Features for Detection of Heart Arrhythmias

Tony Basil

A Thesis Submitted to
Indian Institute of Technology Hyderabad
In Partial Fulfillment of the Requirements for
The Degree of Master of Technology

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Department of Computer Science and Engineering

July, 2013

Declaration

I declare that this written submission represents my ideas in my own words, and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the Institute and can also evoke penal action from the sources that have thus not been properly cited, or from whom proper permission has not been taken when needed.

(Signature)

(Tony Basil)

CS11M10

(Roll No.)

Approval Sheet

This thesis entitled Features for Detection of Heart Arrhythmias by Tony Basil is approved for the degree of Master of Technology from IIT Hyderabad.

C. Krishna Mohan
8/7/2013
C. KRISHNA MOHAN
Supervisor

K S Murty
K. Sri Rama Murty
External Examiner

T. Bheemarjuna Reddy
Internal Examiner

Chandra K Lakshminarayanan
external Supervising Adviser
07/09/2013

M.V. Panduranga Rao
chairman

Acknowledgements

Many individuals contributed in many different ways to the completion of this thesis. I am deeply grateful for their support, and thankful for the unique chances they offered me.

I am greatly thankful to my co-advisers Dr. C. Krishna Mohan and Dr. Choudur K. Lakshminarayan for their valuable guidance, constant encouragement and timely suggestions. I would like to make a special mention of the excellent facility provided by my institute, IIT Hyderabad.

I sincerely thank my thesis committee members Dr. Sri Rama Murty, Dr. T. Bheemarjuna Reddy, and Dr. M.V. Panduranga Rao for their interest in evaluating the progress of my research work.

I sincerely thank Jitendra Kumar Dhiman, Bollepalli S. Chandra and Mettu Srinivas for their help in conducting experimental studies during the research work.

Abstract

Cardiovascular diseases (CVD) are a leading cause of unnecessary hospital admissions as well as fatalities placing an immense burden on the healthcare industry. A process to provide timely intervention can reduce the morbidity rate as well as control rising costs. Patients with cardiovascular diseases require quick intervention. Towards that end, automated detection of abnormal heartbeats captured by electronic cardiogram (ECG) signals is vital. While cardiologists can identify different heartbeat morphologies quite accurately among different patients, the manual evaluation is tedious and time consuming. As the ability to capture and store vast amounts of data in the cloud is becoming affordable and commonplace, automated pattern recognition tools can be applied to detect occurrence of complex events in human bio-signals. Bio-signals such as the heart rhythms exhibit complex dynamics and therefore extracting pertinent features is critical to downstream pattern recognition. Prior researches placed a significant emphasis on feature extraction and selection. In this thesis, we addressed several aspects of automatic detection, most notably, feature extraction, and a careful assessment of two types of arrhythmia, ventricular ectopic beats (VEB), and supraventricular ectopic beats (SVEB). We utilized a feature selection methodology which selects the most significant features from a feature set based on time, and frequency domains. Furthermore, we evaluated the performance of several classifiers; Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and Artificial Neural Networks (ANN). The methodology and framework we proposed is fairly automated and does not require the intervention of the domain expert (cardiologist) in the extraction of the features and labeling of the segments of the ECG signals. Our results bear testimony to the improvements as seen in the commonly used metrics such as classifier sensitivity, specificity, F-measure, and positive predictive value in the evaluation of classifier performance for detecting arrhythmias.

Contents

Chapter 1

Introduction

Cardiovascular diseases (CVD) are a leading cause of fatality representing 30% of all global deaths [19]. In 2008, an estimated 17.3 million individuals died of cardiovascular diseases. Third world countries account for 80% of CVD related deaths. In 2010, CVD related illnesses cost the United States healthcare industry $316.4 billion. A large number of admissions to hospitals are unnecessary and avoidable. Due to inadequate preventive measures, CVD related fatalities continue to rise. It is imperative that we find a solution that reduces these fatalities. One way is to identify high risk patients is using simple and inexpensive tools. An automated system that can identify potential risks of patients can aid optimizing the usage of medical resources. Such systems must be able to identify patterns in cardiovascular activity that can pose a threat to the patients. Furthermore, in rural areas, where access to healthcare facilities is poor, early detection systems can be potentially lifesaving and cost effective. Electrocardiogram (ECG) is a widely used device to monitor heart function irregularities. At present, an expert cardiologist analyzes ECG plots to detect abnormalities. However, such an analysis is over short durations of an ECG signal. Since certain kinds of heartbeat arrhythmias are time consuming to detect, the patient may require long term monitoring. In health care, patients with heart problems require quick responsiveness in a clinical setting or in the operating theatre. Towards that end, automated classification of heartbeats is vital as some heartbeat irregularities are time consuming to detect. Therefore, analysis of electro-cardiogram (ECG) signals is an active area of research.

In this thesis, we propose techniques to detect two types of heartbeat arrhythmias – Ventricular Ectopic Beats (VEB) and Supra Ventricular Ectopic Beats (SVEB). We propose new features from the time and frequency domains and furthermore, a data normalization technique to reduce inter-patient and intra-patient variations. Our results are comparable to those reported in existing literature and in most cases deliver improved performance.

Chapter 2

Related Work

[1]Classification of ECG signals is a very difficult problem and current research is focused on careful extraction of heartbeat features. A major problem encountered by machine learning techniques for classifying ECG signals is due to the large inter-patient and intra-patient variability in the timing profiles and morphology of damaged cardio-vascular processes (see Figure 2.1). The effect of this behavior is that classifiers trained using traditional methods fail when applied to new patients.

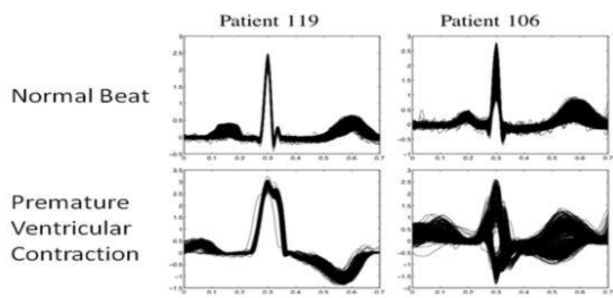

Figure 2.1: Example of heartbeat shapes from the MIT-BIH data set. Each column represents a patient and each row the beats for that specific class. Note the variations in the beat morphology across patients as well as within a patient (Source Alvarado et al [9])

Several approaches have been proposed. Hu et al [2] used a mixture of experts approach which combines global and local classifiers. The global classifier uses heartbeat signatures from a vast collection of labeled data. The local classifier trains on patient-specific ECG recordings. A gating function weights the classification results of the global and local experts and combines them to make a decision. Chazal et al [4] proposed an approach based on heartbeat morphology features, heartbeat interval features, and RR interval features and utilizes the linear discriminant classifier (LDA) for classification. Ince et al [5] proposed a patient-specific classification methodology for accurate classification[i] of heartbeat patterns.

[1]Source Alvarado et al [9]

Wiens et al [11] proposed active learning which reduces the amount of patient-specific (annotated) data needed for classification. It is therefore evident that heartbeat classification researches have focused exclusively on patient-specific ECG signatures and features.

Chapter 3

Classifiers

3.1 Linear Discriminant Analysis

Several classifiers have been used in existing literature, such as Linear Discriminant classifier [4], Neural Networks [16], Support Vector Machine (SVM) [11] etc. We implemented Linear Discriminant classifier described in Chazal et al [4]. A Linear Discriminant classifier assumes that the underlying probability density function of the data or feature vector is Gaussian. By calculating the posterior probability of class membership of a new example, Linear Discriminant classifier classifies the example into one of the k classes, $k \in \{1,2 \ldots .c\}$.The classifier choses the class with highest posterior probability. The posterior probability of an example x belonging to class k is given by

$$P(k|x) = \frac{P(x|k)P(k)}{P(x)}$$

where $P(k)$ is the prior distribution for class k, calculated as the proportion of heartbeats of class k, $P(x|k)$ is the class conditional distribution of feature vector x. Under the assumption that the data is Gaussian (i.e. Feature vector x follow Gaussian distribution), $P(x|k)$ is given by

$$P(x|k) = \frac{1}{|\Sigma|^{\frac{1}{2}}(2\pi)^{\frac{c}{2}}} \exp(-\frac{1}{2}\,(x-\mu_k)^T \Sigma_p^{-1}(x-\mu_k)$$

where μ_k is the mean of class k and Σ_p is the pooled covariance matrix given by

$$\Sigma_p = \frac{(n_1-1)\Sigma_1 + (n2-1)\Sigma_2 \ldots\ldots + (n_c-1)\Sigma_c}{n_1+n_2+\ldots\ldots n_c - c}$$

where n_k is the number of heartbeats of class k, $k \in \{1,2 \ldots .c\}$ in the training set. Covariance of each class was weighted as per the procedure described in Chazal et al [4]. The weights for classes N, SVEB, VEB, F and Q was set to $\frac{400}{n_k}$, where n_k is the number of heartbeats of class k in the training set. The classes are weighted in order to reduce the contribution of dominant classes (e.g. Class N) to the likelihood function. If the number of beats is less than 400, no weighting is applied to that class (e.g. Class Q). The weights for classes N, SVEB, VEB, F and Q are 400/45868, 400/942, 400/3787, 400/415 and 1 respectively. The final classification decision is based on the highest posterior probability $P(k|x)$. The decision function $d_i(x)$ is given by:

$\underset{i}{argmax}\, d_i(x)$ where

$$d_i(x) = -\frac{c}{2}\log(2\pi) - \frac{1}{2}\log\left|\Sigma_p\right| - \frac{1}{2}(x-\mu_k)^T \Sigma_p^{-1}(x-\mu_k) + \log(P_k)$$

and P_k is the prior probability for class k. μ_k and Σ_p are calculated from the training data. A new testing example x is classified to the class with highest $d_i(x)$. P_k for classes N, SVEB, VEB and F were set to 10/41. Prior probability for class Q was set to 1/41 due to its rare occurrence.

