

Development of Cardiovascular Disease Prediction System

*A Thesis Submitted in fulfilment
of the requirements for the award of the degree of*

DOCTOR OF PHILOSOPHY IN BIOTECHNOLOGY



THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Submitted by

RENU NARAIN

(Registration No:- 950800007)

Under the supervision of

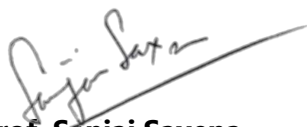
Dr. SANJAI SAXENA
Professor
Thapar Institute of Engineering
& Technology, Patiala, Punjab

Dr. ACHAL KUMAR GOYAL
Professor
Gurukul Kangri University
Haridwar, Uttarakhand

**DEPARTMENT OF BIOTECHNOLOGY
THAPAR INSTITUTE OF ENGINEERING & TECHNOLOGY
(Deemed to be University)
PATIALA -147004, PUNJAB, INDIA
July, 2018**

Certificate

It is hereby certified that the thesis "**DEVELOPMENT OF CARDIOVASCULAR DISEASE PREDICTION SYSTEM**" which is submitted by **Mrs. Renu Narain (Regd. No. 950800007)**, in fulfilment of the requirement for the award of the degree of **Doctor of Philosophy** in the Department of Biotechnology, Thapar Institute of Engineering and Technology (Deemed University), Patiala, Punjab, India, is a record of the candidate's own independent and original research work carried out by her under my supervision and guidance. The matter embodied in this thesis has not been submitted in part or full to any other University or Institute for the award of any degree in India or abroad.



Prof. Sanjai Saxena

Professor

Department of Biotechnology,

Thapar Institute of Engineering and Technology,

Patiala, Punjab.

Supervisors



Prof. Achal Kumar Goyal

Head, University Computer Center,

Gurukul Kangri University,

Haridwar, Uttarakhand.

Candidate's Declaration

I hereby declare that the work which is being presented in the thesis "**DEVELOPMENT OF CARDIOVASCULAR DISEASE PREDICTION SYSTEM**" submitted by me for the award of the degree of **Doctor of Philosophy** in the Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala, is true and original record of my own independent and original research work carried out under the kind supervision of Prof. Sanjai Saxena, Department of Biotechnology, Thapar Institute of Engineering and Technology (Deemed University), Patiala, India and Prof. Achal Kumar Goyal, Head, University Computer Center, Gurukul Kangri University, Haridwar, Uttarakhand. The matter embodied in this thesis has not been submitted in part or full to any other University or Institute for the award of any degree in India or abroad.

Date: 26th July, 2018

Renu Narain
(Renu Narain)

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisors for their valuable guidance during the research work. As my supervisor, I would also express my great appreciation with regards to Dr. Sanjai Saxena for his valuable guidance and suggestions helped me to establish the overall direction for the research and contribution immensely for the success of this research work. As my supervisor his insight, observation and quick guidance helped me to establish the overall direction for the research and contribution immensely. His broad knowledge, interest, enthusiasm and constant encouragement has been essential in my development as a researcher. Foremost, I express my gratitude to my Ph.D supervisor Dr. Achal Kumar Goyal, Head, University Computer Center, Gurukul Kangri University, Haridwar, Uttarakhand India. During this research his continuous guidance nurtured me a lot. The valuable knowledge, which I gained from him, is valuable for me. I am thankful for his patient guidance and encouragement during this doctoral degree.

I am grateful to the Thapar University, Patiala, Punjab, INDIA, especially to the Department of Biotechnology for providing me the opportunity to do this research work.

I am very thankful to the Director, Thapar University for providing me university resources and facilities necessary to carry out this research work.

I am also grateful to Prof. O.P Pandey, Dean (Research and Sponsored Projects) for his continuous support, guidance and providing me university resources and facilities necessary to carry out this research work.

I am also grateful to Dr. M Ghosh, Head, Department of Biotechnology, for her valuable guidance, help and assistance, given to me during the research work.

I express my gratitude to the Doctoral Committee comprising Dr. N. Tejo Prakash, Head, School of Energy and Environment, Dr. M.S. Reddy and Dean (Research and Sponsored Projects) for monitoring the progress and providing valuable suggestions for

improvement of my research work. I am also thankful to the entire faculty and staff of Department of Biotechnology for healthy discussions, suggestions and necessary cooperation to complete this work.

With deepest gratitude I dedicate my PhD research work to my Father, Er. Sushil Kumar Sharma, my Mother Mrs. Asha Sharma, my Aunt Mrs. Sarla Sharma and my husband Mr. Anshul Goswami for the unconditional continuous support and guidance given. Without them this would have been impossible to successfully complete this work.

I must give big thanks to my beloved brothers Dr. Ravi Narayan, Er. Rohit Narain and sister-in-law Dr. Mansi Arora for their continuous help in editing this thesis for successfully completing this work.

I also want to give thanks to my son Vinayak Goswami for his understandability during the days I ignored him due to busy in this research work.

Above all, I pay my reverence to the almighty God as well as to my beloved grandparents (Late.) Mr. Ram Narain Sharma and (Late.) Mrs. Rajeshwari Devi, whose blessings bring this attainment.

Renu Narain
(Renu Narain)

CONTENTS

Certificate.....	i
Candidate's Declaration.....	ii
Acknowledgements.....	iii
Contents.....	v
List of Figures.....	viii
List of Tables.....	ix
Notations and Abbreviations.....	xi
Executive Summary.....	xiii
Chapter 1: Introduction.....	1
1.1. Cardiovascular Diseases.....	3
1.2. Information technology and health care sector.....	5
1.3. Existing cardiovascular disease prediction models with their drawbacks.	5
1.4. Research Gap.....	9
1.5. Aim of the Research.....	11
1.6. Objectives of the Research.....	14
1.7. Organization of the Thesis.....	15
Chapter 2: Literature Review.....	18
2.1. Information Technology and Healthcare.....	18
2.2. Cardiovascular Disease Prediction Systems/ Scores.....	19
2.2.1.Framingham Risk Score.....	19
2.2.2.The Joint British Cardiac Society (BCS) / British Hypertension Society (BHS) / British Hyperlipidemia Association (BHA) Risk Score.....	22

2.2.3.The European SCORE.....	22
2.2.4.FINRISK Model.....	23
2.2.5.Centres for disease control and prevention (CDC)	23
2.2.6.Weinstein’s model	23
2.2.7.CANFIS Model.....	24
2.2.8.Other Approaches	24
Chapter 3: Materials and Method.	33
3.1. Introduction.....	33
3.2. Risk Factors.....	33
3.3. Data Collection.....	44
3.3.1. Pre-processing of Data.....	45
3.4. Quantum Theory and Quantum Neural Network.....	45
3.4.1.Quantum Theory.....	46
3.4.2.Quantum Superposition Principle.....	46
3.4.3.Quantum Neural Network.....	47
3.5. Advantages of Quantum Neural Network.....	48
Chapter 4: Proposed Predictive and Adoptive Algorithm(s) for	
Atherosclerosis and Myocardial Infarction.....	49
4.1. Introduction.....	49
4.2. Methodology used to implement proposed Cardiovascular Disease Prediction System.....	49
4.2.1.Data Description - Patients Data.....	50
4.3. Architecture of Proposed Cardiovascular Disease Prediction System.....	53
4.3.1. Proposed Algorithm of QNN based Cardiovascular Disease	

Prediction System	56
4.4. Conclusion.....	60
Chapter 5: Results and Discussions.....	61
5.1. Introduction.....	61
5.2. Experiment and Results.....	61
5.2.1. Experiment with proposed Cardiovascular Disease Prediction System using Physical parameters.....	63
5.2.2. Experiment with proposed Cardiovascular Disease Prediction System using Physiological parameters.....	77
5.2.3. Experiment with proposed Cardiovascular Disease Prediction System using Clinical parameters.....	82
5.2.4. Experiment with proposed Cardiovascular Disease Prediction System using Physical, Physiological and Clinical parameters.....	84
5.2.5. Experiment with proposed Cardiovascular Disease Prediction System using the parameters used in Framingham Risk Score.....	89
5.3. Methods used for Evaluation and Validation of proposed QNN based Cardiovascular Disease Prediction System.....	96
5.4. Conclusion.....	98
Chapter 6: Conclusions and Future Scope.....	99
6.1. Conclusion.....	99
6.2 Future Scope.....	100
Bibliography.....	101
List of Publications.....	122

