Next-Generation Computing and Communication Engineering

MEDICAL IMAGING AND HEALTH INFORMATICS

Edited By
Tushar H. Jaware
K. Sarat Kumar
Ravindra D. Badgujar
Svetlin Antonov

Scrivener Publishing

WILEY

Medical Imaging and Health Informatics

Scrivener Publishing
100 Cummings Center, Suite 541J
Beverly, MA 01915-6106

Next-Generation Computing and Communication Engineering

Series Editors: Dr. G. R. Kanagachidambaresan and Dr. Kolla Bhanu Prakash

Developments in articial intelligence are made more challenging because the involvement of multi-domain technology creates new problems for researchers. Therefore, in order to help meet the challenge, this book series concentrates on next generation computing and communication methodologies involving smart and ambient environment design. It is an effective publishing platform for monographs, handbooks, and edited volumes on Industry 4.0, agriculture, smart city development, new computing and communication paradigms. Although the series mainly focuses on design, it also addresses analytics and investigation of industry-related real-time problems.

Publishers at Scrivener
Martin Scrivener (martin@scrivenerpublishing.com)
Phillip Carmical (pcarmical@scrivenerpublishing.com)

Medical Imaging and Health Informatics

Edited by

Tushar H. Jaware

K. Sarat Kumar

Ravindra D. Badgujar

and

Svetlin Antonov

Scrivener Publishing

WILEY

This edition first published 2022 by John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA and Scrivener Publishing LLC, 100 Cummings Center, Suite 541J, Beverly, MA 01915, USA

For more information about Scrivener publications please visit www.scrivenerpublishing.com.

Wiley Global Headquarters
111 River Street, Hoboken, NJ 07030, USA

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Library of Congress Cataloging-in-Publication Data

ISBN 978-1-119-81913-4

Cover image: Pixabay.Com
Cover design by Russell Richardson

Set in size of 11pt and Minion Pro by Manila Typesetting Company, Makati, Philippines

Printed in the USA

10 9 8 7 6 5 4 3 2 1

Contents

Preface

There are many aspects to medical imaging and health informatics, including how they can be applied to real-world biomedical and healthcare challenges. Therefore, a collection of cutting-edge artificial intelligence (AI) and other allied approaches for healthcare and biomedical applications are provided in this book. Moreover, a diverse collection of state-of-the-art techniques and recent advancements in AI approaches are given, which are geared toward the challenges that healthcare institutions and hospitals face in terms of early detection of diseases, data processing, healthcare monitoring and prognosis of diseases.

Medical imaging and health informatics is a subfield of science and engineering which applies informatics to medicine and includes the study of design, development, and application of computational innovations to improve healthcare. The health domain has a wide range of challenges that can be addressed using computational approaches; therefore, the use of AI and associated technologies is becoming more common in society and healthcare. Currently, deep learning algorithms are a promising option for automated disease detection with high accuracy. Clinical data analysis employing these deep learning algorithms allows physicians to detect diseases earlier and treat patients more efficiently. Since these technologies have the potential to transform many aspects of patient care, disease detection, disease progression and pharmaceutical organization, approaches such as deep learning algorithms, convolutional neural networks, and image processing techniques are explored in this book.

This book also delves into a wide range of image segmentation, classification, registration, computer-aided analysis applications, methodologies, algorithms, platforms, and tools; and gives a holistic approach to the application of AI in healthcare through case studies and innovative applications. It also shows how image processing, machine learning and deep learning techniques can be applied for medical diagnostics in several specific health scenarios such as COVID-19, lung cancer, cardiovascular diseases, breast cancer, liver tumor, bone fractures, etc. Also highlighted are the significant issues and concerns regarding the use of AI in healthcare together with other allied areas, such as the internet of things (IoT) and medical informatics, to construct a global multidisciplinary forum.

Since elements resulting from the growing profusion and complexity of data in the healthcare sector are emphasized in this book, it will assist scholars in focusing on future research problems and objectives. Our principal goal is to leverage AI, biomedical and health informatics for effective analysis and application to provide a tangible contribution to innovative breakthroughs in healthcare.