For description and details of Quadratic Discriminant Analysis (QDA), refer [12]. Quadratic Discriminant Analysis (QDA) differs from Linear Discriminant Analysis (LDA) in that the pooled covariance matrix is replaced by within class covariance matrices.

3.2 Artificial Neural Networks

By definition a neural network is a massively distributed parallel processor that acquires knowledge by optimizing the inter-neuron synaptic strengths by a learning process. The learning process consists of an algorithm that systematically modifies the synaptic strengths to achieve the desired objective. The architecture of a neural network is a simplified analogue of the network of neurons within a human brain responsible for memory, pattern recognition and a host of other activities. The computational elements that comprise a neural

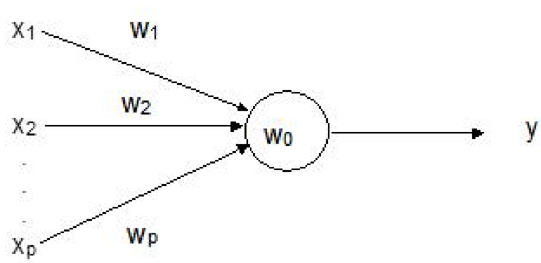

Figure 3.1: Single layered feedforward network

network, akin to the neuron are known as nodes, units, or processing elements. A neural network is an arrangement of neurons organized in layers. The simplest neural network is a single layer network consisting of neurons in the input layer feeding into an output layer. A schematic of a single layer *feedforward* network is given in Figure 3.1. The cumulative effect of the input layer nodes at the output node is a sum of the product of input values and their corresponding weights (synaptic strengths) plus a bias term.

It is feedforward in that the connections between inputs and outputs are in one (forward) direction only. A natural extension is a *multi-layered* feedforward network which consists of an input layer of neurons, a hidden layer of neurons feeding into an output layer of neurons. The number of hidden layers can be greater than one. A schematic of a fully connected, multi-layered feedforward neural network with one hidden layer is illustrated in Figure 3.2.

3.2.1 The back propagation algorithm

The goal of a neural network algorithm is to establish a relationship between the inputs and their corresponding responses. Neural networks are chosen often when it is difficult to mathematically express a relationship between the inputs (signatures) and the outputs (classes). We will, in the following, formally define the constituents and the mechanics of the back propagation network that is usually effective in establishing a relationship or a mapping between the input signatures and the output classes. Suppose we have a set of data of p input-output pairs (the input-output pairs correspond to a vector of electrical parameter measurements and the probable cause of failure), $(x_1, y_1), (x_2, y_2), \ldots, (x_p, y_p)$. The input-output pairs are such that $x \in \Re^p$, and $y \in \Re^K$. Back propagation algorithm is used to train the neural networks to develop an approximate relationship $y \simeq f(x)$. Training the neural network corresponds to finding the appropriate set of weights also known as synaptic strengths. The back propagation method usually achieves this objective given a sufficient number of training samples and correctly chosen input-output pairs. A detailed discussion of back propagation algorithm is given in [15]. Back propagation algorithm is a method to minimize the objective function $E_p = \frac{1}{2}\sum_{j=1}^{p}(d_{pj} - o_{pj})^2$, where E_p is the error due to the p^{th} signature vector, d_{pj} is the desired value for the j^{th} output neuron, and o_{pj} is the actual output of the j^{th} output neuron. We observe that each term in the sum is the contribution to the total error from a single output neuron. The minimization of E_p is achieved by finding its derivative with respect to the weights w_{ji} and computing the changes in weights in the direction of its negative gradient, in other words: $\Delta_p w_{ji} \alpha - \frac{\partial E_p}{\partial w_{ji}}$, where w_{ji} is the weight connecting the j^{th} node in layer l and the i^{th} node in layer l-1.

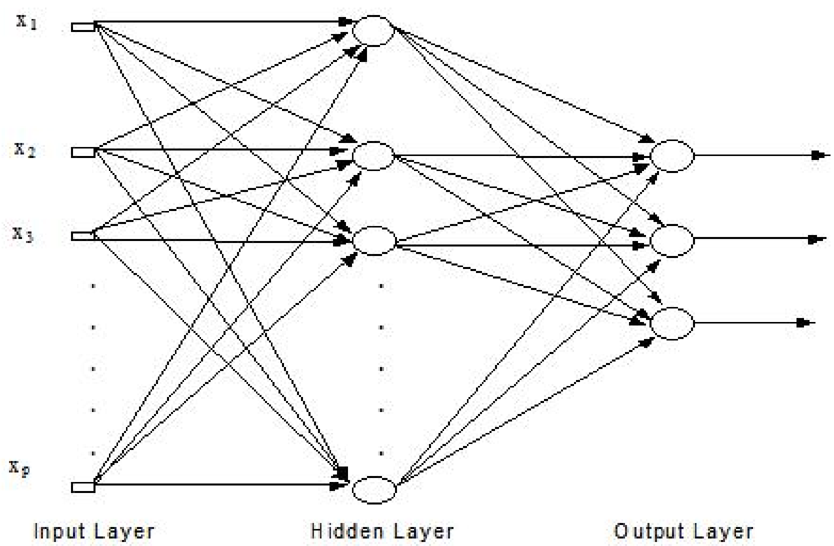

Figure 3.2: Architecture of a multi-layer feed forward neural network

3.2.2 Training with back propagation

The back propagation algorithm uses an iterative optimization technique to determine the best set of hidden-layer and output layer weights. Given an initial set of hidden-layer weights ($w_{ji}^h, (j = 1,,....,N_h)$), and an initial set of output-layer weights ($w_{kj}^o, (k = 1,,....,N_o)$), generated from the Gaussian or the uniform distribution, the method updates the weights by minimizing the sums of square of error. At each iteration,

- A multi layered feed-forward neural network with one hidden layer is preferable
- The elements of the input vector are chosen after a careful review of important input features.
- The data is scaled such that the domain of any given input variable is the unit interval (0,1).
- Sufficient training data in each labeled class is available prior to training.

• The number of hidden layer units is typically a third or half of the sum of input and output layer units.

• The hidden layer and output layer weights (synaptic strengths) are generated from a Gaussian (0,1)/Uniform(0,1) distribution. It is recommended a small value for the variance is chosen. The leaning rate η is usually chosen between (0.1-0.3).

• All the bias units may be set to zero in the entire network.

• The training data is split into two sets, namely the training set, and validation set.

• The *k-fold* cross-validation may be applied for training and validation.

• The training set is used to train several preliminary network architectures

• The validation set is used to identify the network with the least sums of squares of error.

• The network is trained using the pattern by pattern approach. The patterns are selected randomly from each class to eliminate any biases during learning.

Chapter 4

Feature Extraction and Data Description

4.1 Data Description

This In an ECG signal, heartbeat patterns are identified by a cardiac cycle consisting of P-QRS-T waveforms. The P-QRS-T waveforms represent 5 successive deflections in amplitude, known as P, Q, R, S and T waves as shown in Figure 4.1. These patterns tend to vary within a patient recording resulting in intra-patient variations. In addition to intra-patient variations, these patterns exhibit inter-patient variations. This makes heartbeat classification a challenging problem. To effectively classify a heartbeat, a classifier must be able to take into account both inter-patient and intra-patient variations in ECG signal. Figure 2.1 shows the inter-patient variation of heartbeat pattern for patient 119 and 106 relative to normal beats, and intra-patient variations relative to premature ventricular contractions.