LIST OF FIGURES

1.1	Death Rate due to different Diseases	1
1.2	Atherosclerosis	4
1.3	Heart with Myocardial Infarction.....	5
1.4	Architecture diagram of proposed system.....	12
1.5	Activity diagram of proposed system.....	13
3.1	Risk Levels for Homocysteine.....	40
3.2	Risk Levels for Fibrinogen.....	41
3.3	Risk Levels for Platelet Count.....	42
3.4	Risk Levels for Aspartate transaminase.....	43
3.5	Risk Levels for Creatine kinase.....	43
3.6	Architecture of Quantum Neural Network.....	47
4.1	Data Partition Set.....	51
4.2	Bar Diagram showing percentage of patients with various parameters.....	53
4.3	Architecture diagram of proposed system.....	54
4.4	Activity diagram of proposed system.....	55
4.5	Architecture of Quantum Neural Network.....	56
4.6	Flow chart of Quantum Neural Network for Heart Disease Prediction System.....	59
5.1	Mean square Error versus number of hidden nodes in QNN for proposed CVD prediction System.....	62

LIST OF TABLES

1.1	Forecast of number of cases (both male and female) of CVD in India.....	2
1.2	Characteristics comparison and their drawbacks of various CVD prediction systems.....	8
1.3	Objective wise distribution of Chapters.....	14
3.1	Parameters used for the prediction of heart disease.....	33
3.2	Risk Levels for Total Cholesterol.....	35
3.3	Risk Levels for High Density Lipoprotein.....	35
3.4	Age wise Blood Pressure range.....	36
3.5	BMI range.....	37
3.6	Risk levels for C- Reactive Protein.....	39
3.7	Risk levels for Brain Natriuretic Peptide.....	39
3.8	Risk levels for Electrolytes.....	41
3.9	Risk levels for Troponin.....	44
4.1	Parameters used for the prediction of heart disease.....	50
4.2	Percentage level of CVD risk categories.....	52
5.1	Results of Performance parameter for QNN and ANN.....	62
5.2	Random Testing of different experimental values using Physical parameters (Part-1)	63

	(Part-2)	70
5.3	Impact ranking of CVD parameters affecting classification.....	77
5.4	Random Testing of different experimental values using Physiological parameters	
	(Part-1)	78
	(Part-2)	79
5.5	Impact ranking of CVD parameters affecting classification.....	81
5.6	Random Testing of different experimental values using Clinical parameters.....	82
5.7	Impact ranking of CVD parameters affecting classification.....	84
5.8	Random Testing of different experimental values using all parameters (Physical, Physiological and Clinical Parameters)	
	(Part-1)	85
	(Part-2)	86
	(Part-3)	88
5.9	Random Testing of different experimental values using the parameters used in Framingham Risk Score.	
	(Part-1)	90
	(Part-2)	93

NOTATIONS AND ABBREVIATIONS

AMI	Acute Myocardial Infarction
ANN	Artificial Neural Network
BMI	Body Mass Index
BNP	Brain Natriuretic Peptide
CHD	Coronary Heart Disease
CVD	Cardiovascular Disease
CDC	The Center for Disease Control
DBP	Diastolic Blood Pressure
DM	Diabetes Mellitus
FRS	Framingham Risk Score
HCY	Homocysteine
HDL	High Density Lipoprotein
LDL	Low Density Lipoprotein
MI	Myocardial Infarction
QNN	Quantum Neural Network
SBP	Systolic Blood Pressure
SGOT	Aspartate transaminase
TC	Total Cholesterol
WHO	World Health Organization

θ^r	Quantum Interval
r	Quantum Level
δ_j	Error Rate
η	Learning Rate
δ_k	Error Rate of Output layer
δ_j	Error rate of Hidden layer
O_j	Output of hidden layer
O_k	Output of Output layer
W_{ij}	Weights between Input and Hidden layers
W_{kj}	Weights between Hidden and Output layers
n_s	Number of Excitation levels

Executive Summary

This work presents the intelligent Cardiovascular Disease (CVD) prediction system based on machine learning, which uses Quantum Neural Network for machine learning. Early medical diagnosis of Heart disease is very important and should be performed accurately and efficiently. Unfortunately, the physicians don't have the enough time to analyze past history of patient in depth.

This Intelligent system would enhance the medical care and reduce costs, by quick analysis of past data of patients with percentage of risk prediction. The accuracy of this intelligent system is significantly higher than other existing prognostic systems. All available Physical, Physiological, Clinical parameters have been considered in this study. The data of 815 Patients suffering with the symptoms of Heart disease has been collected from hospital and used for training and evaluation. Furthermore, the dataset of famous Framingham study consisting 5209 CVD patients' data has been used for validation purpose. All the patients' reports have been diagnosed and analysed by medical practitioner previously. This system uses the Quantum Neural Network for machine learning. The results obtained have high degree of sensitivity and specificity that matches with the expert's opinion with 98.5% accuracy. Such an expert system would be very useful when incorporated with other systems to provide diagnostic and predictive medical opinions in a timely manner. This system will work as an aid to physician for prognosis of heart disease. Using this system, medical practitioners may plan better medication and treatment strategy. The overall accuracy of this intelligent heart disease prognostic system is 98.5%, which is significantly higher than other existing approaches.

CHAPTER 1

INTRODUCTION

Introduction

Cardiovascular diseases (CVD's) are the primary reason of major deaths around the globe as per WHO report (WHO, October 2008). As per this report, approximately 17.3 million deaths have been attributed to CVD's which account for 30% of all global deaths (WHO Report, October 2005) (Fig1.1). This number is further expected to be increased to 23.6 million by 2030 (WHO Report, 2013).

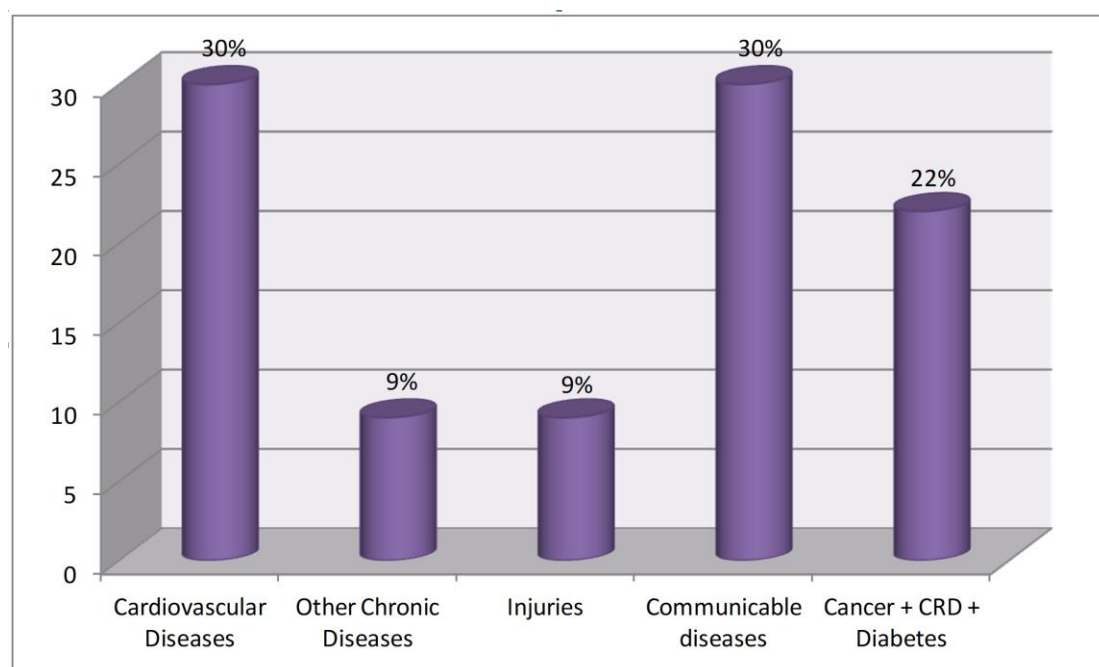


Figure 1.1: Death Rate due to different Diseases

The Center for Disease Control (CDC) and prevention also recognized heart diseases as the leading cause of death equally for men and women. According to their study in 2008 around half of the deaths due to heart disease were in men. In United States more than 616,000 people died due to various heart diseases and out of which 405,309 death were due to coronary heart disease in 2008. Overall, one fourth of deaths in United States happened due to heart disease (Minino, 2011). The forecast for Indian population also

Introduction

exhibits several cases comprising of both male and females suffering from coronary heart diseases in India is shown in Table 1.1 (Indrayan, 2005).