Dr. Tushar H. Jaware
Dr. K. Sarat Kumar
Dr. Ravindra D. Badgujar
Dr. Svetlin Antonov
April 2022

1

Machine Learning Approach for Medical Diagnosis Based on Prediction Model

Hemant Kasturiwale[1]*, Rajesh Karhe[2] and Sujata N. Kale[3]

[1]Thakur College of Engineering and Technology, Kandivali (East), Mumbai, MS, India
[2]Shri Gulabrao Deokar College of Engineering, Jalgaon, MS, India
[3]Department of Applied Electronics, Sant Gadge Baba University, Amravati, MS, India

Abstract

The electrocardiography is the most crucial biosignals for critical analysis of the heart. The heart is the human body's most vital and variety of control mechanisms that regulate the heart's activities. The heart rate is an essential measure of cardiac function. The heart rate is represented as a time interval equal between two corresponding electrocardiogram (ECG) "R" peaks. The heart rate varies with the heart's state. A machine learning technique is used to categorize the statistical parameters mentioned above to predict the individual's physical state, including sleep, examination, and exercise, based on a physiologically important factor known as HRV. The chapter is focused on uses of manual classified data. Each hospital, clinic, and diagnostic center produces massive quantities of information such as patient records and test results to predict the presence of heart disease and provide care for the early stages. The results are validated and compared with predictions obtained from different algorithms. Classification and prediction are a mining technique that uses training data to construct a model, and then, that model is applied to test data to predict outcomes. Different algorithms are employed to disease datasets to diagnose chronic disease, and the findings have been positive. There is a need to establish an appropriate technique for the diagnosis of chronic diseases. This chapter discusses with insight various kinds of classification schemes for chronic disease prediction. Here, readers will come to choice know machine learning and classifiers made to get knowledge out of datasets.

***Keywords*:** ECG, biosignals, machine learning, HRV, classification, prediction, cardiac diseases

**Corresponding author*: hemantkasturiwale@gmail.com

Tushar H. Jaware, K. Sarat Kumar, Ravindra D. Badgujar and Svetlin Antonov (eds.) Medical Imaging and Health Informatics, (1–22) © 2022 Scrivener Publishing LLC

1.1 Introduction

Biosignals are being used in various medical data, such as the electroencephalography (EEG), capturing electric fields created by brain cell activity, and magnetoencephalography (MEG) capturing magnet fields produced by electrical brain cell activity. The electrical stimulation comes from biological activity in various parts of the body. The most popular types of methods currently used to record biosignals in clinical research are described below, along with a brief overview of their functionality and related clinical application signals [1].

1.1.1 Heart System and Major Cardiac Diseases

The electrical activity generates the following types of signals:

- Magnetoencephalography (MEG) signals
- Electromyography (EMG) signals
- Electrooculography (EOG)signals
- Phonocardiography (PCG) signals
- Electrocorticography (ECoG) signals
- Electrocardiography (ECG or EKG) signals

Intervals between the waves are used as indicators of irregular cardiac operation, e.g., a prolonged PR interval from atrial activation to the start of ventricular activation may indicate cardiac failure [2, 3]. In addition, ECGs are used to study arrhythmias [4], coronary artery disease [5], and other heart failure disorders. In biosignals, the sampling frequency (or sampling rate) and the recording period are directly proportional to the data size and the data acquisition process speed. The ECG will be essential for the heart rhythm and disease research. The different heart conditions are as follows:

a) Arrhythmias
b) Coronary heart disease
c) Various types of heart blocks
d) Fibrillations
e) Congestive heart failure (CHF)
f) Myocardial infarction (MI)
g) Premature ventricular contraction (PVC)

1.1.2 ECG for Heart Rate Variability Analysis

Electrocardiogram (ECG) is a waveform pattern that describes the state of cardiac activity and cardiac safety. The ECG signal is non-stationary and non-linear. The ECG has a spectrum of frequencies between 0.05 and 100 Hz [6]. ECG analysis methods, including the heart rate variability (HRV), QRS identification, and ECG post-processing, have advanced considerably since device implementation. The word HRV reflects the interval difference between successive heartbeats.

