

# MEDICAL IMAGING AND HEALTH INFORMATICS

Edited By

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



# 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 artificial 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](mailto:martin@scrivenerpublishing.com))  
Phillip Carmical ([pcarmical@scrivenerpublishing.com](mailto:pcarmical@scrivenerpublishing.com))

# **Medical Imaging and Health Informatics**

Edited by  
**Tushar H. Jaware**  
**K. Sarat Kumar**  
**Ravindra D. Badgujar**  
and  
**Svetlin Antonov**



**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  
© 2022 Scrivener Publishing LLC  
For more information about Scrivener publications please visit [www.scrivenerpublishing.com](http://www.scrivenerpublishing.com).

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, except as permitted by law. Advice on how to obtain permission to reuse material from this title is available at <http://www.wiley.com/go/permissions>.

**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](http://www.wiley.com).

**Limit of Liability/Disclaimer of Warranty**

While the publisher and authors have used their best efforts in preparing this work, they make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials, or promotional statements for this work. The fact that an organization, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and authors endorse the information or services the organization, website, or product may provide or recommendations it may make. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read.

**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

---

<b>Preface</b>	<b>xvii</b>
<b>1 Machine Learning Approach for Medical Diagnosis Based on Prediction Model</b>	<b>1</b>
<i>Hemant Kasturiwale, Rajesh Karhe and Sujata N. Kale</i>	
1.1 Introduction	2
1.1.1 Heart System and Major Cardiac Diseases	2
1.1.2 ECG for Heart Rate Variability Analysis	2
1.1.3 HRV for Cardiac Analysis	3
1.2 Machine Learning Approach and Prediction	3
1.3 Material and Experimentation	4
1.3.1 Data and HRV	4
1.3.1.1 HRV Data Analysis via ECG Data Acquisition System	5
1.3.2 Methodology and Techniques	6
1.3.2.1 Classifiers and Performance Evaluation	7
1.3.3 Proposed Model With Layer Representation	8
1.3.4 The Model Using Fixed Set of Features and Standard Dataset	11
1.3.4.1 Performance of Classifiers With Feature Selection	11
1.4 Performance Metrics and Evaluation of Classifiers	13
1.4.1 Cardiac Disease Prediction Through Flexi Intra Group Selection Model	13
1.4.2 HRV Model With Flexi Set of Features	14
1.4.3 Performance of the Proposed Modified With ISM-24	15
1.5 Discussion and Conclusion	18
1.5.1 Conclusion and Future Scope	19
References	20
<b>2 Applications of Machine Learning Techniques in Disease Detection</b>	<b>23</b>
<i>M.S. Roobini, Sowmiya M., S. Jancy and L. Suji Helen</i>	
2.1 Introduction	24
2.1.1 Overview of Machine Learning Types	24
2.1.2 Motivation	25
2.1.3 Organization the Chapter	25
2.2 Types of Machine Learning Techniques	25
2.2.1 Supervised Learning	25
2.2.2 Classification Algorithm	25
2.2.3 Regression Analysis	26

2.2.4	Linear Regression	27
2.2.4.1	Applications of Linear Regression	27
2.2.5	KNN Algorithm	28
2.2.5.1	Working of KNN	28
2.2.5.2	Drawbacks of KNN Algorithm	29
2.2.6	Decision Tree Classification Algorithm	29
2.2.6.1	Attribute Selection Measures	29
2.2.6.2	Information Gain	29
2.2.6.3	Gain Ratio	29
2.2.7	Random Forest Algorithm	29
2.2.7.1	How the Random Forest Algorithm Works	29
2.2.7.2	Advantage of Using Random Forest	30
2.2.7.3	Disadvantage of Using the Random Forest	31
2.2.8	Naive Bayes Classifier Algorithm	31
2.2.8.1	For What Reason is it Called Naive Bayes?	31
2.2.8.2	Disservices of Naive Bayes Classifier	31
2.2.9	Logistic Regression	31
2.2.9.1	Logistic Regression for Machine Learning	31
2.2.10	Support Vector Machine	32
2.2.11	Unsupervised Learning	32
2.2.11.1	Clustering	33
2.2.11.2	PCA in Machine Learning	35
2.2.12	Semi-Supervised Learning	38
2.2.12.1	What is Semi-Supervised Clustering?	38
2.2.12.2	How Semi-Supervised Learning Functions?	38
2.2.13	Reinforcement Learning	39
2.2.13.1	Artificial Intelligence	39
2.2.13.2	Deep Learning	40
2.2.13.3	Points of Interest of Machine Learning	41
2.2.13.4	Why Machine Learning is Popular	41
2.2.13.5	Test Utilizations of ML	42
2.3	Future Research Directions	43
2.3.1	Privacy	43
2.3.2	Accuracy	43
	References	43
3	<b>Dengue Incidence Rate Prediction Using Nonlinear Autoregressive Neural Network Time Series Model</b>	47
	<i>S. Dhamodharavadhani and R. Rathipriya</i>	
3.1	Introduction	47
3.2	Related Literature Study	48
3.2.1	Limitations of Existing Works	50
3.2.2	Contributions of Proposed Methodology	50