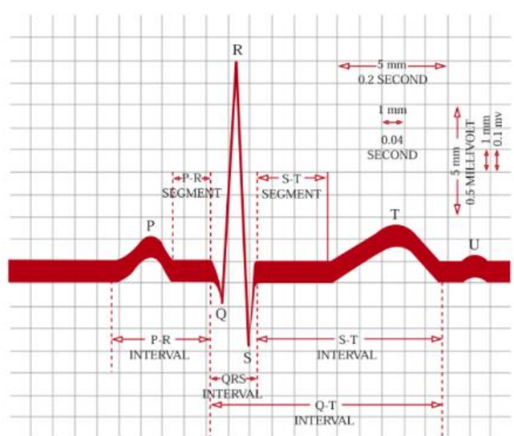

Figure 4.1: Cardiac cycle of a typical heartbeat represented by the P-QRS-T waveform

In order to compare our results with the existing literature, we used MIT/Beth Israel Hospital (BIH) Arrhythmia Database available in PhysioBank archives [1]. The database includes 48 Electrocardiogram (ECG) recordings sampled at 360 Hz for a duration of half hour. The 48 ECG recordings were obtained from 47 subjects. Each ECG recording consists of two ECG lead signals, Lead A and Lead B. Our study was conducted entirely on Lead A. ECG recording is susceptible to noise such as power line interference and baseline wander. Before the feature extraction, we preprocessed the ECG signal to reduce the baseline wander and 60 Hz power line interference. To reduce baseline wander, the signal was passed through median filters of window sizes 200ms and 600ms. The first median filter removes P-waves and QRS complexes and second median filter removes the T-waves leaving behind the baseline wander. By subtracting the baseline wander from the original signal, we obtain the filtered signal. We removed power line interference using a notch filter centered at 60Hz. This type of ECG filtering is conducted in all proposed methods cited in this thesis.

The database has annotations for 20 different types of heartbeats, with each heartbeat annotated by an expert cardiologist. The annotation provides the location of the R-Peak and the corresponding heartbeat label. R-Peak represents the peak of QRS complex (See Figure 4.1) and heartbeat label represents the type of heartbeat. In accordance with the American Association of Medical Instrumentation (AAMI) protocol [3], heartbeats available in MIT-BIH arrhythmia database were grouped into 5 classes (See Table 4.1]). The 5 classes are Normal and bundle branch block beats (N), Supraventricular ectopic beats (SVEBs), Ventricular ectopic beats (VEBs), Fusion of normal and VEBs (F), and Unknown beats (Q). Although AAMI protocol [3] define 5 classes, the protocol defines the detection of SVEB as a binary classification problem. For the detection of SVEB, a heartbeat is classified as either SVEB or not SVEB (N, VEB, F and Q). Similarly, for the detection of VEB, the heartbeat is classified as either VEB or not VEB (N, SVEB, F and Q). The data was divided into two disjoint sets of patients DS1 and DS2, containing 22 patients each. As per the AAMI protocol [3], our study did not consider four patients with paced beats. The training dataset was derived from dataset DS1 and testing dataset was derived from dataset DS2. In other words, training set DS1 is used to train the global classifier, which is then tested on testing set DS2 containing a new set of patients. Note that our approach does not require apriori knowledge of patient specific labeled beats from the testing set, unlike certain other techniques [6,9,11] in existing literature. DS1 and DS2 contains the following recordings:

DS1 = {101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223, 230}

DS2 = {100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, 234}

Paced beats = {102, 104, 107, 217}. Note that paced beats are excluded from analysis.

Table 4.1: Heartbeat classes given by the MIT-BIH Database along with the regrouping defined by the AAMI standard. (Source Alvarado et al [9])

MIT-BIH class	MIT-BIH number	Number of Samples	AAMI groups	Number of samples.
Normal beat	1	75052	N: Beats not found in the class S, V, F and Q.	90631
Left bundle branch block beat	3	7259		
Right bundle branch block beat	2	8075		
atrial escape beats	34	16		
Nodal (junctional) escape beat	11	229		
Atrial premature beat	8	2546	S: Supraventricular ectopic beat.	2781
Aberrated atrial premature beat	4	150		
Nodal(junctional) premature beat	7	83		
Supraventricular premature beat	9	2		
Premature ventricular contraction	5	7130	V: Ventricular ectopic beat.	7236
Ventricular escape beat	10	106		
Fusion of ventricular and normal beat	6	803	Fusion beat (F)	803
Paced beat	12	7028	Q: Unknown beat.	8043
Fusion of paced and normal beat	38	982		
Unclassified beat.	13	33		

4.2 Metrics

A variety of metrics are used in the realm of classification. Adhering to common practice in heartbeat classification, we used the metrics listed below. The classification results are reported in terms of accuracy (Acc), sensitivity (Se), positive predictive value (PPV), and false positive rate (FPR). They are defined as follows:

$$Accuracy\ (ACC) = \frac{TP+TN}{TP+TN+FP+FN}, Sensitivity(Se) = \frac{TP}{TP+FN},$$

$$Positive\ Predictive\ Value\ (PPV) = \frac{TP}{TP+FP}, and\ False\ Positive\ Rate(FPR) = \frac{FP}{TN+FP}$$

where TP (True Positive) is the number of heartbeats of class $'i'$ that are correctly classified; FN (False Negative) is the number of heartbeats of class $'i'$ that are incorrectly classified to class $'j'$; FP (False Positive) is the number of beats of class $'j \neq i'$, that are incorrectly classified to class $'i'$; TN (True Negative) is the number of beats of class $'j'$ that are correctly classified. An additional metric, the F score, was used to evaluate the classification performance. F-Score is a common metric used in the field of Information Retrieval. Wiens et al [11] defined F-Score as

$$F-Score = \frac{2 * Se * PPV}{Se + PPV}$$

4.3 A Comparison of Statistical Machine Learning Techniques in Heartbeat Detection and Classification

4.3.1 Feature Extraction

We used a 13 dimensional feature vector, one for every heartbeat recorded in the 30 minute recording of each patient. The features consist of RR interval duration, the R-peak amplitude, and 5 samples each to the left and right of the R-peak, from the pre-processed ECG signal recordings down-sampled to 114 Hz. The 114 Hz sampling rate is in the vicinity of the average sample rate reported by Alvarado et al [9]. The RR interval features include the Pre-RR Interval and the Post-RR Interval. Pre-RR interval is calculated as the sample count between the current R-Peak and the preceding R-Peak, while Post-RR interval is calculated as the sample count between the current R-Peak and the next R-Peak. We settled for the 13 dimensional feature vector after it was found that having more sample values in the feature vector do not produce significant improvement in performance. It is noted that a lower sampling rate and smaller feature vector is desirable in real time monitoring applications as it translates to lower hardware complexity and power consumption.

4.3.2 Classifiers

We used methods such as linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), mixture of experts (ME), artificial neural networks (ANN) and ensemble networks. For an overview, refer to Chapter 3. In our implementation, the proposed mixture consists of two experts, LDA and QDA. The classification for each patient was performed using LDA and QDA and the winner between the two was chosen for each patient based on the F-Score, described in section 4.2. The LDA assumes that the covariance within the VEB and other generic class is the same, as opposed to the QDA which assumes an unequal covariance between the two classes. The ANN implementation consisted of network ensembles as they are known to exhibit superior performance in applications as widely reported in literature.

4.3.3 Results and Discussion

The classification scheme involved two classes consisting of feature vectors belonging to the VEB class, and the patterns from the remaining classes combined into one set {N, F, Q, S}. We have reported the results of the classification tasks in Table 4.2 and Table 4.3. Table 4.2 reports the gross results for LDA and QDA. Table 4.3 reports the gross results for ME (Mixture of experts) and ANN ensembles. Columns 2-6 contain the total number of heartbeats from each class; columns 7-14 report the gross classifier performance in terms of Acc (Accuracy), Se (Sensitivity), PPV (Positive predictive value) and FPR (False positive rate). Row 1 reports the gross result (Gross) for the 22 patient testing set DS2. Similarly, row 2 reports the gross result (Gross*) for the set of 11 patients overlapping with the testing set from Hu et al [2]. We used the aggregate TP, FN, FP, and TN to calculate the gross results for each classifier.

In the mixture of experts model (ME), the classification for each patient was performed using both LDA and QDA. For each patient, we chose the results for classifier (LDA or QDA) with higher F-Score, which was later used to calculate the gross results for the ME model. The gross statistics for the ANN ensemble classifier was calculated by taking the average of the results reported by the ANN ensemble.

Table 4.2**:** Gross classification results for LDA and QDA

Number of Heartbeats						LDA				QDA			
	N	**S**	**V**	**F**	**Q**	**Acc**	**Se**	**PPV**	**FPR**	**Acc**	**Se**	**PPV**	**FPR**
Gross	44258	1837	3221	388	7	93.4	75.8	61.9	4.8	83.1	97	35.2	18.4
Gross*	23169	203	3174	388	2	94.2	75.8	75.3	3.3	83.6	97	41.6	18.2

Gross: Gross Results for the 22 patient testing set

Gross*: Gross results for the 11 patients coming to the testing set of Hu et al [2]

Table 4.3**:** Gross classification results for Mixture of Experts (ME) and ANN Ensemble

Number of Heartbeats						LDA				QDA			
	N	**S**	**V**	**F**	**Q**	**Acc**	**Se**	**PPV**	**FPR**	**Acc**	**Se**	**PPV**	**FPR**
Gross	44258	1837	3221	388	7	95.4	92.2	63.4	4.3	96.9	79.7	74.6	1.9
Gross*	23169	203	3174	388	2	95.2	93	73.5	4.5	97.1	80.3	94.2	0.7

Gross: Gross Results for the 22 patient testing set

Gross*: Gross results for the 11 patients coming to the testing set of Hu et al [2]

The gross statistics (Gross) reported in Table 4.2 shows that LDA achieved higher Accuracy (93.4%), PPV (61.9%) and FPR (4.8%) while QDA achieved higher Sensitivity (97%). However, since QDA achieved a significantly lower PPV (35.2%) due to high false positives, LDA has clearly outperformed QDA. Comparing the ME model to the ANN ensemble in Table 4.3, the ANN ensemble clearly outperforms the ME model due to higher PPV. ME achieved high sensitivity (92.2%) and low PPV (63.4%), whereas ANN ensemble achieved a modest Sensitivity (79.7%) and PPV (74.6%). Thus after extensive testing, applying the many classifiers, the ANN stood out with the best performance, followed by mixture of experts model (ME), LDA and QDA. ANN ensemble achieved a gross accuracy (Acc) of 96.9% while LDA, QDA and mixture of experts (ME) achieved 93.4%, 83.1% and 95.4% respectively. ANN ensemble achieved sensitivity (Se) of 79.7%, while LDA, QDA and ME achieved sensitivity of 75.8%, 97%, and 92.2% respectively. However, ANN ensemble achieved the highest PPV among the four classifiers (74.6% for ANN, 61.9% for

LDA, 35.2% for QDA and 63.4% for Mixture of experts). Therefore, the performance of the ANN ensemble has been found to be consistently superior to LDA, QDA and the ME.

