are three types of a heart attack: STEMI, NSTEMI, and coronary spasm, refer **Figure 12**.

Several ECG ambulatory monitoring systems are available, but they require time and effort to acquire ECG signals from patients. Furthermore, the ECG data have to be sent to professionals for diagnostic analysis. A wearable ECG device can be a solution to acquire real-time ECG signals from the patient. Then, signals are interpreted using built-in algorithms to observe abnormality events of ECG signals and sudden heart attack. Therefore, proper action can be made.

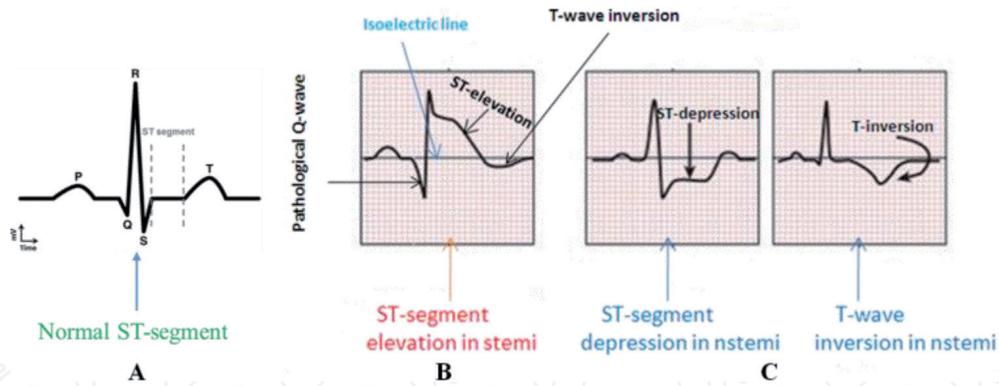
In terms of algorithm, some systems used linear classification and support vector machine (SVM). Some research studies have used the PhysioBank MI database to develop SVM-based real-time MI detection system with a classification accuracy of 90%. Another work came up with the solution for the motion artifact correction for the two-electrode ECG system with an optimized adaptive filter.

In this section, a wearable heart attack detector and alarming system is discussed, which can be used to reduce road accidents that might result from the driver being precipitated by a heart attack, details of which are present in [11]. A portable wearable system that can continuously monitor the heart condition for any early symptoms could inform the patient to pull over the vehicle safely before losing consciousness to avoid potentially fatal consequences. Moreover, medical caregivers could arrive and provide the necessary medical procedures that rescue the consequences of a heart attack.

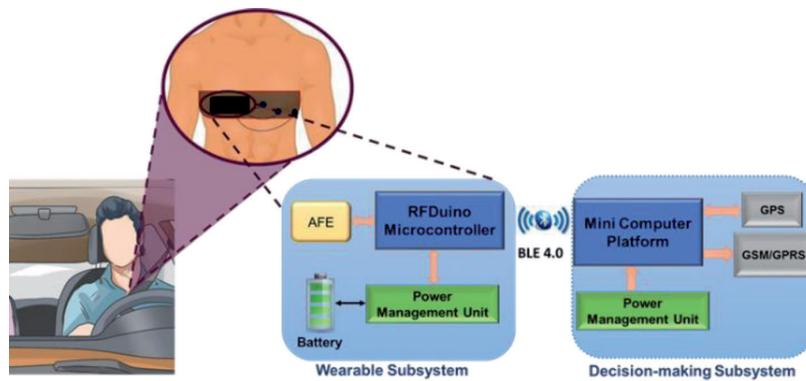
##### 4.1 Wearable system prototype description

The real-time heart attack monitoring wearable prototype system consists of two subsystems as shown in **Figure 13**. The subsystems communicate wirelessly using Bluetooth low energy (BLE) technology.

A wearable sensor subsystem is responsible for real-time ECG signal acquisition, amplification, filtering, digitization, and wireless transmission. The device is attached to a chest belt to be worn by a driver as shown in **Figure 13**. The device



**Figure 12.** Comparison of the ST segment variations in a normal subject (a) and in MI patients with STEMI (b) and NSTEMI (c).



**Figure 13.** Block diagram of the wearable heart attack detection prototype.

includes three dry electrodes (one electrode for reference and two electrodes for differential acquisition) connected to an analog front end (AFE) and an RFduino microcontroller with an embedded BLE module. Furthermore, the designed module incorporates several features to enhance the overall performance including lead-off detection, single-supply operation, adjustable gain control, rail-to-rail output, a three-pole adjustable low-pass filter (LPF), a two-pole adjustable high-pass filter (HPF), and an integrated right-leg drive (RLD). A notch filter in Wien bridge configuration was used to remove the line frequency (50 Hz) from the measured ECG signal. The acquired and preprocessed ECG signal was sent to the decision-making system using BLE.

