

## TABLE OF CONTENTS

---

<b>CHAPTER NO.</b>	<b>TITLE</b>	<b>PAGE NO.</b>
<b>1</b>	<b>INTRODUCTION</b>	<b>1</b>
	1.1 Heart Diseases	1
	1.2 Heart Disease - Causes And Risk Factors	2
	1.3 Diagnosis of CHD	4
	1.4 Treatment of CHD	5
	1.5 Prevention of CHD	7
	1.6 Machine Learning Overview	7
	1.6.1 Types of Machine Learning	8
	1.6.1.1 Classification	9
	1.6.1.2 Clustering	10
	1.6.1.3 Regression	12
	1.6.1.4 Association Analysis	12
	1.7 Application of Classification Methods in the Healthcare Sector	13
	1.8 Feature Selection	14
	1.8.1 Filter method	14
	1.8.2 Wrapper Methods	15
	1.8.3 Embedded Methods	18
	1.9 Ensemble Classification	19
	1.10 Overview of the Machine Learning Process	20
	1.11 Research Motivation	21
	1.12 Problem Statement	22
	1.13 Aim of the Research	22
	1.14 Objectives of the Research	23
	1.15 Contributions of the Research	23
	1.16 Organization of the Thesis	24
	1.17 Chapter Summary	25
<b>2</b>	<b>LITERATURE SURVEY</b>	<b>26</b>
	2.1 Introduction	26
	2.2 Related Works	26

---

<b>CHAPTER NO.</b>	<b>TITLE</b>	<b>PAGE NO.</b>
	2.2.1 Review of Feature Selection Methods for Disease Classification	26
	2.2.2 Review of Machine Learning Methods for Disease Classification	34
	2.2.3 Study on Optimization of classifiers for disease classification	39
	2.2.4 Study on Ensemble Classification for Diseases prediction	44
	2.3 Research Gap	51
	2.4 Chapter Summary	51
<b>3</b>	<b>RESEARCH METHODOLOGY</b>	<b>52</b>
	3.1 Overview	52
	3.2 Proposed Methodology	52
	3.2.1 Selection of Highly Relevant Features for Heart Disease Prediction	52
	3.2.1.1 Feature Selection Using Importance Scores of Features by Gradient Boosting Classifiers	53
	3.2.1.2 Feature Selection by ModifiedBoostARoota and Heart Disease Prediction Using Classification Models	53
	3.2.2 A Novel Super Learner Ensemble Learning Model in Heart Disease Prediction	53
	3.2.3 An Optimized Super Learner Ensemble Model Using Whale Optimization Algorithm	54
	3.3 Methodology Flow	55
	3.4 Dataset Description	55
	3.5 Performance Metrics	60
	3.5.1 Accuracy	60
	3.5.2 Precision	61
	3.5.3 Recall	61
	3.5.4 F-measure	61
	3.6 Chapter Summary	61

<b>CHAPTER NO.</b>	<b>TITLE</b>	<b>PAGE NO.</b>
<b>4</b>	<b>FEATURE SELECTION USING IMPORTANCE SCORES OF FEATURES FOR HEART DISEASE PREDICTION</b>	<b>62</b>
	4.1 Introduction	62
	4.1.1 Feature Importance (FI)	62
	4.1.2 Feature Selection Using Importance Score of Features	63
	4.1.3 Machine Learning Classifiers	64
	4.1.3.1 EXtreme Gradient Boosting (XGBoost)	64
	4.1.3.2 CatBoost	65
	4.1.3.3 Majority Vote Ensemble	65
	4.2 Methodology	66
	4.3. Analysis of FI Score types of Gradient Boosting Algorithms on Heart Datasets	67
	4.4 Evaluation of Performance Accuracy of the Classifiers	73
	4.5 Proposed Search Space of Feature Subsets	77
	4.6 Evaluation of Performance Metrics of the Classifiers	77
	4.7 Comparison of the Proposed and Existing Feature Selection Methods	78
	4.8 Chapter Summary	80
<b>5</b>	<b>FEATURE SELECTION BY MODIFIEDBOOSTAROOTA FOR HEART DISEASE PREDICTION</b>	<b>82</b>
	5.1 Introduction	82
	5.2 Overview Of Feature Selection Algorithms	83
	5.2.1 Tree-based Feature Importance	84
	5.3 Proposed Methodology	85
	5.3.1 ModifiedBoostARoota (MBAR) Algorithm for Feature Selection	85
	5.3.2 Overall framework	86
	5.4 Results and Discussion	87
	5.4.1 Evaluation of MBAR on Low Dimensional Heart Datasets	87
	5.4.2 Evaluation of MBAR on High-Dimensional Heart Datasets	90
	5.5 Chapter Summary	94