Table 1.1: Number of CVD cases (both male and female) in India

Year/ Gender		20-29 Years	30-39 Years	40-49 Years	50-59 Years	60-69 Years	Total
2000	Male	2,514,036	2,828,291	2,514,563	2,481,969	2,647,068	12,985,927
	Female	1,997,156	2,660,975	3,604,884	3,397,327	2,394,643	14,054,985
2005	Male	3,462,627	3,840,800	3,294,851	3,152,508	3,362,421	17,113,207
	Female	2,687,781	3,412,920	5,039,181	4,563,787	3,069,913	18,773,581
2010	Male	4,682,969	5,124,660	4,200,054	3,999,541	4,439,569	22,446,792
	Female	3,634,215	3,970,829	6,774,473	6,041,095	4,101,291	24,521,903
2015	Male	5,585,153	7,057,506	5,274,085	5,009,515	5,796,991	28,723,249
	Female	4,906,858	5,394,036	9,035,966	7,998,625	5,463,608	32,799,094

The main challenge for healthcare organizations lies in timely identification of conditions inducing risk of CVD. Currently the diagnosis are done on the basis of doctor's experience and a couple of clinical investigations, rather than by use of intelligent tools, which affects the decision leading to errors, unwanted biases. Due to this the medical service provided to patients is also affected. The tool may provide fast and accurate predictive information based on database and machine learning algorithms (Jolley et al. 2008, Robert et al. 2008).

As discussed, one of the major cause of deaths is delayed identification and medical diagnosis of conditions inducing risk of CVD.

The approach combining the computer based clinical decision support and patients record system was proposed by Wu, et al. (2002). The diagnosis is currently done based on the past experience and knowledge of the doctor, instead of intelligent tools. The main aim of this approach was to decrease the possibilities of error in doctors' diagnosis and decisions, which may also increase the probability of curing the diseases.

1.1 Cardiovascular Diseases

Worldwide, one of the main causes of deaths is CVD (WHO Report, 2013). The types of diseases under the umbrella of CVDs are myocardial infarction, Coronary Heart Disease (CHD), angina pectoris, stroke, hypertension and peripheral artery occlusive disease. In majority of cases the heart disease starts from atherosclerosis. Atherosclerosis is the condition in which cholesterol and fibrous tissue gets deposits on the walls of large and medium sized arteries. The structure of atherosclerotic plaques consists of proliferated smooth muscle cells and lipid deposit that gradually narrows the arteries. Due to this critical occlusion, the blood supply decreases or stops, which eventually leads to Myocardial Infarction, stroke or spectra of various CVD's (Mendis et. al. 2011)

Atherosclerosis occurs due to endothelial dysfunction, vascular inflammation, and deposition of lipids, cholesterol, calcium, and cellular debris within vessels. This results in formation of plaque, vascular remodeling, acute and chronic luminal obstruction leads to occlusion of blood flow and oxygen supply to myocardium and various other organ systems. Thus, it is the multi-factorial disease of arterial wall with varied etiopathogenesis in which the layer thickens as shown in Figure 1.2. It creates the regions of turbulent blood flow and hence the symptomatology. This phenomenon is very common in people with evidence of hypercholesterolemia (high level of cholesterol in the bloodstream) (Mendis et. al. 2011).

With the age, the quantity and thickness of plaques increases, resulting blood vessels lose its smooth lining and increases formation of blood clots. Also, the parts of blood clots may detach and form emboli. These emboli may flow through the bloodstream and block smaller vessels downstream leading to various other organ dysfunctions.

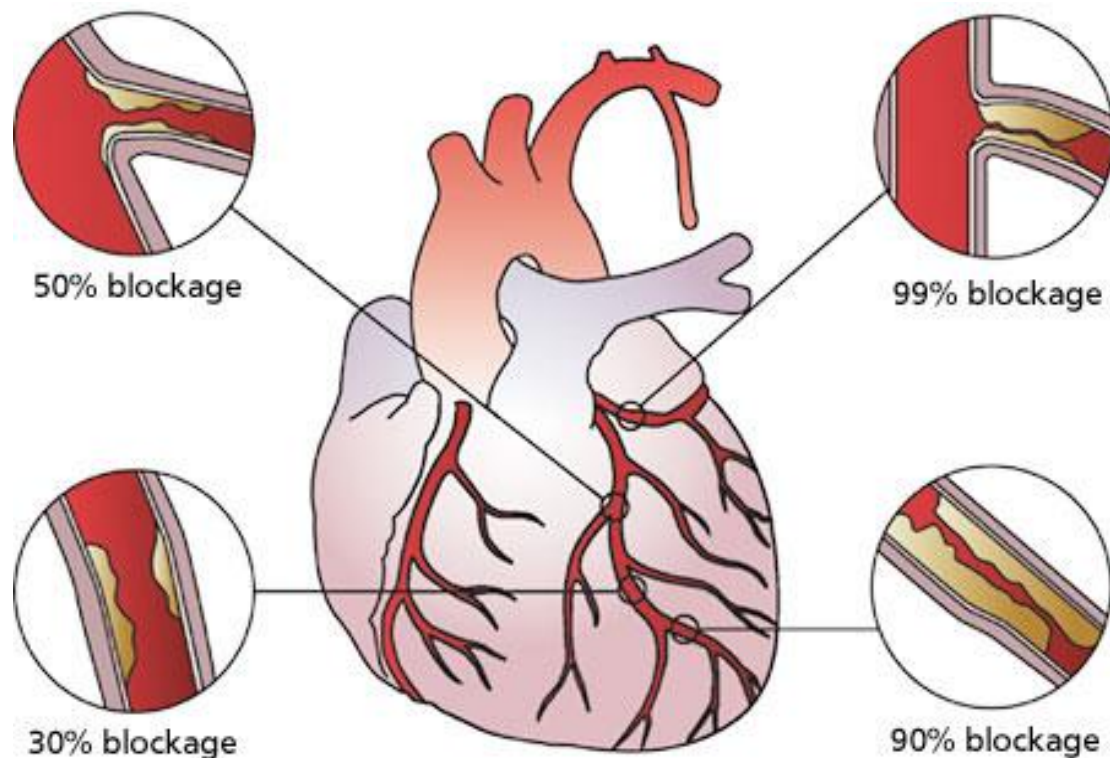


Figure 1.2: Atherosclerosis

Nowadays, due to advancement in medical imaging technology like Computerized Axial Tomography (CAT-scan) and Nuclear Magnetic Resonance (NMR-scanning), echocardiography, the medical practitioners are able to understand pathology and etiology of individual heart problems.

In CHD, the arteries to the heart muscle (myocardium) are narrowed. Due to this the heart gets less blood supply, which may cause chest pain (angina pectoris) or other symptoms, which generally starts at the time of physical exertion or any other stressful condition. If due to blood clot (thrombus) the vessel is completely occluded, area of the heart beyond that blockage may not get oxygenated blood and nourishment, which creates the permanent damage and/or death (infarction) of heart muscle tissue (myocardium). This stage is known as heart attack or Acute Myocardial Infarction (AMI) as shown in Figure 1.3. The pathogenesis of myocardial infarction may include Occlusive intracoronary thrombus as discussed above or other phenomena like Vasospasm or Embolization (Mendis et. al. 2011).