The intelligent heart attack detection and warning subsystems are trained using the Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIH) ST Change Database [12]. This subsystem consists of a single board computer, Raspberry Pi 3 (RPi3). The acquired ECG signal from RFduino is delivered to RPi3 over the Bluetooth low energy (BLE) interface. ECG data are buffered for 10 s, and the baseline drift is corrected, segmented to ECG beats (one ECG trace), and smoothed using a digital filter. Furthermore, the subsystem incorporates SIM 908 global system for mobile communications (GSM) and the global positioning system (GPS) module; the module is directly connected to RPi3 using a low power shield.

Twenty-two different machine learning (ML) algorithms were trained for real-time ECG classification. Experiments were done in two stages. In the first stage, the public labeled database “MIT-BIH ST Change Database” covers both normal and abnormal subjects. Twenty-eight subjects’ ECG recordings with 14 subjects’ ECG recordings were normal, seven patients had T-inversion, and the other patients

experienced long-term recordings that showed ST elevation and depression. The system was tested on normal subjects, and an ECG simulator was used to simulate abnormal ST-elevated MI situations to test the performance of the complete system in real time.

## 4.2 Preprocessing

Due to body movements, even respiration, muscle signal, and power line interference of 50 Hz, filtering stage was essential to filter the signal and eliminate the inherent noises. In addition, ECG was often affected by artifacts constituted through electrodes or the interference of the signal processing hardware. Finite impulse response filter (FIR) based on window method was used to smooth the noisy signal by slicing the array of data into selected length windows, computing averages of the data within that range, and maintaining the process throughout the dataset using the moving-average filter.

Baseline wander correction technique and continuous wavelet transformation were used for synthesizing the ECG signal; the wavelet was able to show how the frequency component varied within certain ranges of time. The ECG signals were segmented into traces. A total of 3500 ECG traces over 28 different subjects were considered. Therefore, the total number of ECG traces considered in this study was 10,500.

## 4.3 Feature extraction

Different features were investigated from time, frequency, and time-frequency domains. **Table 2** shows the t-domain, the f-domain, and the (t, f) domain features extracted.

## 4.4 Training process

A five-fold cross-validation technique was used in the testing and training process. Weighted performance matrix was created by averaging the results of all the iterations. The averages of all scoring matrices were calculated for the four iterations along with their standard deviation. Performance evaluations of three different time-frequency distributions (TFDs) were calculated to identify which TFD produced higher accuracy. In addition, the receiver operating characteristic (ROC) analysis was used to evaluate the performance of the ML algorithm for classification confusion matrix and several standard statistical evaluation parameters were used to evaluate the performance of the algorithms including sensitivity, false positive rate (FPR),  $F_1$  score, and accuracy.

Time domain features	Frequency domain features	Time-frequency domain features
<ul style="list-style-type: none"> <li>• Mean</li> <li>• Variance</li> <li>• Skewness</li> <li>• Kurtosis</li> <li>• Coefficient of variance</li> </ul>	<ul style="list-style-type: none"> <li>• Spectral flux</li> <li>• Spectral entropy</li> <li>• Spectral flatness</li> </ul>	<ul style="list-style-type: none"> <li>• Combines all the prementioned features</li> <li>• Use quadratic time-frequency distribution (QTFD) to find joint (t, f) representation:               <ol style="list-style-type: none"> <li>a. Winger-Ville distribution (WVD)</li> <li>b. Spectrogram (SPEC)</li> <li>c. Extended modified B-distribution (EMBD)</li> </ol> </li> </ul>

**Table 2.**  
*Features extracted from the ECG traces.*

## 4.5 Results

All the features listed in **Table 2** were extracted to evaluate the performance of each feature. AUC were calculated as shown in **Table 3**. All features that scored a minimum of 0.5 and above were useful to apply to the classifiers. It was noticed that all the selected features fulfilled the requirement for all of the distributions. Hence, all the features were used for training, validation, and testing. A five-fold cross-validation was used to compute the ML algorithms validation accuracy.

**Tables 4** and **5** demonstrate the accuracies resulting from classifying ST elevation and T-wave inversion for the three different (t, f) distributions. Also, both tables showed the two best performing algorithms, the support vector machine (SVM) and the k-nearest neighbors (KNNs). It was noticed that SVM outperformed KNN for the extended modified B distribution.

As shown from **Tables 4** and **5**, EMBD outperformed the others in classifying ST elevation and T-wave inversion. Moreover, the variance showed that the variation for different iterations was at the minimum for the EMBD distribution (**Table 3**). Thus, EMBD distribution was more robust to noisy data. EMBD distributions were implemented for real-time classification using Python. The results showed that both the recall and the precision were reasonable for reliable detection, and for both positive and negative classifications. This can be observed in the F score as well.