<b>CHAPTER NO.</b>	<b>TITLE</b>	<b>PAGE NO.</b>
<b>6</b>	<b>SUPER LEARNER ENSEMBLE MODEL IN HEART DISEASE PREDICTION</b>	<b>95</b>
	6.1. Introduction	95
	6.2. Machine Learning Models	96
	6.2.1 Support Vector Machine (SVM)	96
	6.2.2 Logistic Regression (LR)	96
	6.2.3 Gaussian Naive Bayes (GNB)	96
	6.2.4 K-Nearest Neighbors (KNN)	97
	6.2.5 Decision Tree (DT)	97
	6.2.6 Ensemble Classifiers	98
	6.2.6.1 Random Forest (RF)	98
	6.2.6.2 Extreme Gradient boosting (XGB)	98
	6.2.6.3 CatBoost (CatB)	99
	6.2.6.4 Majority Voting Ensemble (MVE):	99
	6.2.6.5 Stacking	99
	6.2.7 Advantages of opting Ensemble classifier over Individual classifier	100
	6.3 Proposed Methodology	101
	6.3.1 Proposed Super Learner for classification of Heart disease	103
	6.4 Results and Discussion	105
	6.4.1 Performance Evaluation on Low and High Dimensional Heart Datasets	105
	6.4.1.1 Evaluation of Classification Accuracy on Low-Dimensional Heart Datasets With All Features	106
	6.4.1.2 Evaluation of Classification Accuracy on Low-Dimensional Heart Datasets With Chosen Features	107
	6.4.1.3 Evaluation of Performance Metrics of SLEM with MBAR on Low -Dimensional Heart Datasets	108
	6.4.1.4 Visualization of the performances of the classifiers on Low dimensional datasets	109
	6.4.1.5 Evaluation of Classification Accuracy on High-Dimensional Heart Datasets With All Features	111
	6.4.1.6 Evaluation of Classification Accuracy on high-dimensional heart datasets with Chosen features	111

<b>CHAPTER NO.</b>	<b>TITLE</b>	<b>PAGE NO.</b>
	6.4.1.7 Evaluation of Performance Metrics of SLEM with MBAR on High -Dimensional Heart Datasets	112
	6.4.1.8 Visualization of the Performances of the Classifiers on High Dimensional Datasets	113
	6.5 Chapter Summary	114
<b>7</b>	<b>OPTIMAL SELECTION OF BASE CLASSIFIERS IN AN ENSEMBLE USING WHALE OPTIMIZATION ALGORITHM AND DIVERSITY MEASURES</b>	<b>115</b>
	7.1 Introduction	115
	7.2 Overview of Optimizing Ensemble Model for Classification Methods	116
	7.3 Proposed Methodology	118
	7.3.1 Constructing the base classifiers	118
	7.3.2 Utilizing WOA for selecting the subset of classifiers	118
	7.3.3 Diversity measures	123
	7.3.4 Overall framework for the proposed OSLEM model	124
	7.4 Results and Discussion	125
	7.4.1 Analysis of OSLEM model on Low-Dimensional Heart Datasets	125
	7.4.2 Analysis of OSLEM model on High-Dimensional Heart Datasets	127
	7.4.3 Visualization of the performances of the classifiers on both low and high dimensional heart datasets	128
	7.5 Chapter Summary	129
<b>8</b>	<b>PERFORMANCE STUDY OF PROPOSED MODEL AND EXISTING MODELS ON HEART DATASETS</b>	<b>130</b>
	8.1 Evaluation of Proposed Model and Related Works on Cleveland Heart Dataset	130
	8.2 Evaluation of Proposed Model and Related Works on Statlog Heart Dataset	132
	8.3 Evaluation of Proposed Model with Related Works on Arrhythmia Heart Dataset	133

---

<b>CHAPTER NO.</b>	<b>TITLE</b>	<b>PAGE NO.</b>
	8.4 Evaluation of Proposed Model with Related Works on Z- Alizadeh Sani Heart Dataset	136
	8.5 Limitations	136
	8.6 Challenges	137
	8.7 Chapter Summary	137
<b>9</b>	<b>CONCLUSION AND FUTURE ENHANCEMENTS</b>	<b>138</b>
	9.1 Conclusion	138
	9.2 Future Work	140
	<b>REFERENCES</b>	<b>142</b>
	<b>APPENDICES</b>	
	<b>PUBLICATION</b>	
	<b>PLAGIARISM REPORT</b>	

---