3.3	Methods and Materials	50
3.3.1	NAR-NNTS	50
3.3.2	Fit/Train the Model	51
3.3.3	Training Algorithms	54
3.3.3.1	Levenberg-Marquardt (LM) Algorithm	54
3.3.3.2	Bayesian Regularization (BR) Algorithm	55
3.3.3.3	Scaled Conjugate Gradient (SCG) Algorithm	55
3.3.4	DIR Prediction	55
3.4	Result Discussions	56
3.4.1	Dataset Description	56
3.4.2	Evaluation Measure for NAR-NNTS Models	57
3.4.3	Analysis of Results	57
3.5	Conclusion and Future Work	65
	Acknowledgment	66
	References	66
<b>4</b>	<b>Early Detection of Breast Cancer Using Machine Learning</b>	<b>69</b>
	<i>G. Lavanya and G. Thilagavathi</i>	
4.1	Introduction	70
4.1.1	Objective	70
4.1.2	Anatomy of Breast	70
4.1.3	Breast Imaging Modalities	71
4.2	Methodology	71
4.2.1	Database	71
4.2.2	Image Pre-Processing	71
4.3	Segmentation	72
4.4	Feature Extraction	72
4.5	Classification	72
4.5.1	Naive Bayes Neural Network Classifier	72
4.5.2	Radial Basis Function Neural Network	73
4.5.2.1	Input	73
4.5.2.2	Hidden Layer	73
4.5.2.3	Output Nodes	74
4.6	Performance Evaluation Methods	74
4.7	Output	75
4.7.1	Dataset	75
4.7.2	Pre-Processing	75
4.7.3	Segmentation	75
4.7.4	Geometric Feature Extraction	77
4.8	Results and Discussion	78
4.8.1	Database	78
4.9	Conclusion and Future Scope	81
	References	81

<b>5 Machine Learning Approach for Prediction of Lung Cancer</b>	<b>83</b>
<i>Hemant Kasturiwale, Swati Bhisikar and Sandhya Save</i>	
5.1 Introduction	84
5.1.1 Disorders in Lungs	84
5.1.2 Background	84
5.1.3 Material, Datasets, and Techniques	85
5.2 Feature Extraction and Lung Cancer Analysis	86
5.3 Methodology	87
5.3.1 Proposed Algorithm Steps	87
5.3.2 Classifiers in Concurrence With Datasets	88
5.4 Proposed System and Implementation	89
5.4.1 Interpretation via Artificial Intelligence	89
5.4.2 Training of Model	90
5.4.3 Implementation and Results	90
5.5 Conclusion	99
5.5.1 Future Scope	99
References	100
<b>6 Segmentation of Liver Tumor Using ANN</b>	<b>103</b>
<i>Hema L. K. and R. Indumathi</i>	
6.1 Introduction	103
6.2 Liver Tumor	104
6.2.1 Overview of Liver Tumor	104
6.2.2 Classification	105
6.2.2.1 Benign	105
6.2.2.2 Malignant	107
6.3 Benefits of CT to Diagnose Liver Cancer	108
6.4 Literature Review	108
6.5 Interactive Liver Tumor Segmentation by Deep Learning	109
6.6 Existing System	109
6.7 Proposed System	110
6.7.1 Pre-Processing	110
6.7.2 Segmentation	111
6.7.3 Feature Extraction	112
6.7.4 GLCM	112
6.7.5 Backpropagation Network	113
6.8 Result and Discussion	113
6.8.1 Processed Images	114
6.8.2 Segmentation	116
6.9 Future Enhancements	117
6.10 Conclusion	118
References	118
<b>7 DMSAN: Deep Multi-Scale Attention Network for Automatic Liver Segmentation From Abdomen CT Images</b>	<b>121</b>
<i>Devidas T. Kushnure and Sanjay N. Talbar</i>	
7.1 Introduction	121