Features	WVD		SPEC		EMBD	
	Original	Joint (t, f)	Original	Joint (t, f)	Original	Joint (t, f)
Mean	0.53	0.67	0.53	0.75	0.53	0.75
Variance	0.51	0.71	0.51	0.66	0.51	0.58
Skewness	0.68	0.57	0.68	0.67	0.68	0.57
Kurtosis	0.75	0.57	0.75	0.69	0.75	0.56
Coefficient of variation	0.50	0.74	0.50	0.65	0.50	0.58
Spectral flux	0.76	0.82	0.76	0.83	0.76	0.55
Spectral flatness	0.54	0.71	0.54	0.79	0.54	0.68
Spectral entropy	0.81	0.71	0.81	0.65	0.81	0.59

**Table 3.**  
 Results of the ROC analysis of the t-domain, f-domain, and (t, f) domain features for the ST elevation detection.

Parameters/ML algorithms	WVD		SPEC		EMBD	
	SVM	KNN	SVM	KNN	SVM	KNN
Recall (TPR) (%)	92	89	89	90	99.1	96.7
FPR (%)	11	12	14	13	1.7	3.8
Precision (%)	89	88	86	87	98.3	96.2
F score (%)	90.5	88.9	87.5	86	98.7	97
Accuracy (%)	87.1	86.4	85.3	84.2	97.4	95.9

**Table 4.**  
 Evaluation parameters in classifying ST elevation.

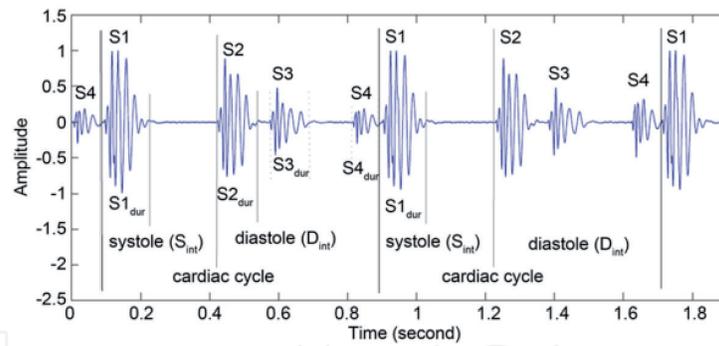
Parameters/ML algorithms	WVD		SPEC		EMBD	
	SVM	KNN	SVM	KNN	SVM	KNN
Recall (TPR) (%)	86	84	83	84	98.5	96.9
FPR (%)	19	20	22	21	1.3	4.3
Precision (%)	81	80	78	79	97.8	95.7
F score (%)	83.4	82.7	75.9	76.2	98.2	96.6
Accuracy (%)	78	76.3 <sup>*</sup>	72.1	74	96.3	95.1

**Table 5.**  
Evaluation parameters in classifying T-wave inversion.

## 5. Real-time smart-digital stethoscope system

Different methods can be used to detect cardiovascular diseases (CVD) including electrocardiogram (ECG), magnetic resonance imaging (MRI), and echocardiogram (echo). ECG is considered the most popular method because it provides an inexpensive, non-invasive, and intuitive method for heart diagnosis. However, it has a limitation when it comes to detecting structural abnormalities and defects in heart valves due to heart murmurs [13]. The previously mentioned technologies for diagnosing cardiovascular diseases could be unaffordable by the majority of the people in low-income countries to detect the CVD in advance [14–16]. Heart sound (HS) has been one of the most primitive and popular methods of detecting early cardiac illnesses with the help of abnormal heart sounds. Phonocardiogram (PCG), also known as heart sound (HS), is a graphical representation of the HS recording signals using an equipment called as phonocardiograph [17, 18]. There are three major limitations of the auscultation of the heart: first, the recorded sounds have very low amplitude which will require the device to be extremely sensitive. Second, the low-amplitude HS signal can be easily corrupted by noise leading to a faulty diagnosis. Finally, the reliability of the auscultation technique mainly depends on the skill, expertise, and capability of hearing of the doctor. Machine learning could be a solution to this problem.