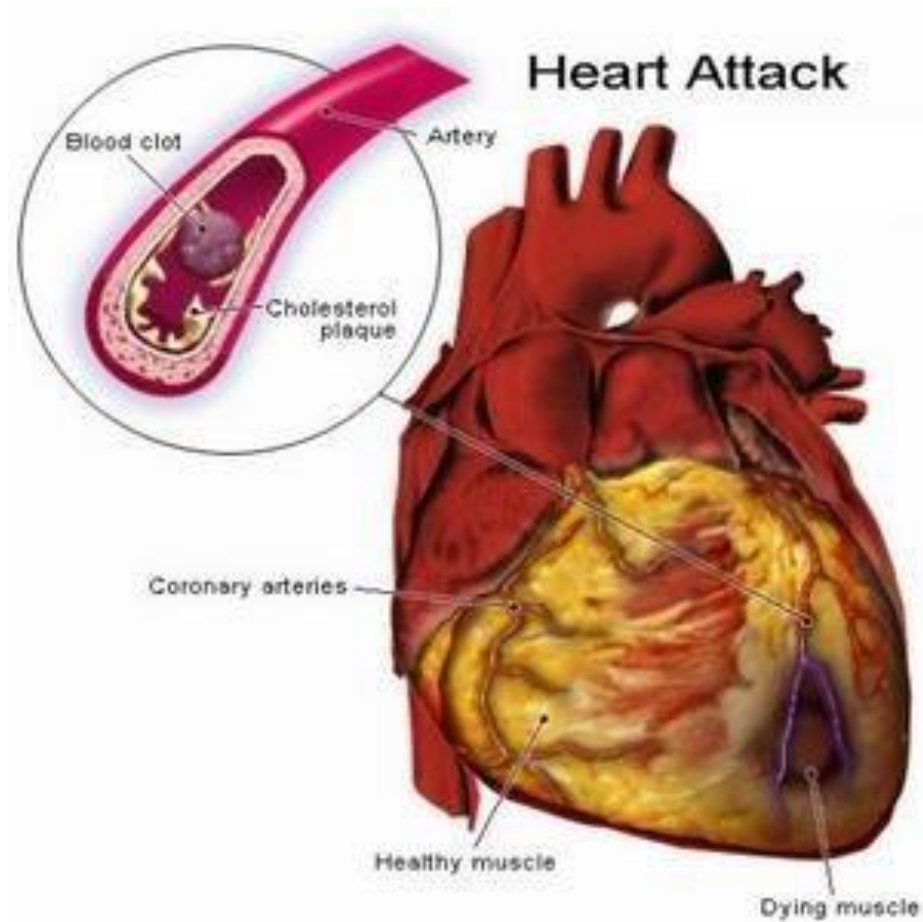


Figure: 1.3 Heart with Myocardial Infarction

1.2 Information Technology and health care sector

Nowadays several information tools are available which are used in order to increase the quality of services in health care sector. Information Technology (IT) has now become very popular in daily life, as it provides accessible, quick and accurate information. Due to recent advancements in IT, the artificially intelligent tools have become very popular for health care sector as well. Recently many researchers have done their research in the field of health care system and provided many tools/ systems like Patient Administration system (PAS) in UK, Electronic Health Care Record (EHCR) in Europe, American hospitals billing systems etc.

1.3 Existing Cardiovascular disease prediction Models with their drawbacks

Many simple strategies are available for assessment of "risk factors" for CVD. Some of the widely used methods are Framingham Risk Score (FRS), the Sheffield risk and the

European Systematic Coronary Risk Evaluation (SCORE) (Montgomery et. al. 2000, Mainous et. al. 2007). The parameters used by different models are given below in detail.

1.3.1 Framingham Risk Score (FRS)

The parameters considered in FRS are Gender, age, TC (Total Cholesterol), HDL (High Density Lipoprotein), Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), diabetes mellitus (DM), Body Mass Index (BMI) and smoking (Mainous et. al. 2007). The data group comprised of information of patients collected in 1960s and 1970s for analysis by Framingham equations. Since then, various risk factors and the lifestyle have changed markedly over a period of time. Moreover, Today's diverse population is entirely different from the study population of Framingham Heart Study which comprised mainly of European Americans and African Americans. Thus, the FRS is outdated for modern population and hence accuracy of FRS is now questionable (Gordon et. al. 1977).

1.3.2 The Joint British Cardiac Society (BCS) / British Hypertension Society (BHS) / British Hyperlipidemia Association (BHA) Risk Score:

The parameters used for this study were Gender, age, TC, HDL, SBP, DBP, DM, Smoking, ECG findings suggestive of LVH if available.

1.3.3 The European SCORE

According to the national guidelines, the European Heart SCORE model is used for treatment and primary prevention of cardiovascular disease (CVD) by several European countries. This study has been conducted for 10 years and the variables used for this study were age, gender, smoking, SBP and Total cholesterol level (Stovring et. al. 2013).

1.3.4 FINRISK Model

This model has been developed for the patients of European origin and it includes Diabetes and LDL as parameters. The FINRISK model was developed based on a 40 Years study, carried out every 5 years since 1972 in eastern and southwestern Finland (Bhopal et. al. 2005).

1.3.5 CDC Model

CDC (Centers for Disease Control and prevention) model based on Markov model was developed for heart disease prediction (Tesvat et al, 1991). As compared to other existing models the CDC model works throughout lifetime of patient, regardless an individual gets CHD or continues to stay "normal" (CDC Diabetes Cost-effectiveness Group, 2002, (Sonnenberg et. al. 1993)), even doesn't stops after first event also, assuming that one can stay in the health states "angina" and "history of CA/MI" for several years (Bhopal et. al. 2005).

1.3.6 Weinstein's model

This is one of the most extensive models and has five health states viz., persons without CHD, persons with new CHD, persons with CHD, persons who survived up to the age of 85 years, and death. Hence this model has two different absorbing states: survival at 85 years and death. Individuals in the state "new CHD" will only be there for first year, and after that transfer to "CHD", "survive until 85" or die (Weinstein et. al. 1987).

1.3.7 CANFIS Model

CANFIS model is based on Neuro-Fuzzy algorithm, due to its ability to adapt and classify. In this model the genetic algorithm has also been integrated with Neuro-Fuzzy approach for more accurate prediction of heart disease (Parthiban et al, 2008).

Although many researches and models have been developed for prediction of heart diseases, but still there is a need of a more accurate prediction system relevant to present day lifestyle. All the studies have been conducted around 30-50 years back and there has been a drastic difference in life style of the people today when compared to those in 1960s and 1970s, several new risk factors are identified which have drastically affected the occurrence or development of CHD. The existing prediction systems thus are outdated for present day population. The characteristics and drawbacks of the CVD prediction systems have been extensively compared in Table 1.2

Table 1.2: Characteristics comparison and their drawbacks of various CVD prediction systems

Various Models	Characteristics	Drawbacks
Framingham Risk Score	The variables for this study are Gender, age, Total Cholesterol (TC), HDL, SBP, DBP, DM, BMI and smoking. The Framingham Equations used are based on Cohort data collected in 1960s and 1970s.	This score is based on risk factors of 1960s and 1970s. The lifestyle has changed completely now. So, Validity of Framingham equation in modern populations may be undermined.
The Joint British Cardiac Society Score	The variables used for this study are Gender, age, Total Cholesterol, HDL, SBP, DBP, DM, Smoking, ECG-LVH if available.	Same drawbacks as that of Framingham Risk Score
The European SCORE	The variables used for this study are Gender, age, TC, TC/HDL, SBP and smoking.	Same drawbacks as that of Framingham Risk Score
FINRISK Model	It is based on study of European origin. It included Diabetes and LDL	Same drawbacks as that of Framingham Risk Score
CANFIS Model	CANFIS model was based on neural network and fuzzy logic and genetic algorithm. Due to these algorithms,	More than one algorithms were used. That's why it is time

	it had adaptive and quantitative features, which was used to diagnose the presence of the disease.	consuming and makes algorithm more complex.
Markov Model	It is a statistical Model. It satisfies the markov property, i.e. the future process always depends on current state. During process the observer tracks the state therefore the state transition is predictable and probabilities are the only parameters.	It is a straight chain model where states depend on number of variables the number of variables increase, the system will become much complex.
Bayesian Inference	It is based on Bayes' theorem to update the process. It is a statistical approach in which evidence or observations are used to update the process (Belder et al, 2001).	It requires us to specify a past knowledge on the parameters. Past knowledge is either vague, or non-existent, which makes it extremely complex to specify a unique previous distribution.

1.4 Research Gap

As highlighted in the previous section, lifestyle changes as well as identification of new risk factors in modern population has hampered the accuracy of prevailing cardiovascular prediction systems. Internationally, the most commonly used tool for evaluating the CVD is Framingham Risk Score (FRS). As discussed above, the Framingham equations have

Introduction

been developed on the data collected in 1960s to 1970s. The parameters used for this study are Gender, age, TC, HDL, SBP, DBP, DM, BMI and smoking (Doll et.al.1994, Steven, 2005). The way of living and other risk factors have markedly changed since then, which affected the rate of CVD. Moreover, various parameters e.g. serum fibrinogen, serum homocysteine etc and some newly identified risk factors, may distort the Framingham predictions to a greater extent. Hence, the FRS is outdated for modern population.