7.2	Related Work	122
7.3	Methodology	123
7.3.1	Proposed Architecture	123
7.3.2	Multi-Scale Feature Characterization Using Res2Net Module	125
7.4	Experimental Analysis	126
7.4.1	Dataset Description	126
7.4.2	Pre-Processing Dataset	127
7.4.3	Training Strategy	128
7.4.4	Loss Function	128
7.4.5	Implementation Platform	129
7.4.6	Data Augmentation	129
7.4.7	Performance Metrics	129
7.5	Results	131
7.6	Result Comparison With Other Methods	135
7.7	Discussion	136
7.8	Conclusion	137
	Acknowledgement	138
	References	138
8	<b>AI-Based Identification and Prediction of Cardiac Disorders</b>	141
	<i>Rajesh Karhe, Hemant Kasturiwale and Sujata N. Kale</i>	
8.1	Introduction	142
8.1.1	Cardiac Electrophysiology and Electrocardiogram	143
8.1.2	Heart Arrhythmia	144
8.1.2.1	Types of Arrhythmias	145
8.1.3	ECG Database	147
8.1.3.1	Association for the Advancement of Medical Instrumentation (AAMI) Standard	147
8.1.4	An Overview of ECG Signal Analysis	148
8.2	Related Work	149
8.3	Classifiers and Methodology	151
8.3.1	Databases for Cardiac Arrhythmia Detection	152
8.3.2	MIT-BIH Normal Sinus Rhythm and Arrhythmia Database	152
8.3.3	Arrhythmia Detection and Classification	153
8.3.4	Methodology	153
8.3.4.1	Database Gathering and Pre-Processing	153
8.3.4.2	QRST Wave Detection	153
8.3.4.3	Features Extraction	154
8.3.4.4	Neural Network	155
8.3.4.5	Performance Evaluation	156
8.4	Result Analysis	156
8.4.1	Arrhythmia Detection and Classification	156
8.4.2	Dataset	156
8.4.3	Evaluations and Results	156
8.4.4	Evaluating the Performance of Various Neural Network Classifiers (Arrhythmia Detection)	157

8.5	Conclusions and Future Scope	159
8.5.1	Arrhythmia Detection and Classification	159
8.5.2	Future Scope	161
	References	161
<b>9</b>	<b>An Implementation of Image Processing Technique for Bone Fracture Detection Including Classification</b>	<b>165</b>
	<i>Rocky Upadhyay, Prakash Singh Tanwar and Sheshang Degadwala</i>	
9.1	Introduction	165
9.2	Existing Technology	166
9.2.1	Pre-Processing	166
9.2.2	Denoise Image	167
9.2.3	Histogram	168
9.3	Image Processing	169
9.3.1	Canny Edge	169
9.4	Overview of System and Steps	170
9.4.1	Workflow	170
9.4.2	Classifiers	171
9.4.2.1	Extra Tree Ensemble Method	171
9.4.2.2	SVM	172
9.4.2.3	Trained Algorithm	173
9.4.3	Feature Extraction	173
9.5	Results	174
9.5.1	Result Analysis	175
9.6	Conclusion	176
	References	176
<b>10</b>	<b>Improved Otsu Algorithm for Segmentation of Malaria Parasite Images</b>	<b>179</b>
	<i>Mosam K. Sangole, Sanjay T. Gandhe and Dipak P. Patil</i>	
10.1	Introduction	179
10.2	Literature Review	180
10.3	Related Works	182
10.4	Proposed Algorithm	183
10.5	Experimental Results	184
10.6	Conclusion	193
	References	193
<b>11</b>	<b>A Reliable and Fully Automated Diagnosis of COVID-19 Based on Computed Tomography</b>	<b>195</b>
	<i>Bramah Hazela, Saad Bin Khalid and Pallavi Asthana</i>	
11.1	Introduction	196
11.2	Background	196
11.3	Methodology	199
11.3.1	Models Used	199
11.3.2	Architecture of the Image Source Classification Model	199
11.3.3	Architecture of the CT Scan Classification Model	200
11.3.4	Architecture of the Ultrasound Image Classification Model	201