In Table 4.4, we compare our methods with the state of the art models. The table shows the results reported by Chazal et al [4], Hu et al [2] and Alvarado et al [9] followed by the results obtained using our approach. Column 1 identifies the study, columns 2-5 represents the gross results for the 22 patient testing set. Columns 6-9 represents the gross results reported by the various studies for the set of 11 patients overlapping with the testing set from Hu et al [2]. Column 10 represents the sampling rate. The mixture of experts model (MoE) proposed by Hu et al [2] consists of a global expert and a local expert. The experts are weighted to make a decision. Hu et al (GE) represents the results obtained using the global expert, while Hu et al (MoE) represents the results obtained using the mixture of experts model described previously.

Table 4.4**:** Comparison of results with the state of the art

Methods	**VEB Gross**				**VEB Gross***				**Sampling Rate**
	Acc	**Se**	**PPV**	**FPR**	**Acc**	**Se**	**PPV**	**FPR**	
Chazal et al[4]	97.4	77.7	81.9	1.2	96.4	77.5	90.6	1.1	360
Hu et al (GE) [2]	-	-	-	-	75.3	69.6	34.6	16.8	180
Hu et al (MoE) [2]	-	-	-	-	93.6	78.9	76	3.2	180
Alvarado et al [9]	-	92.4	94.8	0.4	-	-	-	-	117
Proposed LDA	93.4	75.8	61.9	4.8	94.2	75.8	75.3	3.3	114
Proposed QDA	83.1	97	35.2	18.4	83.6	97	41.6	18.2	114
Proposed ME	95.4	92.2	63.4	4.3	95.2	93	73.5	4.5	114
Proposed ANN Ensemble	96.9	79.7	74.6	1.9	97.1	80.3	94.2	0.7	114

We compared our results for the 22 patient testing set (VEB Gross) with the results reported by Chazal et al [4]. Sensitivity (Se) achieved by LDA, QDA, and ME are comparable to the sensitivity reported by Chazal et al [4]. However, LDA, QDA and ME achieved lower PPV. The performance of ANN ensemble(Acc equal to 96.9%, Se equal to 79.7%, PPV equal to 74.6% and FPR equal to 1.9%) is comparable to the results reported by Chazal et al [4] (Acc equal to 97.4%, Se equal to 77.7%, PPV equal to 81.9% and FPR equal to 1.2%). Note that while our sampling rate was 114 Hz, Chazal et al [4] sampled at 360 Hz. Also, we compared our results with the results reported by Alvarado et al [9] (Se equal to 92.4%, PPV equal to 94.82%, FPR equal to 0.4%). Note that Alvarado et al [9] reported an average sampling rate of 117 Hz, which is in within the range of our sampling rate. Our models were outperformed by the model proposed by Alvarado et al [9], which achieved higher sensitivity (Se) and positive predictive value (PPV). The approach by Alvarado et al [9] comes with many caveats. It may be noted that the proposal by Alvarado et al [9] using the integral and fire (IF) models derives its advantages if the analog signal is sampled as opposed to the digital signal. Hardware implementations for the IF are now not commercially available and are still in the incipient laboratory stages.

Columns 6-9 in Table 4.4 reports the gross results (VEB Gross*) achieved for the 11 patients overlapping with the testing set reported by Hu et al [2]. We have observed that ANN ensemble (Acc equal to 97.1%, Se equal to 80.3%, PPV equal to 94.2% and FPR equal to 0.7%) and our mixture of experts (ME) model (Acc equal to 95.2%, Se equal to 93%, PPV equal to 73.5% and FPR equal to 4.5%) outperforms Hu et al [2] with MoE (Acc equal to 93.6%, Se equal to 78.9%, PPV equal to 76% and FPR equal to 3.2%). The performance of proposed LDA model (Acc equal to 94.2%, Se equal to 75.8%, PPV equal to 75.3% and FPR equal to 3.3%) is comparable to the Mixture of experts model proposed by Hu et al [2]. For the same 11 patients, the results for ANN ensemble is comparable to the results reported by Chazal et al [4] (Acc of 96.4%, Se of 77.5%, PPV of 90.6% and FPR of 1.1%).

The results indicate that artificial neural networks are better suited for the detection of VEB type arrhythmia. It was observed that varying the learning rate and hidden layer nodes had minimal impact on the performance. Also, increasing the sampling rate to 180 Hz did not produce significant gain in performance. Hence a sampling rate of 114 Hz was found to provide enough discriminatory power for the classification task. In short, our approach emulated the performance of the state of the art models at a lower sampling rate and a set of simple features.

4.3.4 Summary

The main contribution of this section is to review the state of the art in classification of heartbeats using ECG recordings. This is a comprehensive study that consisted of a suite of classifiers and variants there of applied to a binary classification task of detecting VEB type arrhythmias using the MIT-BIH patient archives. By performing extensive set of experiments over a range of sampling rates and over a range of tuning parameters specific to various classifiers, we are able to tabulate and compare the performances of individual classifiers. The practitioner based on domain knowledge and comfortable with tolerances relative to detection accuracy can choose an appropriate classifier based on our findings. Our investigation suggests that a simple set of morphological features together with time and amplitude features from the P-QRS-T waveform sampled at a 114 Hz and a well-trained (offline) ensemble of neural networks can yield satisfactory results.

4.4 Detection of Classes of Heart Arrhythmias based on Heartbeat Morphology Patterns

4.4.1 Heartbeat Classes

After careful analysis of the heartbeat morphology, we observed that SVEB follows two different morphological patterns. We grouped together the beats of type N, F and Q into one class named N and split the class SVEB into subclasses S1 and S2, each representing a specific morphological pattern as seen in Figure 4.2. The fourth class is VEB. We trained the classifier to identify N, VEB and the two subclasses S1 and S2, making it a four class classification problem.

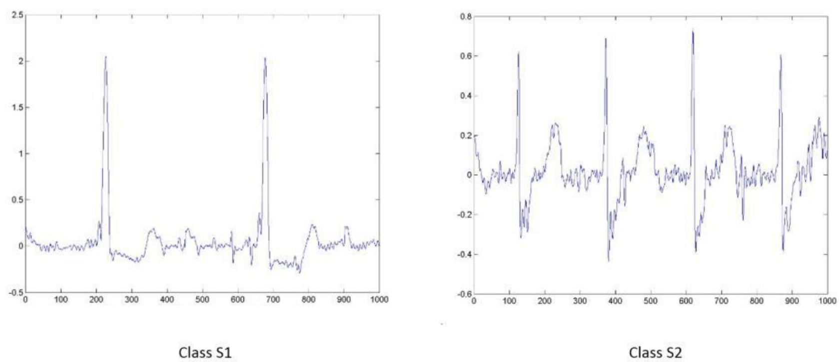

Figure 4.2: Example for variation in heartbeat pattern. Class SVEB was divided into two subclasses S1 and S2

4.4.2 Feature Extraction

Our feature vector was based on the feature set FS3 mentioned in Table VI by Chazal et al [4]. The feature set FS3 consists of RR Interval features and ECG morphology features. The RR Interval features include the Pre-RR Interval, Post-RR Interval, Average RR-Interval and Local avg. RR Interval. Pre-RR interval is time interval between the current R-Peak and the preceding R-Peak and Post-RR interval is the time interval between the current R-Peak and the next R-Peak. Average RR Interval is the average of all the RR intervals in a recording. Local avg. RR Interval is calculated as the average of 10 RR intervals surrounding a heartbeat. The ECG morphology features include fixed interval morphology features from the QRS complex and T wave of a heartbeat cycle. In order to extract the ECG morphology features, two sampling windows were used. The first window was used to extract the features from the QRS complex. The window represents the ECG signal between R-Peak - 50ms to the left of R-Peak and R-Peak + 100ms to the right of R-Peak. The QRS morphology features includes 4 samples from R-Peak - 50ms to R-Peak, and 6 samples from R-Peak to R-Peak + 100ms. The second window was used to extract features from T wave. It represents the ECG signal between R-Peak + 150ms to the right of R-Peak and R-Peak + 500ms to the right of R-Peak. 8 samples were extracted from the second window representing the T wave. The RR Interval features, QRS morphology features and T wave features are then combined to form a 22 dimensional feature vector. The feature vector does not include features from the P wave. We extracted the RR interval features and the ECG morphology features to form the 22 dimensional feature vector. The feature vector was extracted for every heartbeat in the 30 minute recording of each patient.