Similarly, other existing risk scores are also using the risk factors which are identical to those used in FRS. Some of these methods are machine learning based methods that require collection of input output data for training the neural and neuro-fuzzy systems. However, these approaches are not using the crucial physical and clinical risk factors for predicting the cardiovascular diseases such as Alcoholism, stress, Family history, high sensitivity C- Reactive Protein (hs-CRP), Brain Natriuretic Peptide (BNP), Homocysteine levels, Serum electrolytes, various enzymes etc.

To summarize, the main problems/drawbacks with the available CVD Prediction systems / scores have been identified based on the literature review are:

- The modern lifestyle is entirely different from that of 1960-1970 when most of the CVD prediction systems were developed, now these systems are outdated, that's why the accuracy of these systems is questionable in today's time. At present, no other system is available which uses the maximum crucial/ important available physical, physiological and clinical parameters/ risk factors at the same time. The newly available markers are so much accurate and highly specific which may lead to increasing the accuracy of CVD prediction by many folds, by including as risk factors in the parameters. Hence there is a need of a comprehensive, predictive and adoptive system for cardiovascular disease prediction.

1.5 Aim of Research

As there is a need of more rapid and accurate system for prediction of cardiovascular diseases and for a medical practitioner it is a tedious task to diagnose or predict the disease probability by analysis of all risk factors or symptoms manually.

To make our system effective and reliable, Quantum neural network has been incorporated which automatically learns and extracts significant patterns which are involved in development of the heart disease. For this we have created various datasets from the data collected from population of patients. The knowledge gained via patterns is applied in developing the automated system for improved prognosis of the cardiovascular disease. Hence the present research approach draws techniques from machine learning as well as an understanding of the underlying biological systems.

The Quantum Neural Network (QNN) based CVD prediction system is potential solution for development of highly accurate and quick CVD prediction system. The QNN has the ability to learn the patterns, more specifically the patterns of the heart disease. As the emphasis in the present research work is to enhance the accuracy of prediction using modern highly specific disease markers, QNN helps in learning the pattern of heart disease based on historical data. The proposed approach uses the QNN for prediction after learning the pattern of heart disease from historical data. The Architecture diagram of the proposed system is represented in Figure 1.4 and the conceptual activity diagram of the proposed CVD prediction system given in the Figure 1.5.