11.3.5	Architecture of the X-Ray Classification Model	201
11.3.6	Dataset	202
11.3.6.1	Training	202
11.4	Results	204
11.5	Conclusion	206
	References	207
<b>12</b>	<b>Multimodality Medical Images for Healthcare Disease Analysis</b>	<b>209</b>
<i>B. Rajalingam, R. Santhoshkumar, P. Santosh Kumar Patra, M. Narayanan, G. Govinda Rajulu and T. Poongothai</i>		
12.1	Introduction	210
12.1.1	Background	210
12.2	Brief Survey of Earlier Works	212
12.3	Medical Imaging Modalities	213
12.3.1	Computed Tomography (CT)	214
12.3.2	Magnetic Resonance Imaging (MRI)	214
12.3.3	Positron Emission Tomography (PET)	214
12.3.4	Single-Photon Emission Computed Tomography (SPECT)	215
12.4	Image Fusion	216
12.4.1	Different Levels of Image Fusion	216
12.4.1.1	Pixel Level Fusion	216
12.4.1.2	Feature Level Fusion	217
12.4.1.3	Decision Level Fusion	217
12.5	Clinical Relevance for Medical Image Fusion	218
12.5.1	Clinical Relevance for Neurocyticercosis (NCC)	218
12.5.2	Clinical Relevance for Neoplastic Disease	218
12.5.2.1	Clinical Relevance for Astrocytoma	218
12.5.2.2	Clinical Relevance for Anaplastic Astrocytoma	219
12.5.2.3	Clinical Relevance for Metastatic Bronchogenic Carcinoma	220
12.5.3	Clinical Relevance for Alzheimer's Disease	221
12.6	Data Sets and Softwares Used	221
12.7	Generalized Image Fusion Scheme	221
12.7.1	Input Image Modalities	222
12.7.2	Image Registration	222
12.7.3	Fusion Process	223
12.7.4	Fusion Rule	223
12.7.5	Evaluation	224
12.7.5.1	Subjective Evaluation	224
12.7.5.2	Objective Evaluation	224
12.8	Medical Image Fusion Methods	224
12.8.1	Traditional Image Fusion Techniques	224
12.8.1.1	Spatial Domain Image Fusion Approach	225
12.8.1.2	Transform Domain Image Fusion Approach	225
12.8.1.3	Fuzzy Logic-Based Image Fusion Approach	227
12.8.1.4	Filtering Technique-Based Image Fusion Approach	227

12.8.1.5	Neural Network-Based Image Fusion Approach	227
12.8.2	Hybrid Image Fusion Techniques	228
12.8.2.1	Transforms with Fuzzy Logic-Based Medical Image Fusion	228
12.8.2.2	Transforms With Guided Image Filtering-Based Medical Image Fusion	229
12.8.2.3	Transforms With Neural Network-Based Image Fusion	229
12.9	Conclusions	233
12.9.1	Future Work	234
	References	234
<b>13</b>	<b>Health Detection System for COVID-19 Patients Using IoT</b>	<b>237</b>
<i>Dipak P. Patil, Kishor Badane, Amit Kumar Mishra and Vishal A. Wankhede</i>		
13.1	Introduction	237
13.1.1	Overview	237
13.1.2	Preventions	238
13.1.3	Symptoms	238
13.1.4	Present Situation	238
13.2	Related Works	239
13.3	System Design	239
13.3.1	Hardware Implementation	239
13.3.1.1	NodeMCU	240
13.3.1.2	DHT 11 Sensor	240
13.3.1.3	MAX30100 Oxygen Sensor	241
13.3.1.4	ThingSpeak Server	242
13.3.1.5	Arduino IDE	243
13.4	Proposed System for Detection of Corona Patients	245
13.4.1	Introduction	245
13.4.2	Arduino IDE	246
13.4.3	Hardware Implementation	246
13.5	Results and Performance Analysis	247
13.5.1	Hardware Implementation	247
13.5.1.1	Implementation of NodeMCU With Temperature Sensor	247
13.5.2	Software Implementation	248
13.5.2.1	Simulation of Temperature Sensor With Arduino on Proteus Software	248
13.5.2.2	Interfacing of LCD With Arduino	250
13.6	Conclusion	250
	References	250
<b>14</b>	<b>Intelligent Systems in Healthcare</b>	<b>253</b>
<i>Rajiv Dey and Pankaj Sahu</i>		
14.1	Introduction	253
14.2	Brain Computer Interface	255