1.1.3 HRV for Cardiac Analysis

The biomedical signal is an important health assessment parameter. For example, it has been used to detect and predict human stress [1], stroke, hypertension, sleep disorder, age, gender, and many more. The popular techniques to analyze the HRV fall into three categories as time domain, spectral or frequency domain based on fast Fourier transform (FFT) [7], and nonlinear methods consisting of Markov modeling, entropy-based metrics [8], and probabilistic modeling [9]. There are seven commonly used statistical time domain parameters [10] calculated from HRV segmentation during 5-min recording, comprising of RMSSD, SDNN, SDANN, SDANNi, SDSD, PNN50, and autocorrelation, which are considered for implementation. The HRV is also calculated by a device called PPA (peripheral pulse analyzer); it works based on pulses measured, which is different from HRV measurement using ECG. However, the focus would be on ECG-based HRV measurement, but the validation PPA-based method is considered [11]. Nonlinear measurement approaches aim to calculate the structure and complexity of the time series of RR intervals. HRV signals are non-stationary and nonlinear in nature. Analysis of HRV dynamics by methods based on chaos theory and nonlinear system theory is based on findings indicating that the processes involved in cardiovascular control are likely to interact with each other in a nonlinear manner. The more on indices (features/parameters) are discussed in Section 1.3.2.

1.2 Machine Learning Approach and Prediction

Learning is closely connected to (and sometimes overlaps with) quantitative statistics, which often concentrate on forecasting computers' use. It has close connections with mathematical optimization, which provides the fields of methodology, theory, and implementation. The second sub-area focuses more on the study of exploratory data and is also known as non-monitored learning [2]. Unsupervised machine learning (ML) is also possible [11] and can be used to learn and construct baseline conduct profiles for different entities [12]. To gain knowledge of the past and to detect useful trends from massive, unstructured, and complex databases, machine learning algorithms use a range of statistical, probabilistic, and optimization methods [12]. These algorithms include automatic categorization of texts, network intrusion detection, junk e-mail filtering, credit-card fraud detection, consumer buying behavior, manufacturing optimization, and disease modeling. Most of these applications are performed using managed variants of the algorithms of ML rather than unattended [13].

The heart disease detail includes several features that predict heart disease. This large amount of medical data allowed data mining techniques to discover trends and diagnose patients. The historical medical data is very high, so it requires computational methods to process it. Data mining is a technique that removes the hidden pattern and uses as an analytical tool to analyze historical data. There are several different classification schemes for disease datasets. ML techniques are applied for classifying the statistical parameters above in a cardiological signal analysis to predict the RR interval estimate cannot be overemphasized. A precise method of calculation therefore needs to be developed. It is clear from the existing research theory that the conventional systems for chronic disease prediction are unable to establish reliable diagnostic systems as workers make it difficult to get correct responses and can minimize response time. Adaptive systems, by comparison, can increase the chances of success and can advise clinicians on care decisions. Current healthcare programmers can be enhanced by the efficient use of parallel classification systems, as they promote parallel implementation on multiple systems. Parallel classification systems also have a great potential to increase the predictive performance of diagnostic systems for chronic diseases [13, 14]. Here, classifiers are discussed out of the available are K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Ensemble AdaBoost (EAB), and Random Forest (RF).

1.3 Material and Experimentation

The proposed method comprises of two phases:

a. processing the enrolment database (PEP) and
b. Prediction (P).

Figure 1.1 shows that the research purpose types of database are created based on acquisition units. The standard database has varying sampling frequency which comprises of different age groups of male and female.

A total number of subjects and corresponding signal were acquired with different set conditions. This may comprises of female and male with varying age group with sampling frequency of 256 and 500 Hz [6, 15]. The model will be testing for cardiac HRV-based analysis with both the ECG and non-ECG (PPA). For the research purpose, the congestive heart failure, arrhythmia, sudden cardiac death, ventricular arrhythmia, CHF database data being considered along with externally obtained ECG and non-ECG.

1.3.1 Data and HRV

The research uses the normal and cardiac subject's standard data [16] and externally acquired ECG or non-ECG data. Hence, the proposed techniques for the classification of cardiac diseases use data with varying characteristics. The DAQ cards help in

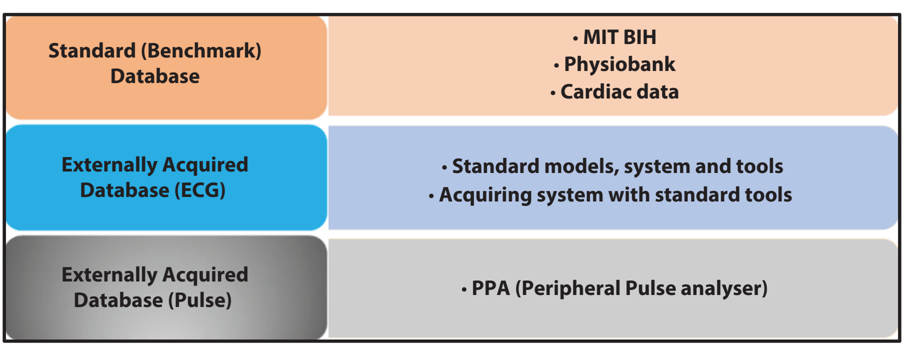

Figure 1.1 Acquisition system and sources [source: 14].