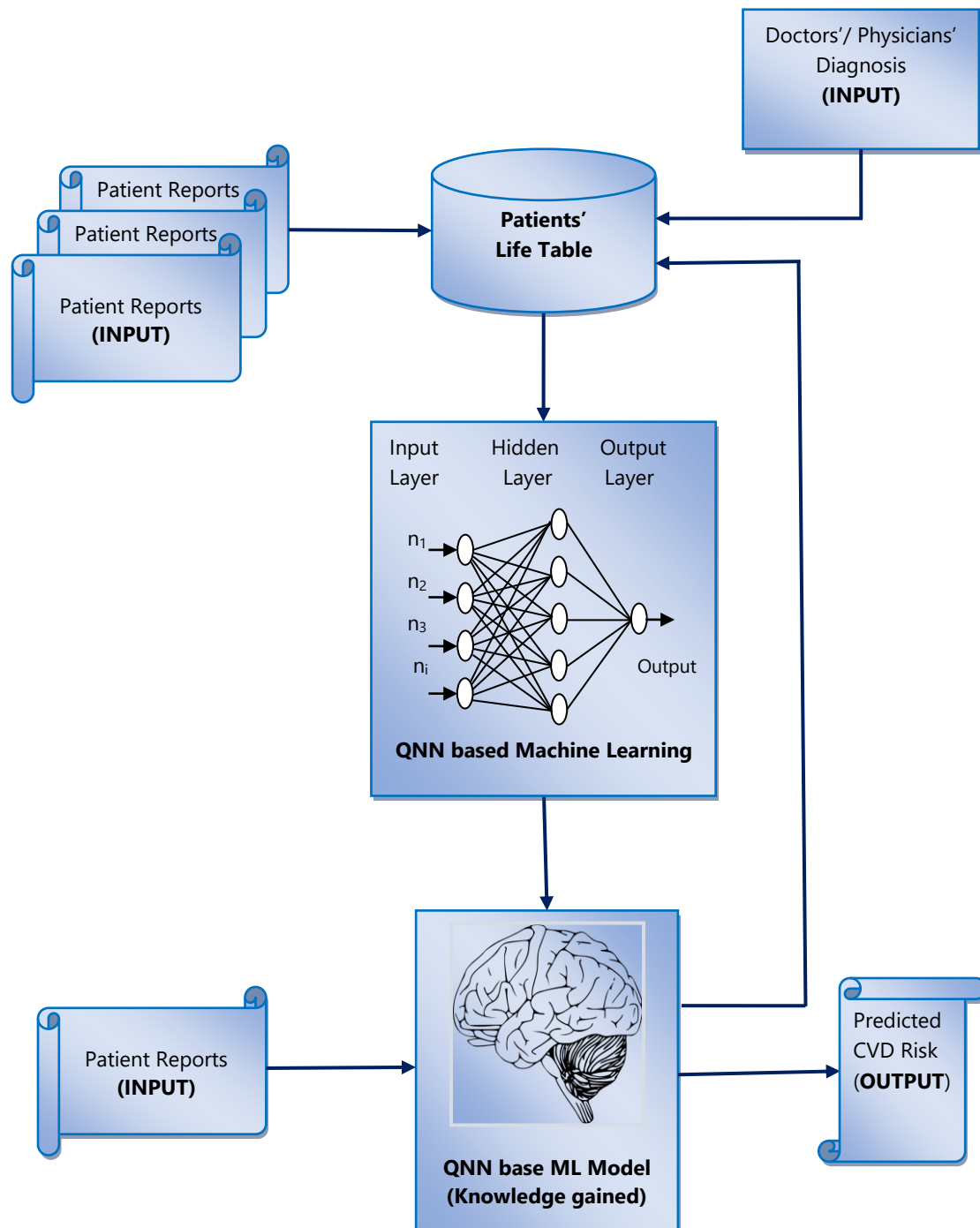


Figure 1.4: Architecture diagram of proposed system

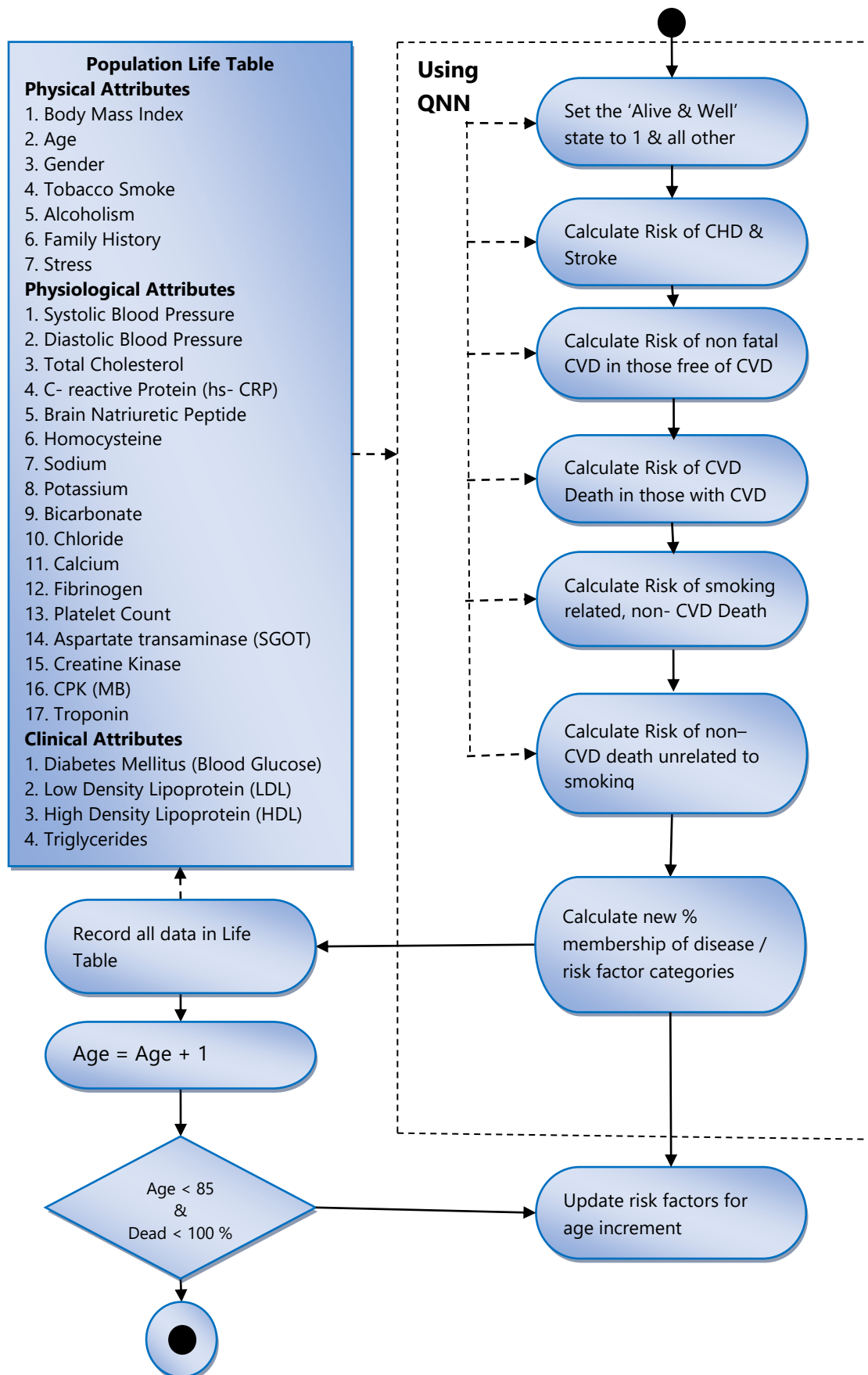


Figure 1.5: Activity diagram of proposed system

1.6 Objectives of the Research

In view of the above gaps, following were the main objectives of the present investigation:

Objective 1. To collect the Primary and Secondary Data of all parameters such as stress, alcoholism, LDL, Triglycerides, C- reactive protein, BNP to construct the Collation of the data.

Objective 2. To Develop the Predictive and Adoptive Algorithm(s) for Atherosclerosis and Myocardial Infarction.

Objective 3. To validate the final system.

All the objectives have been accomplished successfully and are explained in the thesis.

Table 1.3 shows the objective wise distribution of Chapters.

Table 1.3 Objective wise distribution of Chapters

Objective	Chapter
1) To collect the Primary and Secondary Data of all parameters such as stress, alcoholism, LDL, Triglycerides, C- reactive protein, BNP to construct the Collation of the data.	1) Introduction
	2) Literature Review
	3) Materials and Method.
2) To Develop the Predictive and Adoptive Algorithm(s) for Atherosclerosis and Myocardial Infarction.	4) Proposed Predictive and Adoptive Algorithm(s) for Atherosclerosis and Myocardial Infarction.
3) To validate the final system.	5) Results and Discussion
	6) Conclusions and Future Scope

1.7 Organization of the thesis

Chapter 1: Introduction

As mentioned above, the goal is to design a CVD prediction system based on QNN is to predict the cardiovascular disease in timely manner by clinicians, doctors, pathologists as well as by general public also. In First chapter, introduction about the subject along with the basic knowledge of CVD is given. It is known that the field of CVD prediction still has many problems to address, like the newly discovered parameters are not incorporated in the existing CVD prediction systems etc. This chapter explains the challenges and problems of CVD prediction systems.

Chapter 2: Literature Review

In this chapter, the literature survey was carried out, and also included extensive overview of various bioinformatics tools/ models/ approaches which have been recently developed for health care sector. Detailed review of CVD prediction systems has been given and discussed. In this chapter the technological advancements of different machine learning approaches are also mentioned.

Chapter 3: Materials and Method

Before explaining QNN model for CVD, this chapter addresses Risk Factors and Data set first and then the basic concepts of quantum theory with traditional artificial neural network. Quantum theory is the body of scientific principles which expresses the behaviour of matter and its relations with energy on the level of atoms and subatomic particles, and how these phenomena may possibly be connected to daily life. The fundamental principles of quantum theory are superposition principle, measurement principle and unitary evolution. QNN has been developed by various researchers following the basics of quantum theory. As such the QNN training method, architecture etc, are discussed in this chapter.

Chapter 4: Proposed Predictive and Adoptive Algorithm(s) for Atherosclerosis and Myocardial Infarction

This chapter includes details of QNN model for CVD. The complete architecture of the model is developed. Moreover, the method used to implement the CVD by using the QNN is also included. The mechanism used to categorise the CVD risk level has been given in this chapter.

Chapter 5: Results and Discussion

In tests and Results, which are performed on the proposed system for validation with a highlight on its underlying ideas and intrinsic advantages. The performance of proposed system is evaluated and validated on data of patients with the Doctor's diagnosis (predictions). The accuracy achieved by the proposed system is 98.88% which is considerably higher in comparison with the existing approaches like Framingham Risk Score, European heart score etc. In present condition, the accuracy of the Framingham Risk Score is only 19.22%. The major problems with these studies are their ineffectiveness after certain values range like the Framingham risk score can work only for 20 to 100 age range patients and the maximum threshold limit of CVD risk is 30%. Similarly, in European heart score, if the blood pressure is more than 180 or cholesterol over 8 mmol/L, patient is automatically classified as high risk without taking into account the medication condition.

Chapter 6 Conclusions and Future Scope

In this study, all available clinical, physiological and physical parameters have been used, which are available now a days. The proposed model is based on machine learning, which is implemented using QNN. This system is capable to learn on historical data. The proposed system is able to identify the data patterns and based on that, it builds its own knowledge to predict the CVD risk accurately. This system has a built-in capability to learn and adopt the current data pattern to correct itself over the time, as the data increases.

Introduction

The testing and validation have been performed using the data of Indian and American patients. The proposed system has achieved the highest accuracy among existing systems.

Thus, in the present work an intelligent Prognostic System for heart diseases based on Quantum Neural Network has been proposed here. During the experiments the reasonable accuracy of 98.88% has been achieved, which is considerably higher than other existing system. All the results have been verified by the physicians and are found correct. This system will assist the doctors and medical practitioners to provide the better medication. This intelligent aid will increase the doctor's performance and ability to early diagnosis of cardiovascular diseases in any patient.

CHAPTER 2

LITERATURE SURVEY

Literature survey

Literature survey was carried out for the study related to Cardiovascular Disease prediction system, other available information technology tools/ approaches for health care sector, quantum neural network and other machine learning algorithms. In this chapter, various models of Cardiovascular Disease prediction system are explored. The technological advancement in the field of information technology and various predictive systems for health care sector are identified.

2.1. Information Technology and Healthcare

Now a days many information tools are available to increase the quality of services in health care sector. Information Technology (IT) has now become very popular in daily life. IT provides easy to access, quick and accurate information. Due to recent advancement in IT, the artificially intelligent tools now become very popular for health care sector as well. Recently many researchers have done their research in the field of health care and have provided many tools/ systems like Health care information system, Automatic disease prediction, evaluation and Diagnostic systems, Imaging Systems etc. (Pavlopoulos et al., 1999; Sood et al., 2008; Shrimali et al., 2008, 2009; Gautam et al., 2018; Mastoi et al., 2018).

The Agency for Healthcare Research (AHR), 2006 has identified the usefulness of information technology in healthcare. According to their report, the Healthcare with IT will be more:

- Effective with quality
- Productive and efficient
- Accurate, procedurally correct and medically error free
- Cost effective
- Administratively efficient

- Available to real-time healthcare systems
- Accessible to affordable care

2.2. Cardiovascular Disease Prediction Systems/ Scores

Various scoring systems are available for assessment of “risk factors” of cardiovascular diseases. Some of the widely used methods are Framingham Risk Score (FRS), the Sheffield risk and the European Systematic Coronary Risk Evaluation (SCORE) (Mainous et al. 2007). The parameters used by different models are given below in detail.