14.2.1	Types of Signals Used in BCI	256
14.2.2	Components of BCI	257
14.2.3	Applications of BCI in Health Monitoring	258
14.3	Robotic Systems	258
14.3.1	Advantages of Surgical Robots	258
14.3.2	Centralization of the Important Information to the Surgeon	259
14.3.3	Remote-Surgery, Software Development, and High Speed Connectivity Such as 5G	260
14.4	Voice Recognition Systems	260
14.5	Remote Health Monitoring Systems	260
14.5.1	Tele-Medicine Health Concerns	262
14.6	Internet of Things-Based Intelligent Systems	262
14.6.1	Ubiquitous Computing Technologies in Healthcare	264
14.6.2	Patient Bio-Signals and Acquisition Methods	265
14.6.3	Communication Technologies Used in Healthcare Application	267
14.6.4	Communication Technologies Based on Location/Position	269
14.7	Intelligent Electronic Healthcare Systems	270
14.7.1	The Background of Electronic Healthcare Systems	270
14.7.2	Intelligent Agents in Electronic Healthcare System	270
14.7.3	Patient Data Classification Techniques	271
14.8	Conclusion	271
	References	272
15	<b>Design of Antennas for Microwave Imaging Techniques</b>	275
	<i>Dnyaneshwar D. Ahire, Gajanan K. Kharate and Ammar Muthana</i>	
15.1	Introduction	275
15.1.1	Overview	276
15.2	Literature	277
15.2.1	Microstrip Patch Antenna	278
15.2.2	Early Detection of Breast Cancer and Microstrip Patch Antenna for Biomedical Application	279
15.2.3	UWB for Microwave Imaging	279
15.3	Design and Development of Wideband Antenna	280
15.3.1	Overview	280
15.3.2	Design of Rectangular Microstrip Patch Antenna	281
15.3.3	Design of Microstrip Line Feed Rectangular Microstrip Patch Antenna	283
15.3.4	Design of Microstrip Line Feed Rectangular Microstrip Patch Antenna With Partial Ground	285
15.3.5	Key Shape Monopole Rectangular Microstrip Patch Antenna With Rounded Corner in Partial Ground	286
15.4	Results and Inferences	290
15.4.1	Overview	290
15.4.2	Rectangular Microstrip Patch Antenna	290
15.4.2.1	Reflection and VSWR Bandwidth	290
15.4.2.2	Surface Current Distribution	291

15.4.3	Microstrip Line Feed Rectangular Microstrip Patch Antenna With Partial Ground	292
15.4.3.1	Reflection and VSWR Bandwidth	292
15.4.3.2	Surface Current Distribution	292
15.4.3.3	Inference	293
15.4.4	Key Shape Monopole Rectangular Microstrip Patch Antenna with Rounded Corner in Partial Ground	294
15.4.4.1	Reflection and VSWR Bandwidth	294
15.4.4.2	Surface Current Distribution	294
15.4.4.3	Results of the Fabricated Antenna	295
15.4.4.4	Inference	296
15.5	Conclusion	297
	References	298
<b>16</b>	<b>COVID-19: A Global Crisis</b>	<b>303</b>
<i>Savita Mandan and Durgeshwari Kalal</i>		
16.1	Introduction	303
16.1.1	Structure	304
16.1.2	Classification of Corona Virus	304
16.1.3	Types of Human Coronavirus	304
16.1.4	Genome Organization of Corona Virus	305
16.1.5	Coronavirus Replication	305
16.1.6	Host Defenses	306
16.2	Clinical Manifestation and Pathogenesis	306
16.2.1	Symptoms	307
16.2.2	Epidemiology	307
16.3	Diagnosis and Control	308
16.3.1	Molecular Test	308
16.3.2	Serology	308
16.3.3	Concerning Lab Assessments	309
16.3.4	Significantly Improved D-Dimer	309
16.3.5	Imaging	309
16.3.6	HRCT	309
16.3.7	Lung Ultrasound	310
16.4	Control Measures	310
16.4.1	Prevention and Patient Education	311
16.5	Immunization	312
16.5.1	Medications	312
16.6	Conclusion	313
	References	313
<b>17</b>	<b>Smart Healthcare for Pregnant Women in Rural Areas</b>	<b>317</b>
<i>D. Shanthi</i>		
17.1	Introduction	317
17.2	National/International Surveys Reviews	319
17.2.1	National Family Health Survey Review-11	319