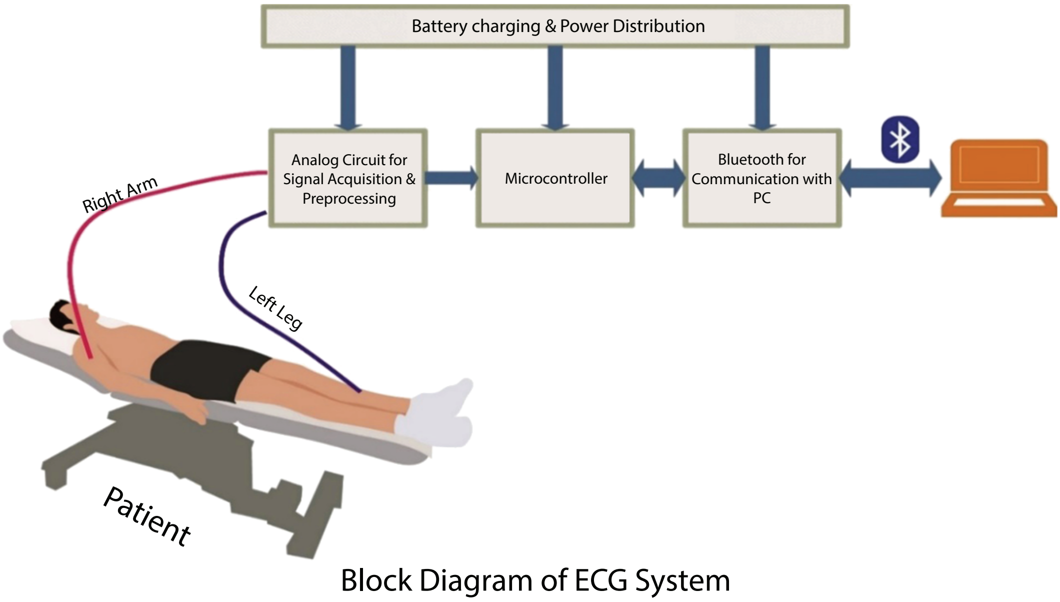

Figure 1.2 ECG acquisition system with connection (low cost).

creating a database of ECG or non-ECG signals. Figure 1.2 shows the HRV data and categorization, ensuring the data obtained is free from significant artifacts or noise. The sources of data and systems related to signal acquiring are a part of system. For more insight, the following methods/techniques are used for data acquisition in support of standard tools.

1.3.1.1 HRV Data Analysis via ECG Data Acquisition System

The analog circuit for ECG acquisition is possible with one or three channels. The analog devices as a signal conditioning circuit are used for data collection electronically.

The three-channel data acquiring system is attached to the body of the subject for recording purpose. These probes collect the ECG signal and give it as input to the ECG kit. The ECG kit comprises low-pass filters and an ECG chip. The electronic assembly is customized to acquired signal and processed further till detection. The myDAQ software, which works with NI DAQ, records the ECG wave, processes it, and provides analysis regarding the subject's heart condition. Three probes are connected to the ECG kit to test the signal and CRO to verify the signal. The extracted ECG signal from the subject is filtered as first step of process. The NIDAQ card processes a pure electrical signal. The front panel of the analog circuit with MCP6004 and instrumentation amplifier with other passive components is preferred as low cost and effective option. The circuit removes the baseline noise, line interference, and extracting data even for a few more seconds. It is effective for short duration records and has low storage capacity. The mechanism to reduce noise or artifact is effective to some extent for this circuit. The wireless connectivity is also major advantage of chapter system.

1.3.2 Methodology and Techniques

The proposed method breaks down HRV signals with the collection of features and checks consistency. Features are derived from the HRV signals components. Eventually, the classification is done with the classification unit. The classifier is used here for inspection and checking earlier. The best classifier is selected based on the classification parameters.