4.4.3 Classifier

We used a classifier based on Linear Discriminant function (See Chapter 3). We implemented the classifier proposed by Chazal et al [4], which was used later for the classification purpose. Classes N, V and SVEB (Class S1 and Class S2) were assigned equal prior probabilities as mentioned in Chazal et al [4]. The prior probabilities for class N and class V were set to 10/30 while that of S1 and S2 were set to 5/30.

4.4.4 Results and Discussion

We have reported the classification results in Table 4.5. Column 1 identifies the study, Columns 2-5 represent the number of beats from the classes N, S1, S2, and V respectively (see Section 4.1). Columns 6, 11 and 16, represent the results for the entire class SVEB (S). The results reported for SVEB represent the gross result obtained from both S1 and S2. They are used to compare our results with the results reported by Chazal et al [4]. Columns 7-10 represent the Sensitivity (Se) for each class while columns 12-15 represent the Positive Predictive Value (PPV) for each class. We performed four class classification task for detecting heartbeat arrhythmias using MIT-BIH patient archives (see Section 4.1). The four classes are N, S1, S2 and V. We trained the classifier using the training set DS1 (see Section 4.1: DS1) and tested the classifier on the testing set DS2 (see Section 4.1: DS2). Results for this configuration has been reported in row 1 of Table 4.5 (Configuration 1). The sensitivity (Se) and positive predictive value (PPV) for class S1 is 56.9% and 13.8% respectively. The sensitivity and PPV for class S2 is 20.4%and 29% respectively. The gross Sensitivity (Se) for the Class SVEB (S), which is a combination of S1 and S2 is 27.6% and Positive Predictive

Value (PPV) is 20.1%. Table 4.6 reports the confusion matrix for Configuration 1. Clearly, misclassification of S2 beats is significantly higher when compared to other classes. We observed that the training set had inadequate number of training examples from Class S2, resulting in large number of false negatives. This had significant impact on the overall classification performance. However, when we updated the training set to include the first three minutes of data from the test patient 232 (see Section 4.1), we observed significant improvement in the classification performance for class S2. The motivation for adding first three minutes of data to the training set is that the recording (Test Patient 232) contained large number of heartbeats from Class S2. The resulting training set had adequate number of training examples from class S2. With the updated training set, we achieved a Sensitivity and PPV of 80.5% and 73.8% respectively.

Table 4.5: Classification Results for Configuration 1 and Configuration 2

	Number of Beats					**Sensitivity (Se)**					**Positive Predictive Value (PPV)**				
	N	**S1**	**S2**	**V**	**S**	**N**	**S1**	**S2**	**V**	**S**	**N**	**S1**	**S2**	**V**	**S**
Config. 1	44624	362	1474	3221	1837	95.9	60.2	20.6	81.6	28.4	97.3	14	28	85.2	19.7
Config. 2	44624	362	1331	3221	1693	95.8	71.5	80.8	88.1	78.8	99.3	14.5	77.2	85.8	42
Chazal et al [4]	-	-	-	3221	1837	-	-		77.7	75.9	-	-	-	81.9	38.5

For the latter case, the gross Sensitivity for the class SVEB is 77.1%, while PPV is 41.1%. This indicates that with sufficient representation in the training set, class S1 and S2 can act as two independent subclasses within the class SVEB, while improving the classification performance. We have reported the results for the latter case in row 2 of Table 4.5 (Configuration 2). The confusion matrix for Configuration 2 has been reported in Table 4.7. By comparing with the confusion matrix of Configuration 1 (see Table 4.6), it is clear that the misclassification of S2 beats reduced significantly with Configuration 2. For class S2, Configuration 2 produced fewer false negatives (260) when compared to Configuration 1 (1172). We compared our results with the results reported by Chazal et al [4]. Row 3 in Table 4.5 present the results reported by Chazal et al [4]. Our results for SVEB (Sensitivity = 77.1, PPV = 41.1%) is comparable to the results reported by Chazal et al [4] (Sensitivity = 75.9%, PPV = 38.5%). Our results for VEB improved upon the results reported by Chazal et al [4]. While Chazal et al [4] achieved a sensitivity of 77.5% and Positive Predictive Value (PPV) of 81.9%, we achieved a Sensitivity of 87% and Positive Predictive Value (PPV) of 81.9%. Our investigation has shown that it is possible to divide the class SVEB into multiple subclasses on the basis of their morphological patterns and classify the SVEB heartbeats into one of these subclasses. This can have a non-trivial impact on the classifier performance as the classifier trains on more a coherent set of examples from each subclass, leading to better classification performance. Preliminary results show that our method emulate the results reported by Chazal et al [4].

Table 4.6 Confusion Matrix for Configuration 1

		Algorithm			
		N	S1	V	S2
Reference	N	42792	1213	374	245
	S1	51	218	25	68
	V	67	53	2628	470
	S2	1043	71	56	304

Table 4.7 Confusion Matrix for Configuration 2

		Algorithm			
		N	S1	V	S2
Reference	N	42692	1334	378	163
	S1	58	259	36	9
	V	58	177	2837	146
	S2	182	19	54	1076

4.4.5 Summary

The main contribution of the section is the study of various heartbeat patterns within the SVEB class and its impact on the classification performance. We analyzed the performance of the classifier in detecting the SVEB type of arrhythmia and compared it with the state of the art. We subdivided SVEB into two subclasses based on their morphological patterns and trained the classifier to detect these subclasses. Our investigation suggests that these subclasses accommodates for the inter-patient variations and have a non-trivial influence on the classification performance. The results reported are comparable to the state of the art.

4.5 Feature Extraction and Selection for Detecting Certain Types of Heart Arrhythmias

4.5.1 Feature Extraction

We extracted time domain features, ECG morphology features and frequency domain features from Lead A of each ECG recording. Out of the 18 features extracted, 12 features are time domain features, 2 are ECG morphology features and 3 are frequency domain features. The last feature is a flag indicating 0 or 1. Time domain features include RR Interval features, QRS duration, QR duration, RS duration and T wave duration, energy of QRS complex, energy of QR segment, energy of RS segment and energy of T wave. Energy of a signal is calculated as the sum of squares of magnitude of samples in that segment. We extracted the RR Interval features proposed by Chazal et al [4]. They are Pre-RR Interval, Post-RR Interval, Average RR-Interval and Local average RR Interval. Pre-RR interval is time interval between the R-Peak of current heartbeat and the R-Peak of preceding heartbeat. Post-RR interval is the time interval between the R-Peak of current heartbeat and the R-Peak of subsequent heartbeat. Average RR Interval is the average RR interval of all the heartbeats in a recording. Local average RR Interval is the average RR interval of 10 heartbeats surrounding a heartbeat. QRS duration is the time interval between the QRS onset and QRS offset. R duration is the time interval between the QRS onset and the R-Peak of QRS complex. RS duration is the time interval between the R-Peak of QRS complex and QRS offset. ECG morphology features include features extracted from the QRS complex; and the T wave of a heartbeat cycle. Before extracting the ECG morphology features from QRS complex, the ECG signal was down sampled to 120 Hz. Once down sampled, magnitude of 2 samples to the left of R-Peak, the magnitude of sample at R-Peak and the magnitude of 2 samples to the right of R-Peak were extracted. In total, 5 samples were extracted from the QRS complex. In order to obtain the ECG morphology features from T wave, linear interpolation was applied as in [4] before extracting the magnitudes of 9 samples representing the T. For ease of presentation, we counted the 5 features extracted from QRS complex as a single set. Similarly, 9 features extracted for T wave was treated as a single set. The frequency domain features include maximum Fourier coefficients of QRS complex, QR segment of QRS complex and RS segment of QRS complex. In addition to time domain features, ECG morphology features and Frequency domain features, we also extracted the P wave flag, which is a binary flag representing the presence or absence of P wave associated with a beat. In total, we extracted 18 different types of features from Lead A. The features were extracted for every heartbeat in the 30 minute recording of each patient. Table 4.8 is a listing of the different features used.

Table 4.8: List of features extracted from ECG signal

Group Label	Features	
Time Domain	• Pre RR Interval • Post RR Interval • Average RR Interval • Local avg. RR Interval • QRS duration • QR duration	• RS duration • T wave duration • Energy of QRS complex • Energy of QR segment • Energy of RS segment • Energy of T wave
ECG Morphology	• ECG Morphology of QRS complex (5 samples) • ECG Morphology of T wave (9 samples)	
Frequency Domain	• Max. Fourier coefficient of QR segment • Max. Fourier coefficient of RS segment • Max. Fourier coefficient of QRS complex	
Others	• P wave flag	

4.5.2 Classification Methodology

4.5.2.1 Classifier

We used Linear Discriminant classifier (See Chapter 3) throughout our study.