17.2.2	National Family Health Survey Review-2.2	319
17.2.3	National Family Health Survey Reviews-3	320
17.3	Architecture	320
17.4	Anganwadi's Collaborative Work	321
17.5	Schemes Offered by Central/State Governments	321
17.5.1	AAH (Anna Amrutha Hastham)	321
17.5.2	Programme Arogya Laxmi	323
17.5.3	Balamrutham-Kids' Weaning Food from 7 Months to 3 Years	323
17.5.4	Nutri TASC (Tracking of Group Responsibility for Services)	323
17.5.5	Akshyapatra Foundation (ISKCON)	324
17.5.6	Mahila Sishu Chaitanyam	324
17.5.7	Community Management of Acute Malnutrition	325
17.5.8	Child Health Nutrition Committee	325
17.5.9	Bharat Ratna APJ Abdul Kalam Amruth Yojna	325
17.6	Smart Healthcare System	326
17.7	Data Collection	328
17.8	Hardware and Software Features of HCS	328
17.9	Implementation	329
17.9.1	Modules	329
17.9.2	Modules Description	329
17.9.2.1	Data Preprocessing	329
17.9.2.2	Component Features Extraction	329
17.9.2.3	User Sentimental Measurement	330
17.9.2.4	Sentiment Evaluation	330
17.10	Results and Analysis	331
17.11	Conclusion	333
	References	333
<b>18</b>	<b>Computer-Aided Interpretation of ECG Signal—A Challenge</b>	<b>335</b>
	<i>Shalini Sahay and A.K. Wadhwanı</i>	
18.1	Introduction	336
18.1.1	Electrical Activity of the Heart	336
18.2	The Cardiovascular System	338
18.3	Electrocardiogram Leads	340
18.4	Artifacts/Noises Affecting the ECG	342
18.4.1	Baseline Wander	343
18.4.2	Power Line Interference	343
18.4.3	Motion Artifacts	344
18.4.4	Muscle Noise	344
18.4.5	Instrumentation Noise	344
18.4.6	Other Interferences	345
18.5	The ECG Waveform	346
18.5.1	Normal Sinus Rhythm	347
18.6	Cardiac Arrhythmias	347
18.6.1	Sinus Bradycardia	347
18.6.2	Sinus Tachycardia	348

xvi CONTENTS

18.6.3	Atrial Flutter	348
18.6.4	Atrial Fibrillation	349
18.6.5	Ventricular Tachycardia	349
18.6.6	AV Block 2 First Degree	350
18.6.7	Asystole	350
18.7	Electrocardiogram Databases	351
18.8	Computer-Aided Interpretation (CAD)	351
18.9	Computational Techniques	354
18.10	Conclusion	356
	References	357
	<b>Index</b>	<b>359</b>

## 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

# Machine Learning Approach for Medical Diagnosis Based on Prediction Model

Hemant Kasturiwale<sup>1\*</sup>, Rajesh Karhe<sup>2</sup> and Sujata N. Kale<sup>3</sup>

<sup>1</sup>*Thakur College of Engineering and Technology, Kandivali (East), Mumbai, MS, India*

<sup>2</sup>*Shri Gulabrai Deokar College of Engineering, Jalgaon, MS, India*

<sup>3</sup>*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

## 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:

- processing the enrolment database (PEP) and
- 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

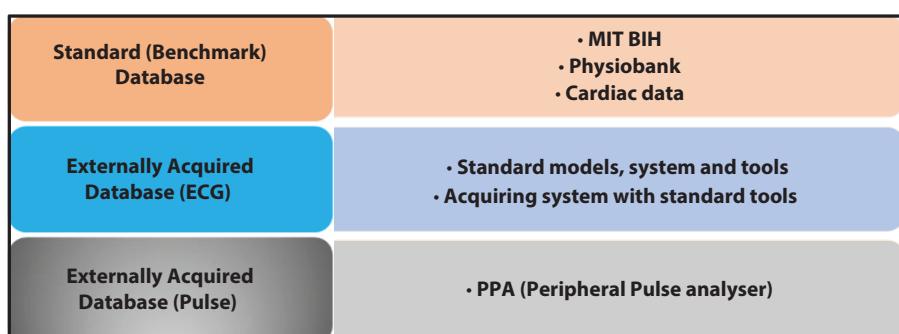


Figure 1.1 Acquisition system and sources [source: 14].