The approach proposed comprises two phases: Enrolment Database Processing (PEP) and Prediction and Identification (PI). All available data samples and ECG signals obtained by the units are fed for analysis as shown in Figure 1.3. However, the proposed model is compatible with age, gender, and feature (static) as other input conditions [17]. The proposed model design, heart disease dataset, data pre-processing, and performance measurement are critical and have been taken care of. The proposed methods are developed based on the following essential parameters:

- Short-term and long-term analysis
- Feature's indices and their sequencing
- Mathematical indices
- The technique for a standard database and classifiers

- ▪ The noise and impact study on ECG-HRV
- ▪ Noise impact on non–ECG-HRV

- The technique for an acquired database (ECG and non-ECG HRV) and classifiers
- Performance evaluation criteria and validation

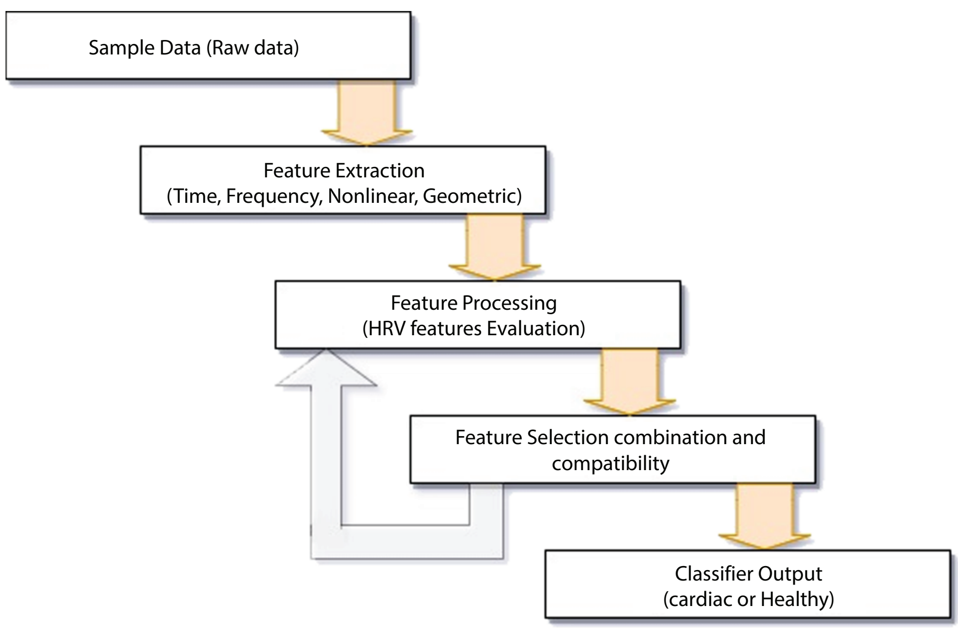

Figure 1.3 Cardiac diseases identification model for cardiac diseases [source: 24].

1.3.2.1 Classifiers and Performance Evaluation

The parameters used for evaluating the algorithm's performance are accuracy, precision, F-measures, recall, and execution time [18, 19]. These parameters are defined using four measures: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) [20, 21].

1.3.2.1.1 Performance Metrics

Performance metrics calculate how well a given algorithm performs with accuracy, precision, sensitivity, specificity, and other parameters. The different performance metrics are as below.

1.3.2.1.1.1 Confusion Matrix

The confusion matrix shows the performance of the algorithm. It depicts how the classifier is confused while predicting. The rows indicate the class label's actual instance, while the columns indicate the predicted class instances. Table 1.1 shows a confusion matrix for binary classification. TP value means the positive value is correctly predicted, FP means positive value is falsely classified, FN means the negative value is falsely predicted, while the TN means negative value is correctly classified. A confusion matrix table used to calculate different performance metrics.

1.3.2.1.2 Major Features Contributors With Analysis

The details added mathematical features are an essential part of the total of 24 features in Tables 1.2A and B. The formulation of Table 1.2 is a restructured table looking into the needs of research to analyze short term and long-term duration data size. The mathematical features Dalton DSD index, Dalton MABB index, and De Hann LTV index are added and have significant in view of analysis.