The heart produces HS due to the closure of heart valves that produces the normal heart sounds. Mitral and tricuspid valve closure produces the first heart sound (“S1”), and aortic and pulmonic valve closure produces the second heart sound (“S2”) (**Figure 14**). There is no heart sound observed for normal opening of heart valve. Moreover, the flow of blood from one structure inside heart to another typically does not produce any sound. Abnormal heart sounds are caused due to the problems in heart valves or muscles or both. The third HS (S3) (**Figure 14**) is normally caused by a sudden reduction of blood supply from the left atrium to the ventricle, which is normal for children and adults (35–40 years). However, in other age groups and especially in the people above 40 years, it is abnormal [19]. The failure of heart in the diastolic period can be linked to the fourth heart sound as illustrated in **Figure 13**. Heart sounds can be categorized in terms of frequency as they differ from each other in frequency. The frequency of S1 is smaller than S2. The low-amplitude abnormal heart sounds S3, and S4 typically occurs 0.1–0.2 s after S2 and about 0.07–0.1 s before S1 respectively [19]. S1 and S2 sounds are high-amplitude signals and well identified by the diaphragm of the stethoscope. The frequency ranges of S1 and S2 are 50–60 and 80–90 Hz, respectively [19]. S3 is a low-amplitude signal with a bandwidth of 20–30 Hz [19]. S4 typically occurs at the end of diastole and can be identified easily by stethoscope, and the frequency is below 20 Hz [19]. A detailed review and state-of-the-art techniques of HS processing and classification



**Figure 14.**  
*Different heart sounds.*

are discussed in [20–22]. The abnormal characteristics of the HS signal were stated in [23], while the different signal processing techniques involved in the HS signal analysis are discussed in [24]. In [25], recent works related to feature selection and classification were discussed.

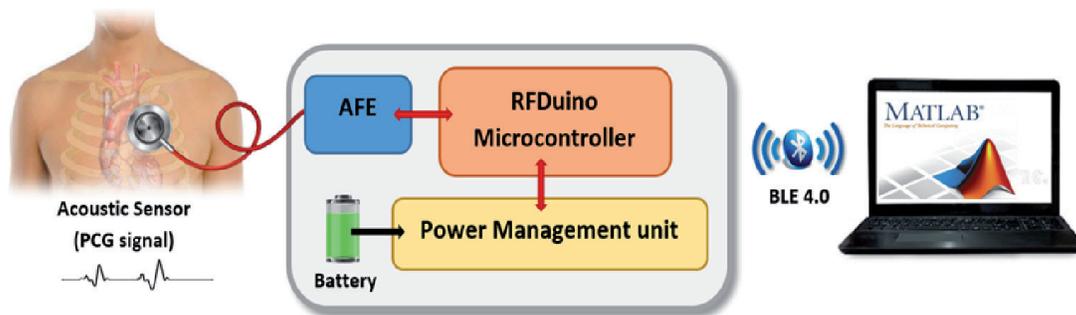
### 5.1 Wearable system

The proposed system consists of two subsystems: the two systems communicate wirelessly via BLE technology as shown in **Figure 15**, details of which are found in [26, 27]. The sensor subsystem consists of the acoustic sensor that acquires the heart sound signal and feeds to analog front end. A custom sensor was designed and implemented on a traditional stethoscope chest piece to amplify the heart sound waveform. The sensor bandwidth is 20–600 Hz to perform the conversion of the heart sound to electrical signal; the microphone was placed in the rubber tube very close to the chest piece as shown in **Figure 16**. The signal is then pre-amplified and filtered. After that, the signal is converted by ADC in the RFDuino microcontroller and transmitted wirelessly into an intelligent detection subsystem that consists of personal computer (PC) where the signal will be processed and classified using MATLAB.

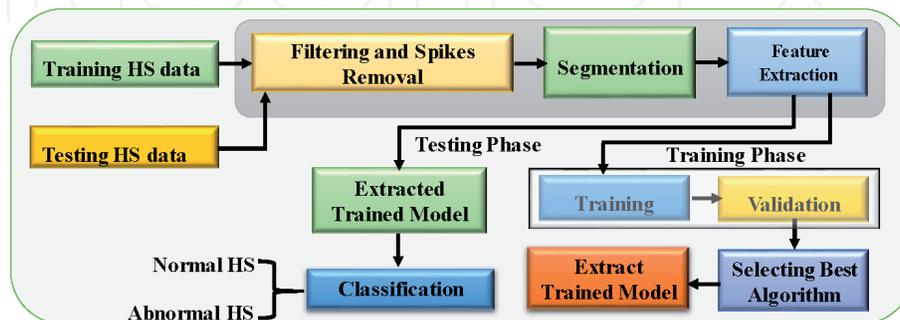
In this system, a dataset from PhysioNet challenge 2016 was utilized, which includes 3126 heart sound recordings divided into five databases (A through E); each record lasted from 5 s to just over 120 s. These HS data were recorded from clinical and nonclinical environment from both healthy and pathological patients (e.g., children and adults) from four different locations—aortic, pulmonic, tricuspid, and mitral areas. The dataset includes normal and abnormal recordings with a higher number of normal than abnormal readings.