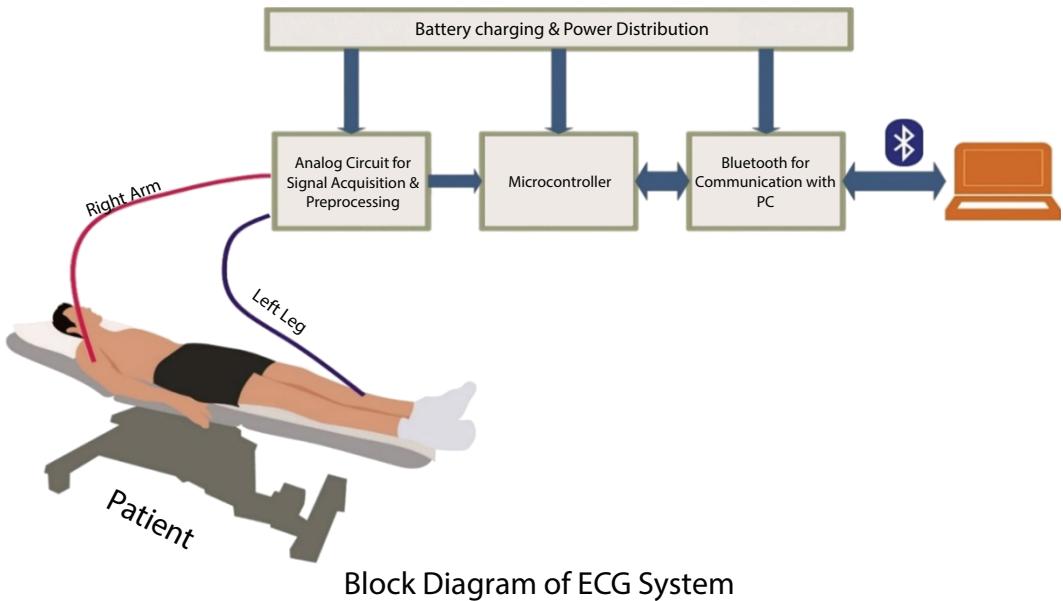


Figure 1.2 ECG acquisition system with connection (low cost).

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

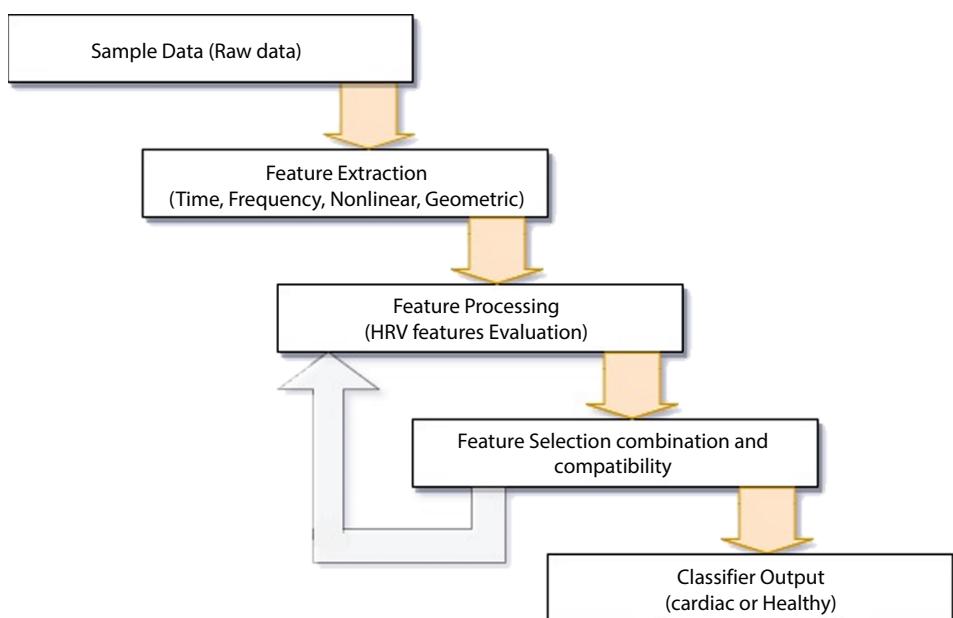


Figure 1.3 Cardiac diseases identification model for cardiac diseases [source: 24].