4.5.2.2 Feature Selection

In our study, the universe of features that we considered are given in Table 4.8. Subsets of those features are used for classification purpose. Use of all these features in training the classifier may not guarantee the best performance. During feature selection, we select the subset that maximize classification performance. A set of N features can have 2^N-1 possible feature subsets. We used Incremental Wrapper Approach [7] with forward selection algorithm, to find the best subset of features for Classifiers 1, 2 and 3 (See sections 4.5.2.4 and 4.5.2.5). We start with an empty subset. During each step, we add to the subset the feature with maximum impact on classification performance. A metric called F Score, described in section 4.2 was used to evaluate the performance of each feature. The feature that maximizes F-Score when combined with the existing feature set, is added to the subset. Though this approach does not guarantee an optimal solution, it has been proven to be reliable and efficient. We ran the algorithm to find three features sets. Classifier 1 was trained on Feature Set 1 and Classifier 2 was trained on Feature Set 2. Feature Set 3 was used to train Classifier 3. The F-Scores for classifiers 1 and 2 are computed using the sensitivity and PPV of Normal beats and similarly, F-Score for classifier 3 is computed using the sensitivity and PPV of SVEB, as our primary focus is on detecting SVEB. The feature sets have been reported in Table 4.9.

Table 4.9: Features selected for Feature set 1, Feature Set 2 and Feature Set 3

Group Label	Feature Set 1 (10 dimension)	Feature Set 2 (13 dimension)	Feature Set 3 (10 dimension)
Time Domain	• Pre RR Interval • Local Avg. RR Interval • Energy of T wave	• QRS duration • Local avg. RR Interval • Pre RR Interval • Post RR Interval • Energy of QR segment • Energy of T wave	• Normalized Pre RR Interval • Post RR Interval • T wave duration • Energy of T wave
ECG Morphology	• ECG Morphology of QRS complex (5 samples)	• ECG Morphology of QRS complex (5 samples)	• ECG Morphology of QRS complex (5 samples)
Frequency Domain	• Max. Fourier coefficient of QR Segment • Max. Fourier coefficient of RS segment	• Max. Fourier coefficient of QR Segment • Max. Fourier coefficient of RS segment	• Max. Fourier coefficient of RS segment

4.5.2.3 RR Interval Normalization

Pre RR Interval is the time duration between the R-Peaks of consecutive heartbeats, measured in milliseconds or number of samples. We measured Pre RR Interval as the number of samples between consecutive beats in a signal sampled at 360 Hz. As an example, we chose Patient Records 209, 222 and 232 from MIT-BIH to demonstrate inter and intra-patient variations. Lannoy et al [7] demonstrated that Pre RR Interval is an important feature for classification. However, Pre RR Interval undergoes significant inter-patient and intra-patient variation. To reduce the impact of inter-patient and intra-patient variations, we propose a normalization technique that achieves significant improvement over existing Pre RR Interval normalization techniques [7]. Table 4.10 shows the behavior of Pre RR Interval. Each case represents a set of 5 consecutive heartbeats, with the Pre RR Interval reported on the bottom of each beat. Case 1 represents the variation in Pre RR Interval of 5 consecutive Normal beats from Patient Record 222 with a mean of 317 samples and standard deviation of 10.42 samples. With a standard deviation of 10.42 samples, the variation is minimal. Case 2 shows the variation in Pre RR interval of 5 consecutive SVEB's from Patient Record 209 with a mean of 141.8 and standard deviation of 2.39 samples. With a standard deviation of 2.39 samples, the variation is minimal for SVEB. Our analysis show that when we have a combination of heartbeats, as shown in Case 3 and Case 4, Pre RR Interval of SVEB is always lower than that of Normal beats. Case 3 and Case 4 represents a set of 5 consecutive beats from Patient Records 222 and 232. Our experiments have shown that behavior of Pre RR Interval of VEB is similar to that of SVEB, i.e. Pre RR Interval of VEB is lower than that of Normal beats. This makes Pre RR Interval an effective feature for separating SVEB or VEB from a Normal beat. However, Pre RR Interval exhibits inter and intra-patient variations, due to variations in heart rate, as shown in Table 4.11. Table 4.11 shows the overlap of Pre RR Intervals due to inter and intra-patient variations. Cases 3 and 4 in Table 4.10 demonstrates the same.

We propose a normalization technique to reduce overlap, resulting in a feature that captures the relation between SVEB and Normal beats. The normalized feature must be such that any occurrence of SVEB must have a Pre RR Interval that is lower than that of Normal beats and this property should hold in all cases. The global and local averages of Pre RR Interval proposed by [4, 11] is ineffective due to intra-patient variations. However, the local averaging of Pre RR Interval [11] is superior to global averaging [4].

Our approach identifies Normal beats in the neighborhood of each heartbeat and averages the Pre RR Interval of Normal beats detected. Normalization includes dividing Pre RR Interval of current heartbeat by the average of Pre RR Interval values in the neighborhood. The size of the neighborhood is determined empirically. The details are given in Section 4.5.3. In case no normal beats are identified, nearest average Pre RR Interval is used. Case 5 and Case 6 in Table 4.10 are normalized values of Pre RR Interval corresponding to Cases 3 and 4. Average Pre RR Interval of Normal beats for Cases 3 and 4 are 432.33 and 684.5 samples respectively. Each Pre RR Interval in Case 3 is divided by 432.33 and that in Case 4 is divided by 684.5, resulting in values reported in Cases 5 and 6. Such a normalization technique ensures that SVEB has a Pre RR interval lower than that of Normal beats, hence reducing overlap between SVEB and Normal beats. Comparing unnormalized and normalized values of SVEB in column 3 of Table 4.10, we noticed that while the unnormalized values are substantially different (165 and 247), the normalized values are close to each other (0.38 and 0.36). This reduces inter and intra-patient variations inherent in the Pre RR interval feature. Note that since VEB has Pre RR Interval lower than that of Normal beats, the same process helps separate VEB from a Normal beat. Section 4.5.2.4 describes the procedure used to detect the normal beats surrounding each heartbeat.

Table 4.10: Each pair of rows represents a case study

	Case 1	**Normal**	**Normal**	**Normal**	**Normal**	**Normal**
		323	299	318	325	320
	Case 2	**SVEB**	**SVEB**	**SVEB**	**SVEB**	**SVEB**
Before Normalization		142	145	143	140	139
	Case 3	**Normal**	**SVEB**	**Normal**	**SVEB**	**Normal**
		411	165	452	175	434
	Case 4	**Normal**	**SVEB**	**SVEB**	**Normal**	**SVEB**
		665	247	268	704	250
	Case 5	**Normal**	**SVEB**	**Normal**	**SVEB**	**Normal**
After Normalization		0.95	.38	1.05	.40	1.00
	Case 6	**Normal**	**SVEB**	**SVEB**	**Normal**	**SVEB**
		0.97	0.36	0.39	1.03	0.37

Table 4.11: Table demonstrates the overlap of Pre RR Interval of SVEB and Normal beats due to inter and intra-patient variations. With a range of 159.17 – 182.65 and 159.5 – 187.66 for SVEB and Normal beats respectively, Pre RR Intervals of Patient Record 222 clearly has overlap due to intra patient variations (Bold entries in first row of Table 4.11). With a range of 257.52 – 270.48 for SVEB in Patient Record 232 and 265.66 – 292.98 for Normal beats in Patient Record 222, Pre RR Interval of Normal beats clearly overlaps with the Pre RR Interval of SVEB (Bold entries in last two rows of Table 4.11).

Range of Pre RR Interval (In samples)	**Patient Record**	**Total # of beats**	**# of SVEB**	**Mean (SVEB)**	**Std (SVEB)**	**Range of Pre RR Interval for SVEB (In samples)**	**# of N beats**	**Mean (Normal)**	**Std (Normal)**	**Range of Pre RR Interval for Normal beats (In Samples)**
151 - 200	222	712	182	170.91	11.74	**159.17–182.65**	530	173.58	14.08	**159.5–187.66**
	232	2	2	173.50	12.02	161.48 -185.52	0	-	-	-
201 - 250	222	269	23	218.09	14.66	203.43–232.75	246	222.9	13.92	208.98–236.82
	232	209	207	244	5.62	238.38–249.62	2	247	2.12	244.88–249.12
251 - 300	222	749	4	262	10.58	251.42–272.58	745	279.32	13.66	**265.66–292.98**
	232	1170	1170	264	6.48	**257.52–270.48**	0	-	-	-

4.5.2.4 Classifier 1 and Classifier 2

Stated formally, if we define a window of size k, the purpose of Classifiers 1 and 2 is to detect the Normal beats in that window of size k. Figure 4.3 shows an example of a window of size 5. Once the Normal beats are detected within the window, the average Pre RR Interval of Normal beats within the window is used to divide the Pre RR Interval of the current heartbeat, to normalize the feature. This is repeated for each heartbeat in a patient record. Our analysis show that Normal beats are harder to detect when occurring at high heart rate.