The brain of the whole system is the intelligent abnormal heart sound and warning subsystem. It is comprised of three modules: data acquisition and logging, Bluetooth module, and classification. This subsystem acquires real-time HS signals and detects the abnormality of the signal using trained ML algorithm. Normal and abnormal heart sound data from a public database were used for training and testing of the machine learning algorithm in the MATLAB environment. The best performing algorithm was identified and ported in Python 3.5 to be used as a real-time classifier in the testing phase. The detailed block diagram of the machine learning approach can be seen in **Figure 16**.

The HS data were preprocessed and segmented into training and testing data. The algorithm to classify MI, the support vector machine (SVM) algorithm, was implemented initially using MATLAB 2018a and later on using Python 3.5 in personal computer (PC) platform for real-time classification. Signal preprocessing was accomplished using the signal processing toolbox, and training and classification



**Figure 15.**  
System overall with modified stethoscope chest.



**Figure 16.**  
Blocks of the machine learning-based abnormality detection algorithm.

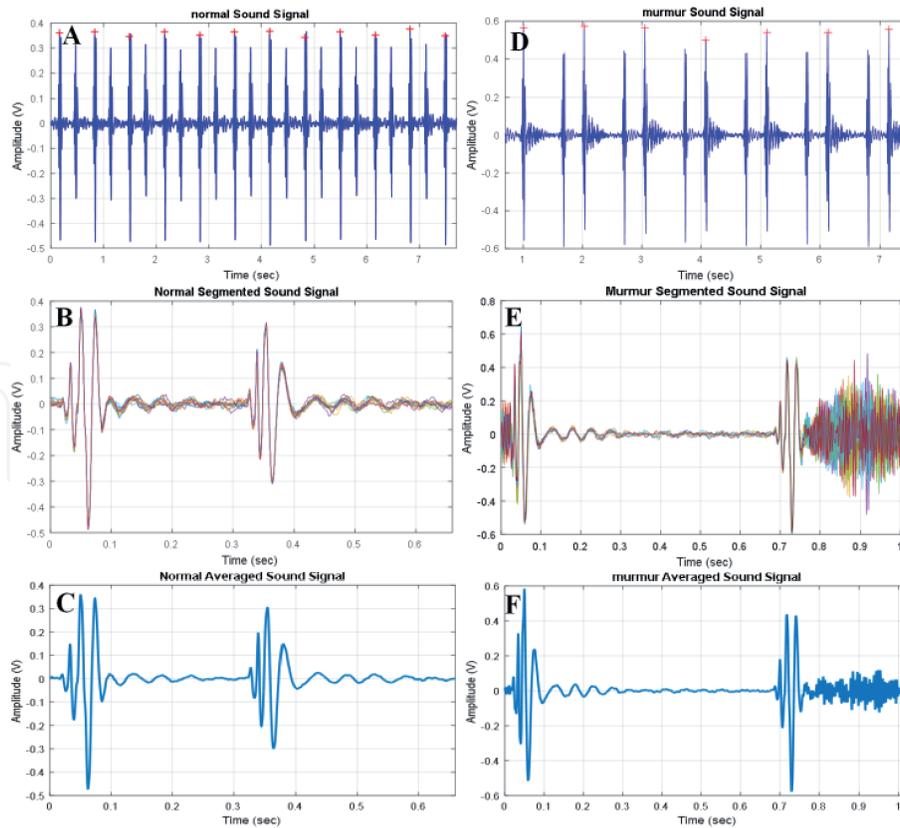
were done by Statistics and Machine Learning Toolbox in the MATLAB and using Numpy (v1.13.3), Matplotlib (v3.0.2), PyBrain (v0.31), and Scikit learn (v0.20) libraries in Python.

## 5.2 Preprocessing

In the preprocessing, a sixth-order bandpass IIR filter with lower 3-dB frequency of 20 Hz, higher 3-dB frequency of 600 Hz, and the sample rate of 2000 Hz was used to remove any high or low noise. After filtering, the signal is segmented. Segmentation of the PCG signals into heart cycles and marking of cycle starting instances are very important to generate the epoch of interest for the machine learning.

An automatic code was used to identify the S1 peaks of the PCG signal. One complete cycle of PCG signal is from one S1 peak to another S1 peak. This was segmented along with a time offset to capture the beginning of S1 and the ending of S2 as shown in **Figure 17**. In this study, MIT-BIH benchmark dataset was used and randomly partitioned into two subsets. The first set is for training and validation (80% of data), and the second set is for testing (20% of data). The ML models were trained to classify a two-class problem (normal and abnormal). For evaluation, confusion matrix and standard statistical evaluation parameters for each algorithm were calculated for each fold. After five-fold cross-validation, these parameters were used to evaluate the performance of the algorithms.