2.2.1. Framingham Risk Score

The FRS was first described based on data obtained from the Framingham Heart Study, which was conducted in United States during a span of 10 years (1960 to 1970) to estimate the risk of developing coronary heart diseases. This study has been conducted using the cohort of 5209 patients, mainly aged between 30-62 Years. The parameters considered in this study were Gender, age, Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), diabetes mellitus (DM), Body Mass Index (BMI), total cholesterol (TC), High Density Lipoprotein (HDL) and smoking. The FRS uses the Cox-proportional hazards model for scoring (Mainous et al. 2007).

This scoring system gives an indication of the likely benefits of prevention, hence useful for both the individuals i.e. the patient and the clinician, as it helps in deciding whether preventive medical treatment and lifestyle modification will lead to decrease in incidence of CHD or not in future, individuals in low risk group have 10% or less risk of CHD at 10 years, with intermediate group having risk of 10-20%, and with high risk group having 20% or more chances of CHD.

In current scenario the FRS is outdated for modern population, as the lifestyle and habits of population now is entirely different from what it used to be a long ago. In FRS, only few parameters have been considered, and due to advancement in medical science many new parameters have been discovered which are very accurate. That's why the FRS may not be so much accurate and effective in present scenario. Moreover, the Scoring system

could overestimate (or underestimate) the risk among various group of populations mainly the non-US population (Capewell et al. 2010)

According to D' Agostino et al. (2008) the FRS does not perform well with all ethnicity, like with Native American women, Japanese American and Hispanic men. On the other hand, it performs well with white and blacks. The FRS overestimates the risk with North European population, which are in higher risk and on south European population with low risk.

The FRS also overestimates the risk with the Chinese population. This study was performed during a span of 10 years period in the Chinese group comprising of 30,121 adults aged between 35 to 65 years. They were diagnosed on the baseline parameters (Liu et al. 2004).

Artigao-Rodenas et al. (2013) proposed an algorithm based on the Framingham study which depends on three parameters, the first parameter included the risk factors with the highest weight (cholesterol, diabetes, blood pressure and smoking), secondly the prediction of the cardiovascular diseases and lastly integrating with rest the idea of vascular age. In this, total 959 persons were randomly selected in the sample, the age of those persons were ranging from 30-74 years. There were two types of examination cycles, first necessary data was collected for the new function and in the second cycle, recording of data on all events was done. ROC curves were used for the discrimination. The algorithm produced the results of risks, associated with female and male as 11.3% and 19.7% respectively and the areas under the ROC curve were 0.789 (95% CI, 0.716-0.863) and 0.780 (95% CI, 0.713-0.847) ($P < 0.001$, both). The main limitations of the algorithm were in terms of used data as it was only used for individuals, not for vascular age.

Norhayati et al. (2013) identified various factors that were associated with the 10 years risk of cardiovascular disease. The cardiovascular disease risk was according to WHO risk prediction chart. In around 196 patients, the ratio of high-risk group for 10-year cardiovascular disease risk ($\geq 20\%$) depending on the WHO risk prediction chart was 7.7%. The limitation with the above system was that the majority of social factors were

not included in the study and the size of sample estimation depending on the smallest detectable difference.

Bergoderi et al. (2012) set the validity and use of various other prediction models for calculating the risk associated with the cardiovascular disease. Three types of databases were reviewed: Ovid MEDLINE, LILACS and EMBASE. The identification of 1,655 patients were identified, from five cohorts with total 13,142 subjects meet the criteria. The predicted/ observed event-rate ratio for FRS was 1.68 (95% CI, 1.26–2.11); incident MI, 1.36 (95% CI, 0.90–1.83); and CHD death, 1.21 (95% CI, 0.43–2.00) as represented by the Mexican cohort. The main limitations during the review were the heterogeneity of the Latin American population, the lack of ability to do further subgroup analysis with the intermediate risk population and the capability to divide subgroup analysis between the intermediate risk populations.

Wannamethee et al. (2005) showed the comparison between the Metabolic Syndrome (MetS) with the FRS, Type 2 diabetes mellitus (DM2) in middle-aged men, stroke and CHD were used as predictors. The experiment was performed on 5,128 men age ranging from 40-59 years. The 26% men with metabolic syndrome with borderline parameters showed significantly high relative risk than the men without metabolic syndrome.

Soofi et al. (2015) described the prevalence of various risk factors related with the CHD and predicted the hard-coronary artery events in 10 years in the Saudi Arabia. For prediction, various factors were included such as Demographic factors and Hard Coronary Events (HCE) risk factors. The risk associated with each individual was measured by the means of FRS. According to the results, the examination was done on total 4932 people, out of which 2,215 were males and 2,717 females, the majority of 85% people were less than 40 years old.

The risk associated with the HCE within 10 years was found to be low in 92.6% of people, in 3.2% people, it was intermediate and 4.1% had high risk. The total prevalence of coronary artery disease was calculated as 5.5% as per the CADIS study. The population involved in CADIS was in the age range between 30-70 years. The model described in the

FRS, HCE prediction equation used the categorical approach that described the predictability of the model, which was similar to the continuous variable approach.

2.2.2 The Joint British Cardiac Society (BCS) / British Hypertension Society (BHS) / British Hyperlipidemia Association (BHA) Risk Score

The parameters used for this study are Gender, age, TC, HDL, SBP, DBP, DM, Smoking, ECG-LVH if available.

2.2.3 The European Heart SCORE

According to the national guidelines, the European Heart SCORE model is used for treatment and primary prevention of CVD by several European countries. For this score, the study has been conducted for a span of 10 years and the variables used for this study were age, gender, smoking, SBP and Total cholesterol levels (Stovring et al. 2013).

According to Stovring et al. (2013) the SCORE was not able to predict all causes of CVD mortality consistently. To overcome this, the effect of other factors were also considered like concomitant medication usage and the fact that non-CVD mortality might also play a role and vary according to factors like smoking status etc. After proposed modification, they had achieved the consistency while assuming that non-CVD mortality is unaffected by smoking. They found that the total risk dropped due to medication was ranged from 0.0% to 3.5%. The effect of this was that during scoring the expected lifetime increased from 3 to 11 months.

Conroya et al. (2003) developed a SCORE project for risk scoring system for clinical management of CVD risk. Total 12 data sets of different European cohort studies was used in building the project. The project used Weibull model for estimating the 10-year risk of serious CVD in which exposure time to risk was calculated by age as a measure rather than as a risk factor. There were 7,934 cardiovascular deaths, of which 5,652 deaths were from CHD. For CHD and non-CHD two separate estimation equations were developed. For estimation of risk, two different prediction models were developed depending on the TC and TC/HDL cholesterol ratio. The risk calculations were represented graphically by risk charts. The main limitation of the above project was that

the underlying functions were based on single risk factor measurements and not on the usual levels of the persons.

2.2.4. FINRISK Model

This model has been developed for the patients of European origin. It included Diabetes and LDL (Bhopal et al. 2005). The FINRISK model was developed based on the 10 Years study, performed during 1982 to 1992 in eastern and southwestern Finland. The paramets used for FINRISK model were smoking, SBP, TC, HDL, diabetes, and family history.

Bhopal et al. (2005) identified during his study that the FINRISK model was not so much effective for non-European population.

2.2.5. Centers for disease control and prevention (CDC)

CDC model based on Markov model was developed for heart disease prediction. At the time of development of this model as compared to other existing models, the CDC model was working on the fact that whether an individual gets CHD or continues to stay "normal". This model doesn't stop at first event, assuming that one can stay in the health states "angina" and "history of CA/MI" for several years.

2.2.6. Weinstein's model

Weinstein's model is possibly the most extensive of those mentioned here. This model has five health states; persons without CHD, persons with new CHD, persons with CHD, persons who survived up to the age of 85 years, and death. Hence this model has two different absorbing states; survive 85 and death. Individuals in the state "new CHD" will only be there for the first year, and after that will be transferred to "CHD", "survive until 85" or die states.

Weinstein et al. (1987) prepared the guidelines to modelers to accomplish and studies. The quality of model used in the report was based on the criteria that lie between model structure, data used as inputs to models, and model validation. The report suggested that the results of the various models was explained as the aid of decision-making

conditions rather than based on the scientific fact statement. The model assumptions for the estimations for casual structure and parameter were continuously assessed against data. The expected value of the additional information and cost of information was based on the balance between these two for obtaining the additional data to inform a model. For performing an empirical study or trial, the resource cost was used in the cost of the information.