The HRV indices are known as HRV parameters or HRV features. The feature acronym and feature name reflected in Table 1.2. The time domain and frequency domain, and linear and some of the nonlinear indices are part of the research [22]. The indices are divided into four groups. The proposed model works on the development of a new set of indices group [23, 24]. The mathematical indices, namely, Dalton and De Hann, are being figured as part of the feature group (Group 4), as shown in Table 1.2B. The feature group's novelty is that they are created as per their essential characteristics to improve the model's performance. The mathematical equation explains the dependency of variables with features [25].

Table 1.1 Confusion matrix.

Actual label	**Predicted label**	
	+(1)	−(0)
+(1)	True Positive	False Negative
−(0)	False Positive	True Negative

Table 1.2 (A) New proposed HRV indices with groups.

Sr. no.	Feature acronym	Feature name
Time domain (Group 1)		
1	meanRR	Mean value RR period
2	SDNN	Standard deviation of intervals (NN)
3	Mean	Mean value of the heart rate (HR)
4	sdHR	Standard heart rate
5	NNx	Total number of interval successive NN intervals greater than "x" ms
6	HRVTi	Integral of the density of the RR interval histogram divided by its height
7	TINN	Baseline width of the RR interval histogram
8	pNNx	Percentage of successive RR intervals differ by more than "x" ms
9	RMSSD	Root mean square of successive RR interval differences
Frequency domain (Group 2)		
10	aHF	Areas within a higher frequency band (0.15–0.4 Hz)
11	aLF	Areas within a lower frequency band (0.04–0.15 Hz)
12	Raio (aLF/aHF)	The ration of LF to HF
Nonlinear (Group 3)		
13	Ent	Sample entropy
14	Hval	Hurst component
15	avgpsdf	Average power spectral density
16	hfdf	Higuchi fractal dimension
17	D	The factor of the dimension of time series
18	Alpha	Scaling exponent for alignment of series points
Other features (parameters/indices) (Group 4)		
19	SD1	The standard deviation of the distance of each point from the y-axis
20	SD2	The standard deviation of the distance of each point from the x-axis

1.3.3 Proposed Model With Layer Representation

The HRV analysis for cardiac diseases is complicated, so the step-by-step processes are defined as a part of HRV analysis. The research work has come up with the development of a robust model via the layer model. The features contribution and impact is a significant

Table 1.2 **(B)** New proposed HRV indices with groups.

21	CD	Correlation dimension
22	Dalton DSD index	The standard deviation of RR of length HR signal (long-term variability index)
23	Dalton MABB index	Absolute of one-half of arithmetic mean value of differences of subsequent RR intervals (short-term variability index)
24	De Hann LTV index	As an interquartile range of radius location of particular RR intervals (long-term variability index)

contribution of research that helps in the classification of cardiac diseases. The model development has a three-part fixed set feature model (FSM), flexi intra group selection model, and qualitative analysis, as shown in Figure 1.4. The long-term and short-term analyses are unique with the development of the model. The model performs under all conditions, and so the results obtained are encouraging for future growth. Figures 1.4 and 1.5 show a novel approach to identify and predict cardiac diseases with many features like feature extraction, feature concatenation, and combination. The responsiveness of the algorithm is on HRV parameters (linear, nonlinear, time, and frequency). Here, research work has included mathematical parameters like Dalton and Higuchi to enhance the method's efficiency.