For feature extraction, 27 features encompassing t-domain, f-domain, and mel-frequency cepstral coefficient (MFCC) features were extracted for each heart sound cycle. The t-domain features were mean, median, standard deviation, signal twenty-fifth percentile, signal seventy-fifth percentile, signal interquartile range, mean absolute deviation, skewness, kurtosis, and Shannon's entropy, while the f-domain features were spectral entropy, signal magnitude at maximum frequency,



**Figure 17.** Normal and abnormal HS: (a and d) detection of peaks; (b and e) overlaid segments; (c and f) average of the segments.

the maximum frequency in the power spectrum, and ratio of signal energy between the maximum frequency range and the overall signal. Other features were mel-frequency cepstral features.

### 5.3 Feature reduction

Neighborhood component analysis (NCA) is a non-parametric and embedded method for selecting features to provide maximum prediction accuracy of classification algorithms. NCA with built-in functions in Statistics and Machine Learning Toolbox™ of MATLAB was used. It can be noted that 15 features were the most important features out of the 27 features, which were kurtosis, maximum frequency value, and all the MFCC features. In addition, the parameters for the trained model were tuned to optimize their hyperparameters. The best performing algorithms were optimized to calculate the performance measures. Statistical measures were then calculated for the testing dataset (20% of the whole database).

### 5.4 Results

Twenty-two different algorithms [three decision trees, two discriminant analyses, six support vector machines (SVMs), six k-nearest neighbors (KNNs), and five ensemble classifiers] were trained for the testing dataset with 27 features. The validation accuracy and their corresponding performance measures are listed in **Table 6**.

It is obvious from the above table that the best validation accuracy was observed for “Fine Tree” classifier. Moreover, the accuracy of classifying normal is higher

than abnormal and this is because of the imbalanced dataset. Therefore, to reduce the potential over-fitting of the features, a reduction in the number of features used in the training process. The training dataset was re-trained with the reduced number of features (15), and all evaluation parameters were calculated. **Table 7** summarizes the evaluation measures for identifying the best algorithm with feature selection.

From **Table 6**, the overall accuracy was reduced as well as classifying normal and abnormal both were also reduced even though the same algorithms were performing best in the classification after feature reduction. Therefore, it can be said that the features used for classification are optimized and cannot be reduced.

To improve the performance of the best performing algorithms by optimizing the hyperparameters of the algorithms, it was observed that the performance of the ensemble algorithm can be improved. Two important parameters were optimized for the ensemble algorithms: “Distance” and “Number of neighbors.”

Items	Fine KNN	Weighted KNN	Ensemble subspace discriminant
Accuracy (%)	94.63	93.72	93.17
Accuracy: abnormal (%)	88, 12	85, 15	87, 13
Accuracy: normal (%)	96.6, 3.4	97, 3	95, 5
Error (%)	5.37	6.28	6.83
Sensitivity (%)	96.32	95.24	95.67
Specificity (%)	89.34	88.72	85.49
Precision (%)	96.62	96.54	95.29
FPR (%)	10.66	11.28	14.51
F <sub>1</sub> score (%)	96.46	95.88	95.48
MCC (%)	85.34	82.7	81.5

**Table 6.**  
*Performance measures of three best performing algorithms for full-feature set.*

Items	Fine KNN	Weighted KNN	Ensemble subspace discriminant
Accuracy (%)	92.36	92.02	92.89
Accuracy: abnormal (%)	84, 16	82, 18	83, 17
Accuracy: normal (%)	95, 5	95, 5	96, 4
Error (%)	7.64	7.98	7.11
Sensitivity (%)	94.85	94.30	94.77
Specificity (%)	84.52	84.62	86.71
Precision (%)	95.08	95.22	95.90
FPR (%)	15.48	15.38	13.29
F <sub>1</sub> score (%)	94.96	94.76	95.33
MCC (%)	79.17	78.09	80.42

**Table 7.**  
*Performance matrix for the three best performing algorithms on reduced feature set.*

## 6. Discussion

This chapter summarizes the findings of four different applications of wearable devices to tackle four critical clinical problems. The smart wearable devices reported here can help patients in different settings to manage their diseases, which will reduce frequent hospital visit requirement and can elevate their living standard. The summary of the results from each of the case study can be provided as below.

The FSR-based smart insole can acquire high-quality vGRF for different gait cycles, and it was found that the flexible piezoelectric sensors were performing poor in calibration due to their sensitivity to 3D forces requiring special force calibration machines to control the applied force in  $x$ ,  $y$ , or  $z$  directions. Therefore, piezoelectric sensors cannot be utilized as a substitute for FSR in smart insole application.