### 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
<b>Time domain (Group 1)</b>		
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
<b>Frequency domain (Group 2)</b>		
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
<b>Nonlinear (Group 3)</b>		
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
<b>Other features (parameters/indices) (Group 4)</b>		
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

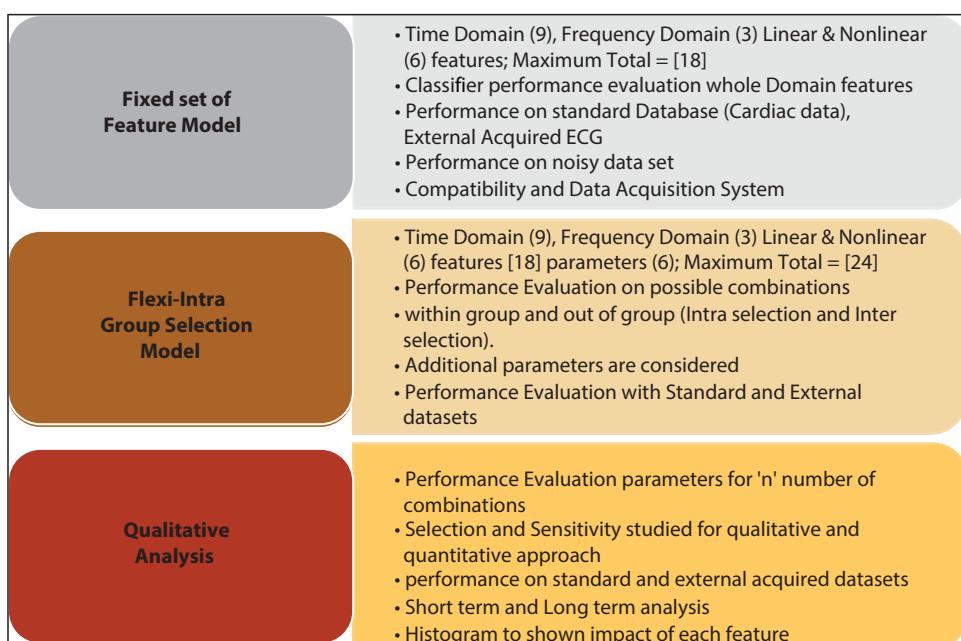
### 1.3.3 Proposed Model With Layer Representation

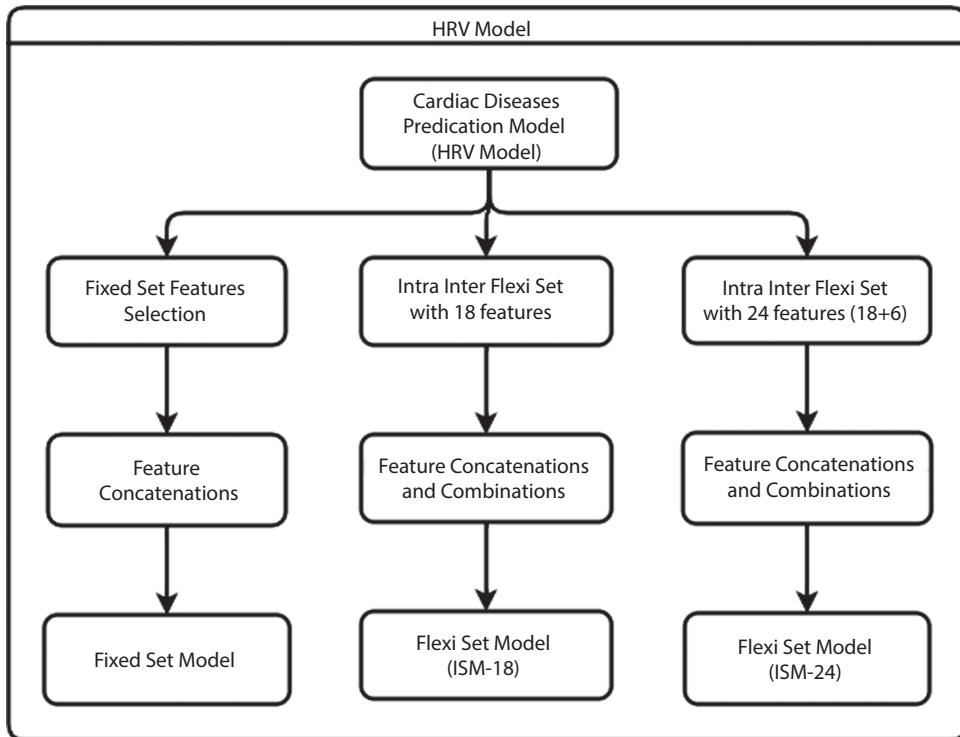
The HRV analysis for cardiac diseases is complicated, so the step-by-step processes are defined as a part of HRV analysis. The research work has come up with the development of a robust model via the layer model. The features contribution and impact is a significant

**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