To solve this problem, we trained Classifier 1 to detect Normal beats at low heart rate and Classifier 2 to detect the Normal beats at high heart rate. Any heartbeat with Pre RR Interval less than 230 samples are considered as occurring at high heart rate and any heartbeat with Pre RR Interval greater than or equal to 230 samples if considered as occurring at low heart rate. The threshold of 230 was determined empirically. Classifier 1 was trained on dataset DS1. Classifier 2 was trained on dataset DS1 after removing from DS1 all the heartbeats with Pre RR Interval greater than or equal to 230 samples. Such a classifier is fine-tuned for classification at high heart rate. Separate feature sets were extracted from dataset DS1 for training Classifiers 1 and 2. Using the wrapper algorithm (See section 4.5.2.2), the best set of features is Feature Set 1 in Table 4.9 for detecting Normal beats at low heart rate. Similarly, Feature Set 2 from Table 4.9 is the best set of features determined by the wrapper algorithm to detect Normal heartbeat at high heart rate.

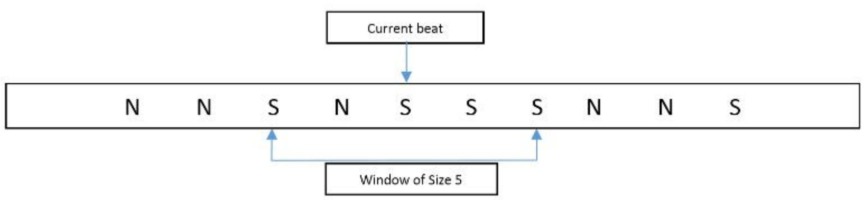

Figure 4.3: Normalization for a window of size 5

We combined Classifier 1 and Classifier 2 into a single system. When a new test heartbeat arrives, we scan a window surrounding the new heartbeat. For each heartbeat in the window, if the Pre RR Interval is greater than or equal to 230 samples, Feature Set 1 is extracted and classified using Classifier 1. If the Pre RR Interval is less than 230 samples, Feature Set 2 is extracted and classified using Classifier 2. The average Pre RR Interval of Normal beats detected within the window is used to normalize the Pre RR Interval of current heartbeat. We tested the system on dataset DS2 with window size fixed at 3, which was determined empirically (See section 4.5.3). Results are reported in Table 4.13. Classifier 1 and Classifier 2 detected 41621 out of 44228 Normal beats, resulting in a sensitivity (See section 4.5.3) of 94.1%.

4.5.2.5 Classifier 3

The application of wrapper algorithm produced Feature Set 3 which was used in conjunction with classifier 3. We extracted Feature Set 3 from each heartbeat in training set DS1 and used it to train Classifier 3. Note that since we have apriori knowledge on the type of each heartbeat in the training set, calculating normalized Pre RR Interval for each heartbeat in the training set is straightforward. This is because we have apriori knowledge of Normal beats within the window surrounding each heartbeat. However to normalize Pre RR Interval of heartbeats in testing set, detection of normal beats within the window is necessary. Once classifier 1 and classifier 2 detects the Normal beats within a window of size 3, the Pre RR Interval of the current heartbeat is normalized using the average Pre RR Interval of detected normal beats. In case normal beats are not detected within the window, the previous mean is used. With the normalization complete, Feature Set 3 is extracted from the current heartbeat. Classifier 3 is then used to perform the final classification. Table 4.12 shows classification algorithm.

Table 4.12: Algorithm to implement our approach

1. Fix a window size, e.g. 3 (In our illustration we chose 3, the user can chose window size 'k' as appropriate)
2. Extract Feature Set 1 from each heartbeat present in training dataset DS1.
3. Train Classifier 1 using the data extracted in step 1
4. Locate the heartbeats in training dataset DS1 that has Pre RR Interval less than 230 sample time.
5. Extract Feature Set 2 from each heartbeat located in training set using Step 3
6. Train Classifiers 1 and 2 using the data extracted in step 3 and step 4 respectively
7. For each heartbeat in testing dataset DS2
    1) For each heartbeat in the window surrounding the heartbeat, extract Feature Set 1 and Feature Set 2
    2) For each heartbeat in the window
        a. If Pre RR Interval is greater than or equal to 230 sample time, use Classifier 1 to classify the heartbeat
        b. If Pre RR Interval is less than 230 sample time, use Classifier 2 to classify the heartbeat
    3) Locate the beats within the window that has been classified as a Normal beat in step 7.2 and find their average Pre RR Interval. If no normal beats are detected in current window, use previous average.
    4) Divide the Pre RR Interval of current heartbeat using the average calculated in step 7.3
    5) Extract Feature Set 3 for the current heartbeat.
    6) Use classifier 3 to classify the heartbeat into one of the 5 classes {Normal, SVEB, VEB, Fusion and Paced}

4.5.3 Results and Discussion

Table 4.13 reports the classification results for the combined classifier (Classifier 1 and Classifier 2) and Classifier 3. Classification performance is measured in terms of sensitivity (Se), positive predictive value (PPV) and F-Score. Column 1 in Table 4.13 represents the classifier, Column 2 represents the sensitivity (Se) for SVEB and column 3 represents the Positive predictive value (PPV) for SVEB.

A combined system of classifier 1 and classifier 2 achieved a sensitivity of 94.1% and PPV of 47.7% in detecting Normal beats when tested on dataset DS1. For detecting SVEB, classifier 3 achieved the highest F-Score of 77.86 when window size was set to 3, with a sensitivity of 91.94% and PPV of 67.52%. We tested for windows of size 3, 5,7,11,13,15,17 and 21. It was observed that as the window size increases, F-Score decreases. The decrease in F-Score is triggered by the reduction in PPV. This confirms the observation that Pre RR Interval of a heartbeat is a localized phenomenon and dependent on the local heart rate. For detecting VEB, classifier 3 achieved a sensitivity of 81.98% and PPV of 96.63%.

Table 4.13: Classification Results. Classifier 1 and Classifier 2 represents the classification results for detecting Normal beats. Classifier 3 represents the classification results for detecting SVEB and VEB

Classifier	**Se**	**PPV**	**F-Score**
Classifier 1 and Classifier 2	94.1	98.7	96.34
Classifier 3 (SVEB)	91.94	67.52	77.86
Classifier 3 (VEB)	81.98	96.63	88.7

Table 4.14 reports the gross classification performance of classifier 3. Throughout the analysis, we used a window of size 3. 'Gross' classification performance is the performance of classifier 3 for the entire dataset DS2. We reported the gross results for two different cases. While Alvarado et al [9] and Chazal et al [6] reported the gross results for the dataset DS2 after ignoring the first 500 heartbeats from each patient record, Chazal et al [4] reported gross results for the entire dataset DS2. Chazal et al [6] and Alvarado et al [9] used a first 500 beats to train the local classifier and therefore did not include them in the testing set. In order to compare our results with Chazal et al [4], Chazal et al [6] and Alvarado et al [9], we reported the classification results for both the cases. The results for the entire dataset DS2 is represented by 'Gross*' and the results after ignoring first 500 heartbeats from each patient record is represented by 'Gross'. The performance is measured in terms of accuracy, sensitivity, positive predictive value and false positive rate (FPR). The sensitivity and PPV were calculated by treating it as a binary classification problem where a heartbeat is classified as either SVEB or a combined class consisting of 'Normal,' 'VEB,' 'Fusion,' and 'Paced'. Table 4.14 summarizes patient level performance of Classifier 3. Column 1 represents the data used for gross statistics, column 2-6 represent the number of beats of each type, and columns 7-10 represent the performance of the classifier 3. For SVEB, classifier 3 achieved a 'Gross' sensitivity of 91.39% and PPV of 65.21%. Classifier 3 achieved a sensitivity of 91.94% and PPV of 67.52% for the entire dataset DS2, reported as 'Gross*' in Table 4.14 (Bold entries in the last two rows of Table 4.14). Note that for VEB, 'Gross' sensitivity is 81.77% and 'Gross' PPV is 96.7%. Classifier 3 achieved a 'Gross*' sensitivity of 81.98% and PPV of 96.63% (Bold entries in the last two rows of Table 4.14).