ReliefF feature selection algorithm produced the best result when combined with Gaussian process regression (GPR) for predicting the systolic and diastolic blood pressure using PPG signal. The feature selected using a combination of ReliefF and GPR performed the best in estimating SBP, while correlation-based feature selection (CFS) and GPR performed best for DBP. It can be noted that this optimized approach can estimate SBP and DBP with the RMSE of 6.74 and 3.57, respectively.

Extended modified B distribution shows the best performance in classifying ST elevation and T-wave inversion in the heart attack detection case study using ECG signals. The variance of the results from EMBD technique showed that the variation for different iterations was at the minimum for the EMBD distribution (**Table 3**). Thus, EMBD distribution was more robust in heart attack detection in case of noisy ECG data.

Heart sound signals can be accurately classified using “Fine Tree” classifier compared with 22 different algorithms [three decision trees, two discriminant analyses, six support vector machines (SVMs), six k-nearest neighbors (KNNs), and five ensemble classifiers]. Feature reduction technique did not help in improving the performance. It was observed that the best performing algorithms can be improved further by optimizing the hyperparameters of the algorithms.

## 7. Conclusion

The chapter discusses in detail with case studies about the different opportunities available in the design and development of wearable medical devices, which can help in real-time healthcare monitoring. The chapter has not only discussed on the study of characterizing sensors for specific wearable device but also how reliable dataset can be utilized to develop trained model for medical diagnosis. It has also shown how to design and develop a complete wearable device with real-time monitoring and alarming in case of emergency. These smart wearable solutions, if properly designed and deployed, can help millions of users to take advantages of the wearable technologies and thereby can monitor their health status in different setting.

## Acknowledgements

This work was made possible by NPRP12S-0227-190164 from the Qatar National Research Fund, a member of Qatar Foundation, Doha, Qatar. The statements made herein are solely the responsibility of the authors.

## **Conflict of interest**

The authors declare no conflict of interest.

IntechOpen

## **Author details**

Muhammad E.H. Chowdhury<sup>1\*</sup>, Amith Khandakar<sup>1,2</sup>, Yazan Qiblawey<sup>1</sup>,  
Mamun Bin Ibne Reaz<sup>2</sup>, Mohammad Tariqul Islam<sup>2</sup> and Farid Touati<sup>1</sup>

1 Department of Electrical Engineering, Qatar University, Doha, Qatar

2 Department of Electrical, Electronic and Systems Engineering, Universiti  
Kebangsaan Malaysia, Bangi, Selangor, Malaysia

\*Address all correspondence to: [mchowdhury@qu.edu.qa](mailto:mchowdhury@qu.edu.qa)

## **IntechOpen**

© 2020 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

## References

- [1] Akkaş MA, Sokullu R, Çetin HE. Healthcare and patient monitoring using IoT. Internet of Things. 2020;**100173**
- [2] Mahajan A, Pottie G, Kaiser W. Transformation in healthcare by wearable devices for diagnostics and guidance of treatment. ACM Transactions on Computing for Healthcare. 2020;**1(1)**:1-12
- [3] Tao W, Liu T, Zheng R, Feng H. Gait analysis using wearable sensors. Sensors. 2012;**12(2)**:2255-2283. DOI: 10.3390/s120202255
- [4] Huang E, Sharp MT, Osborn E, MacLellan A, Mlynash M, Kemp S, et al. Abstract TP173: Feasibility and utility of home-based gait analysis using body-worn sensors. Stroke. 2019;**50**(Suppl 1): ATP173
- [5] Tahir AM, Chowdhury MEH, Khandakar A, Al-Hamouz S, Abdalla M, Awadallah S, et al. A systematic approach to the design and characterization of a smart insole for detecting vertical ground reaction force (vGRF) in gait analysis. Sensors. 2020;**20(4)**:957. DOI: 10.3390/s20040957
- [6] Toğaçar M, Ergen B, Cömert Z. A deep feature learning model for pneumonia detection applying a combination of mRMR feature selection and machine learning models. IRBM. 2019. DOI: 10.1016/j.irbm.2019.10.006
- [7] Aho VP. Insole Energy Harvesting from Human Movement Using Piezoelectric Generators [thesis]. Tampere University of Technology; 2018
- [8] Zhu H, Wertsch JJ, Harris GF, Loftsgaarden JD, Price MB. Foot pressure distribution during walking and shuffling. Archives of Physical Medicine and Rehabilitation. 1991;**72(6)**:390-397
- [9] Chowdhury MH, Shuzan MNI, Chowdhury MEH, Mahbub ZB, Uddin MM, Khandakar A, et al. Estimating blood pressure from photoplethysmogram signal and demographic features using machine learning techniques. Sensors. 2020;**20(11)**:3127. DOI: 10.3390/s20113127
- [10] Liang GLY, Chen Z, Elgendi M. PPG-BP Database [Internet]. 2018. Available from: [https://figshare.com/articles/PPG-BP\\_Database\\_zip/5459299/](https://figshare.com/articles/PPG-BP_Database_zip/5459299/) [Accessed: 21 October 2019]
- [11] Chowdhury MEH, Alzoubi K, Khandakar A, Khallifa R, Abouhasera R, Koubaa S, et al. Wearable real-time heart attack detection and warning system to reduce road accidents. Sensors. 2019;**19(12)**:2780. DOI: 10.3390/s19122780
- [12] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al., PhysioBank, PhysioToolkit, and PhysioNet. Circulation. 2000;**101(23)**:e215–e220. DOI: 10.1161/01.CIR.101.23.e215
- [13] Reed TR, Reed NE, Fritzson P. Heart sound analysis for symptom detection and computer-aided diagnosis. Simulation Modelling Practice and Theory. 2004;**12(2)**:129-146. DOI: 10.1016/j.simpat.2003.11.005
- [14] Kim RJ, Wu E, Rafael A, Chen E-L, Parker MA, Simonetti O, et al. The use of contrast-enhanced magnetic resonance imaging to identify reversible myocardial dysfunction. New England Journal of Medicine. 2000;**343(20)**:1445-1453. DOI: 10.1056/NEJM200011163432003
- [15] Rad MZ, Ghuchani SR, Bahaadinbeigy K, Khalilzadeh MM. Real time recognition of heart attack in a