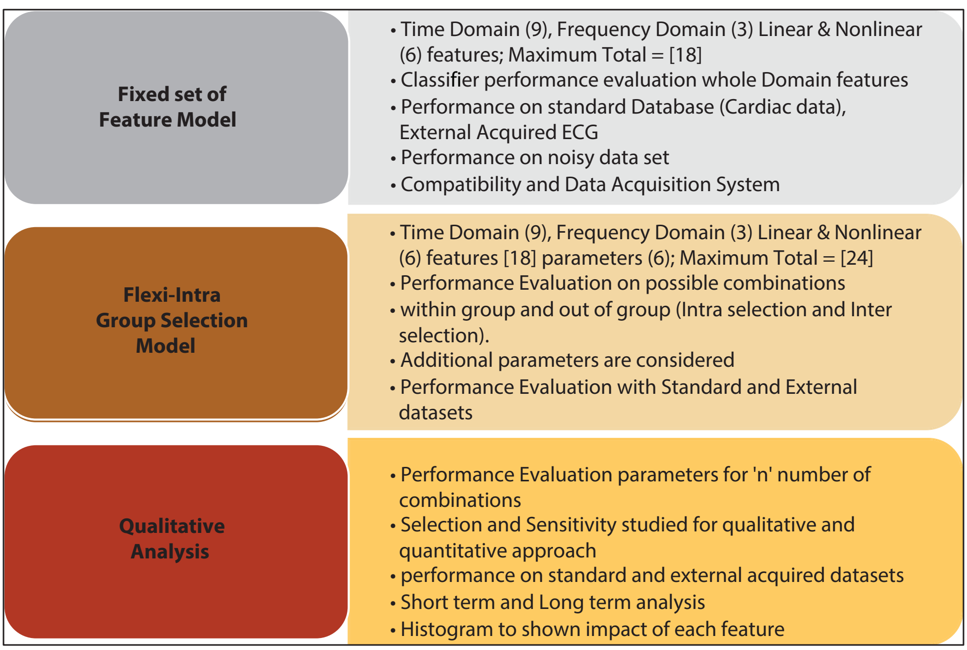

Figure 1.4 Robust model layers.

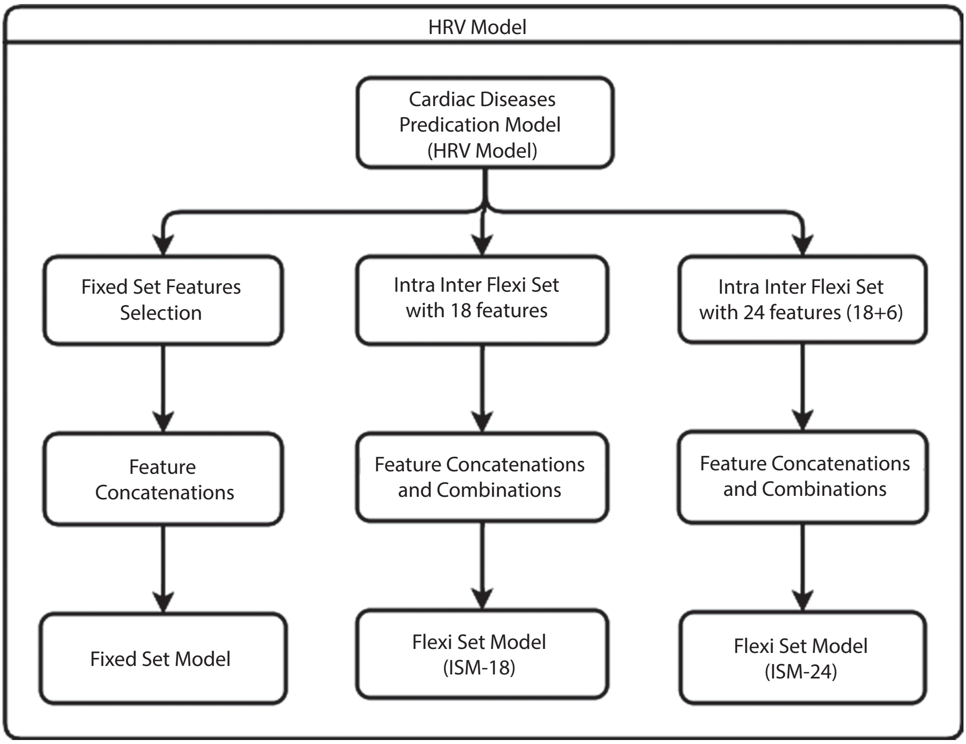

Figure 1.5 HRV model for cardiac prediction.

i) Support Vector Machine (SVM)
   a. SVM linear
   b. SVM polynomial
   c. SVM Gaussian
ii) RF
   a. With variation in the number of trees
iii) KNN
iv) EAB

The research aims to enhance HRV analysis to identify and predict cardiac diseases using a ML algorithm. The model's performance depends on the quality of input data and features for predication cardiac diseases. The research present the model which can be customized looking into needs data size and type of input signal. The model tested all possible subjects' conditions and database like raw and non-ECG signal. The research reviews current perspectives on the prediction of cardiac diseases that needs 24 h, short-term (~1 min), and long term (>1 min) HRV. The research enhances the importance of HRV and its implications for health and performance. The investigation provides an insight into widely used