Table 4.14: Classification Results for Classifier 3

	Number of Beats					SVEB				VEB			
Rec	**N**	**S**	**V**	**F**	**Q**	**Acc**	**Se**	**PPV**	**FPR**	**Acc**	**Se**	**PPV**	**FPR**
100	2237	33	1	0	0	99.82	90.91	96.77	0.04	100	100	100	0
103	2081	2	0	0	0	99.9	50	50	0.05	100	-	-	0
105	2525	0	41	0	5	99.06	-	0	0.94	96.85	12.2	10	1.78
111	2122	0	1	0	0	99.91	-	0	0.09	100	100	100	0
113	1788	6	0	0	0	99.89	100	75	0.11	100	-	-	0
117	1533	1	0	0	0	99.93	100	50	0.07	100	-	-	0
121	1860	1	1	0	0	99.89	100	33.33	0.11	100	100	100	0
123	1513	0	3	0	0	99.93	-	0	0.07	100	100	100	0
200	1740	30	826	2	0	96.12	30	15.25	2.78	97.58	92.49	99.87	0.06
202	2060	55	19	1	0	95.83	87.27	36.92	3.94	99.2	26.32	62.5	0.14
210	2421	22	194	10	0	95.78	90.91	16	4.18	96.11	47.42	98.92	0.04
212	2747	0	0	0	0	98.07	-	0	1.93	100	-	-	0
213	2639	28	220	362	0	99.47	53.57	83.33	0.1	97.62	65	100	0
214	2001	0	256	1	2	98.77	-	0	1.23	98.27	89.84	94.65	0.65
219	2081	7	64	1	0	98.33	14.29	3.33	1.38	97.54	25	76.19	0.24
221	2030	0	396	0	0	97.44	-	0	2.56	99.71	98.23	100	0
222	2272	209	0	0	0	88.84	91.39	42.26	11.4	100	-	-	0
228	1687	3	362	0	0	98.53	33.33	4.17	1.36	99.12	95.3	99.71	0.06
231	1566	1	2	0	0	99.62	0	0	0.32	100	100	100	0
232	398	1381	0	0	0	94.45	98.33	94.63	18.29	98.76	-	0	1.24
233	2229	7	830	11	0	99.34	14.29	10	0.4	93.75	76.99	99.84	0.04
234	2698	50	3	0	0	98.25	10	55.56	0.15	100	100	100	0
Gross	34363	1440	2578	292	5	97.73	**91.39**	**65.21**	2.01	98.6	**81.77**	**96.7**	0.2
Gross*	44228	1836	3219	388	7	97.95	**91.94**	**67.52**	1.81	98.65	**81.98**	**96.63**	0.2

Gross – Classification results for the dataset DS2 ignoring first 500 heartbeats of each record

Gross* - Classification results for the dataset DS2 including the first 500 heartbeats of each record

Table 4.15 compares our algorithm with the state of the art classification techniques. Results are summarized in Table 4.15. Chazal et al [6] achieved a sensitivity of 87.7% and PPV of 47% for SVEB, while Alvarado et al [9] achieved a sensitivity and PPV of 86.19% and 56.68% respectively. Clearly, our algorithm outperform both Chazal et al [6] and Alvarado et al [9] with a 'Gross' sensitivity of 91.39% and PPV of 65.21%. We compare 'Gross*' with Chazal et al [4], Ince et al [5] and Wiens et al [11]. While our results for SVEB outperform Chazal et al [4] and Ince et al [5], Wiens et al [11] outperformed our method with a sensitivity of 92% and PPV of 99.5%. However, Wiens et al [11] based on an active learning methodology use patient specific data to aid the classifier, i.e. a subset of the patient data is required to be labeled by a domain expert. More explicitly, while [5] and [6] used roughly 350 and 500 beats respectively per patient and Wiens et al [11] requires 45 patient specific labeled samples, our method is fully automated and does not require a domain expert to label patient specific heartbeats. Therefore, Wiens et al [11] approach produces improved performance relative to sensitivity and positive predictive value. Although our results for VEB are comparable to the state of the art, our classifier achieved a comparatively lower sensitivity. Comprehensive results and implementation details, including the source code, are available at [20].

Table 4.15: Comparison Results compared to the state of the art

Methods	SVEB		VEB	
Chazal et al [4]	75.9	38.5	77.7	81.9
Chazal et al [6]	87.7	47	94.3	96.2
Alvarado et al [9]	86.19	56.68	92.43	94.82
Ince et al [5]	63.5	53.7	84.6	87.4
Wiens et al [11]	92	99.5	99.6	99.3
Proposed (Gross)	**91.39**	**65.21**	**81.77**	**96.7**
Proposed (Gross*)	**91.94**	**67.52**	**81.98**	**96.63**

4.5.4 Summary

We have shown that by addressing the problems related to inter-patient and intra-patient variations, and selecting the most significant features that have measureable impact on classification metrics, overall performance can be improved significantly. We proposed a set of new features in the time domain and frequency domain, and demonstrated the significance of Pre RR Interval. We introduced a normalization technique, which when applied to Pre RR Interval improved classification rates. Furthermore, our technique is fully automated and eliminates the requirement for patient specific labeled data. Furthermore, these algorithms can be used in a real-time setting, and as continuous monitoring of patients via sensors is becoming a reality, our algorithms can be used in conjunction to provide timely interdiction to prevent CVD fatalities and reduce unnecessary admissions to hospitals. In addition, while others have leveraged the MIT-BIH database for arrhythmia detection, we examined the heartbeat patterns extensively and painstakingly to tease out subtle variations in order to extract features that significantly improve classification performance. This we believe should benefit the machine learning community as they address CVD related automated detection.

Chapter 4

Conclusion

Heartbeat arrhythmias needless to say are a major cause of unnecessary fatalities and any system that improves responsiveness for timely interdiction is not only helpful in reducing costs, but a moral imperative. In this thesis, we addressed several aspects of automatic detection, most notably, feature extraction, and a careful assessment of two types of arrhythmia, ventricular ectopic beats (VEB), and supraventricular ectopic beats (SVEB). We utilized a feature selection methodology which selects the most significant features from a feature set based on time, and frequency domains. Furthermore, we evaluated the performance of several classifiers; Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and Artificial Neural Networks (ANN). The methodology and framework we proposed is fairly automated and does not require the intervention of the domain expert (cardiologist) in the extraction of the features and labeling of the segments of the ECG signals.

Our results bear testimony to the improvements as seen in the commonly used metrics such as classifier sensitivity, specificity, F-measure, and positive predictive value in the evaluation of classifier performance for detecting arrhythmias. In the follow up studies, we propose to build a framework for the on-line detection of heartbeat irregularities, while improving the performance of the classifiers by extending our work on feature selection and modifications to the classifiers.

References

[1] 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. 215-220, 2000

[2] Y. H. Hu, S. Palreddy, and W. J. Tompkins, “A patient adaptable ECG beat classifier using a mixture of experts approach,” IEEE Trans. on Biomedical Engineering, vol. 44, no. 9, pp. 891-900, 1997

[3] R. Mark and R. Wallen, “AAMI-recommended practice: testing and reporting performance results of ventricular arrhythmia detection algorithms,” Tech. Rep. AAMI ECAR, 1987

[4] P. de Chazal, M. O’Dwyer, and R. Reilly, “Automatic classification of heartbeats using ecg morphology and heartbeat interval features,” Biomedical Engineering, IEEE Transactions on, vol. 51, no. 7, pp. 1196-1206, 2004

[5] T. Ince, S. Kiranyaz, and M. Gabbouj, “A generic and robust system for automated patient-specific classification of ecg signals,” Biomedical Engineering, IEEE Transactions on, vol. 56, no. 5, pp. 1415-1426, 2009

[6] P. de Chazal and R. Reilly, “A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features,” Biomedical Engineering, IEEE Transactions on, vol. 53, no. 12, pp. 2535 –2543, Dec. 2006.

[7] Gauthier Doquire, Gael de Lannoy, Damien François, Michel Verleysen, “Feature Selection for Interpatient Supervised Heart Beat Classification,” Comp. Int. and Neurosc, 2011

[8] P. N. Tan, M. Steinbach, and V. Kumar, “Introduction to Data Mining”, Addison-Wesley, 2005

[9] Alexander Singh Alvarado, Choudur Lakshminarayan, Jose C. Principe, "Time-based Compression and Classification of Heartbeats", IEEE Transactions on Biomedical Engineering, vol. 59, no. 6, 2012

[10] H. Feichtinger, J. Principe, J. Romero, A. Singh Alvarado, and G. Velasco, “Approximate reconstruction of bandlimited functions for the integrate and fire sampler,” Advances in Computational Mathematics, vol. 36, pp. 67-78, 2012

[11] J. Wiens and J. Guttag, “Active learning applied to patient-adaptive heartbeat classification,” in Advances in Neural Information Processing Systems 23, J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, Eds., pp. 2442—2450, 2010

[12] R.A. Johnson, and D.W. Wichern, Applied Multivariate Statistical Analysis, 3rd edition, Englewood Cliffs, New Jersey: Prentice Hall, 1992

[13] R. O. Duda, and P.E. Hart, Pattern Classification and Scene Analysis, New York: John Wiley & Sons, 1973

[14] G.J. McLachlan.: Discriminant Analysis and Statistical Pattern Recognition, New York: John Wiley & Sons, 1992

[15] J. A. Freeman and D.M. Skapura.: Neural Networks, Algorithms, Applications, and Programming Techniques, Computation and Neural systems Series, Reading, Massachusetts: Addision Wesley, 1991

[16] Tony Basil, Bollepalli S. Chandra, and Choudur Lakshminarayan, A Comparison of Statistical Machine Learning Methods in Heartbeat Detection and Classification, Lecture Notes in Computer Science, vol. 7678, pp. 16-25, 2012

[17] Tony Basil, Choudur Lakshminarayan, and C. Krishna Mohan, "Detection of Classes of Heart Arrhythmias based on Heartbeat Morphology Patterns," SIAM 2nd International Workshop on Analytics for Cyber-Physical Systems, 2013

[18] Tony Basil, Choudur Lakshminarayan and C. Krishna Mohan, "Feature Extraction and Selection for Detecting Certain Types of Heart Arrhythmias," IEEE International Conference on Big Data, 2013 (Submitted)

[19] http://www.who.int/mediacentre/factsheets/fs317/en/index.html

[20] https://sourceforge.net/projects/ecganalysis/