- smart phone. *Acta Informatica Medica*. 2015;**23**(3):151-154. DOI: 10.5455/aim.2015.23.151-154
- [16] Gaziano TA, Bitton A, Anand S, Abrahams-Gessel S, Murphy A. Growing epidemic of coronary heart disease in low- and middle-income countries. *Current Problems in Cardiology*. 2010;**35**(2):72-115. DOI: 10.1016/j.cpcardiol.2009.10.002
- [17] Roy JK, Roy TS, Mukhopadhyay SC. Heart sound: Detection and analytical approach towards diseases. In: Mukhopadhyay SC, Jayasundera KP, Postolache OA, editors. *Modern Sensing Technologies*. Cham: Springer International Publishing; 2019. pp. 103-145. DOI: 10.1007/978-3-319-99540-3\_7
- [18] Lindsay T. *Medical Conditions as a Contributing Factor in Crash Causation*. University of Adelaide. 2018
- [19] de Lima Hedayioglu F. *Heart Sound Segmentation for Digital Stethoscope Integration* [thesis]. University of Porto; 2011
- [20] Leng S, Tan RS, Chai KTC, Wang C, Ghista D, Zhong L. The electronic stethoscope. *Biomedical Engineering*. 2015;**14**(1):66. DOI: 10.1186/s12938-015-0056-y
- [21] Gupta CN, Palaniappan R, Swaminathan S, Krishnan SM. Neural network classification of homomorphic segmented heart sounds. *Applied Soft Computing*. 2007;**7**(1):286-297. DOI: 10.1016/j.asoc.2005.06.006
- [22] Noponen A-L, Lukkarinen S, Angerla A, Sepponen R. Phono-spectrographic analysis of heart murmur in children. *BMC Pediatrics*. 2007;**7**:23. DOI: 10.1186/1471-2431-7-23
- [23] Shen C-H. *Acoustic Based Condition Monitoring* [thesis]. University of Akron; 2012
- [24] Abbas AK, Bassam R. *Phonocardiography Signal Processing. Synthesis Lectures on Biomedical Engineering*. Morgan & Claypool Publishers. 2009;**4**(1):1-194. DOI: 10.2200/S00187ED1V01Y200904BME031
- [25] Liu C, Springer D, Li Q, Moody B, Juan RA, Chorro FJ, et al. An open access database for the evaluation of heart sound algorithms. *Physiological Measurement*. 2016;**37**(12):2181-2213. DOI: 10.1088/0967-3334/37/12/2181
- [26] Chowdhury MEH, Khandakar A, Alzoubi K, Mansoor S, Tahir A, Reaz MBI, et al. Real-time smart-digital stethoscope system for heart diseases monitoring. *Sensors*. 2019;**19**:2781. DOI: 10.3390/s19122781
- [27] Ng CL, Reaz MBI, Chowdhury MEH. A low noise capacitive electromyography monitoring system for remote healthcare applications. *IEEE Sensors Journal*. 2019;**20**(6):3333-3342. DOI: 10.1109/JSEN.2019.2957068