2.2.7. CANFIS Model

CANFIS (Coactive Neuro Fuzzy Inference System) model is based on Neuro-Fuzzy algorithm. This Neuro-Fuzzy model has been used due to its ability to adapt and classify. In CANFIS model the genetic algorithm has also been integrated with Neuro-Fuzzy approach for more accurate prediction of heart disease (Parthiban et al. 2008). It combines the neural network's adaptive capabilities with fuzzy logic qualitative ability with genetic algorithm for predicting the possible CVD disease.

Parthiban et al. (2008) proposed a system for automatic diagnosis system. The system was based on the CANFIS which was used for the prediction of heart disease. The neural network adaptive capabilities and the fuzzy logic qualitative approach were combined into the CANFIS model, which was explained earlier. The training performances and classification accuracies were used as a performance measure for the CANFIS model. As per the results, the CANFIS model had a great capability in predicting the heart disease. For auto tuning of CANFIS parameters, Genetic Algorithm (GA) was the most useful technique and for the identification of the nonlinear relationships and mapping among different parameters.

2.2.8. Other Approaches

Since last few decades many researchers have been trying to solve different research problems using the Machine Learning (ML) and statistical approaches like, Neural Network, Probabilistic Neural Network, Genetic algorithm, Fuzzy logic based algorithms, for architectural and structural damage analysis, medical imaging equipments, health care systems, environmental research, Natural Language Processing and for other biotechnology problem areas etc.. (Mayani et al., 2007; Dutta et al., 2010; Panigrahi et al.,

2009; Sahoo et al., 2017; Chang et al., 2013; Mansourvar et al. 2017; Kadiyala et al., 2017; Akkala et al. 2013; Lauraitis et al., 2018; Gauba et al. 2017.).

Similarly, different ML approaches for CVD prediction have been used by many researchers, some of them are, as follows.

2.2.8.1. Neural Network (NN) and Genetic Algorithm (GA) based approach

Gupta et al. (2014) had proposed the Neural Network and genetic Algorithm based approach for predicting the chances of CVD. They had used the Feed-forward neural network algorithm for classification to diagnose the diseases. The dataset has been taken from University of Cleveland for training and testing purpose. The training of the system had been done using the Genetic algorithm. They had achieved the 97.7% accuracy in predicting the CVD. The main problem with their system was only the limited data set, i.e they had used only Cleveland data set for training and testing their system.

Gupta et al. (2009) proposed the approach for the prediction of risk of CVD. The desired accuracy of the system was achieved by integrating the neural network with genetic algorithms. The aim of the system was to determine the weights of the NN with the GA and reduce the number of iterations. During training and testing, dataset from the University of Cleveland was used. The accuracy of the system was 97.75%. Accuracy achieved by the Feed-forward and Fitting neural network was improved by 1.29% and 1.37% respectively.

Wagh et al. (2016) proposed the approach for prediction of CVD, for which various risk factors were used including age, ECG, heart rate diabetes, slope, hypertension, high cholesterol, or physical inactivity, etc. The approach used a Neuro-Fuzzy model (NFS) which was a hybrid model that combined the NN with the GA. NFS integrated the NN adaptive capabilities and the fuzzy logic reasoning approach for predicting the heart disease. For training, an improved genetic algorithm was used that automatically distributed the optimal set of weights. The accuracy obtained by the system during the evaluation was 90% as compared to 82% of the NFS without GA.

Patil et al. (2009) proposed a system for the prediction of the heart diseases. The data was first extracted from the warehouse and then after preprocessing of data, the data was clustered using the K-means algorithm. The MAFIA algorithm was used for mining and for calculating the significant weightage of the extracted data for heart disease prediction. The patterns that were significant to the heart disease prediction, were based on the estimated significant weightage. They also proposed the future work, as design and implementation of system for efficient heart disease attack prediction system with the help of the selected significant patterns using artificial intelligence techniques.

Lewsey et al. (2014) proposed a CVD policy model for the modelling of the remaining life expectancy that included the socioeconomic deprivation as an independent risk factor for CVD as a measurement. The Scottish Heart Health Extended Cohort (SHHEC), which was combined with the Scottish morbidity and death records, was used to develop the state transition model. The main limitation of the system was to ensure the calibration factor for ensuring the difference among the model predictions and life table and second limitation was the modifiable risk factors that were calculated at baseline with no measurable events.

Karthikeyan et al. (2017) proposed a heart disease prediction system that used the Deep Belief Network classification algorithm for predicting the risk associated with the heart disease. The deep learning approach in deep learning network was the most suitable algorithm for the Deep Belief Network. The approach also used the comparison of Convolutional Neural Network [CNN] and Deep Belief Network classification [DBN] algorithms. The Convolutional Neural Network algorithm was one of the unsupervised algorithms and achieved 82% of accuracy for predicting the risks of heart diseases. The Deep Belief Network algorithm achieved 90% accuracy, which improved the accuracy of the heart disease prediction system.

2.2.8.2. Data Mining

Krishnaiah et al. (2014) proposed fuzziness in the measured data by Data Mining (DM) techniques to remove uncertainty in the data. In this, a membership function was designed and integrated with the measured values. An effort was tried to categorize the

patients, which depends on various parameters collected from the medical field. To classify the data into different groups, a minimum distance K-NN classifier was integrated with data. Fuzzy K-NN classifier was compared with other classifiers of parametric techniques.

Palaniappan et al. (2008) proposed a model that was based on the DM techniques using the Decision Trees, Naïve Bayes and NN. As per the results, this technique has noble strength in realizing the complex queries. For predicting the patients getting a heart disease, the model was using the medical profiles such as the age, sex, blood pressure and blood sugar. The model was developed on the .Net platform. For building and accessing the model, the system was using the DMX query language and functions and the system acquired hidden knowledge from a historical heart disease database. The model was trained and validating using the test dataset. The evaluation of the effectiveness of the models was done by the chart and classification matrix methods. These models are able to extract all predictable states.

2.2.8.3. Markov Tree

Martin et al. (2004) proposed a model that was used in those people who were free from heart diseases and evaluated their accuracy against data. The Markov cycle tree was developed using cohort simulation. The comparison in the model was done with the estimation of survival of 20 years for 1,391 members, the age of those members lie between 35 and 65 years respectively. During the evaluation, the estimated survival was 75% and the observed was 75.4%. With the 39 subgroups, the correlation between estimated and observed survival was 0.933.

2.2.8.4. Fuzzy Logic

Nguyen Cong Long et al. (2015) proposed a heart disease diagnosis system that was based on the attribute reduction and interval type-2 fuzzy logic system (IT2FLS). The integration between these two handles high-dimensional dataset and uncertainties. The IT2FLS was based on the hybrid learning process that has fuzzy c-mean clustering algorithm and parameters tuning, the chaos firefly and genetic learning algorithm was

used for tuning. The experimental results described by the significant dominance of the proposed system was compared with the other machine learning methods namely Naive Bayes, support vector machines, and artificial neural network.

Anooj (2012) proposed a machine learning technique that acquires the knowledge automatically from the raw data. The diagnosis of disease was done automatically from patient raw data with the help of a weighted fuzzy rule clinical decision. The system was consisting of two phases that included the generation of weighted fuzzy rule by automated approach and fuzzy rule-based decision support system was developed. The weighted fuzzy rules were obtained by using the data mining technique, attribute selection and attribute method. The datasets were extracted from the UCI repository for the experimentation and comparison of neural network. The K-fold cross validation was used for analyzing the predictive outcomes and their sensitivity, specificity and accuracy values. The neural network-based system achieved 52.4% as compared to the sensitivity of the proposed system which was 45.2%. The specificity of the system was 68.7% whereas the overall accuracy of the system was 68.7% as compared to other systems.

Khatibi et al. (2010) proposed a fuzzy-evidential hybrid inference engine, based on the Dempster-Shafer theory of evidence and fuzzy sets theory. Its working was based on the two phases, during first phase fuzzy sets were used to model the input information's vagueness and after that it extracts or choose the fuzzy set rules and implement those rules on the problems for producing the outputs of first phase. During second phase, the system was using the results from the first stage and then basic belief for the problems and plausibility functions were set. In the next, it acquired the data from different sources and provided integrative results. The system achieved the accuracy of 91.5%.

2.2.8.5. Multilayer Classifier

Bashir et al. (2016) proposed a multi-layer classifier-based approach. The proposed framework solved two different CVD datasets, which included two breast cancer datasets, two diabetes datasets, two liver disease datasets, one Parkinson's disease dataset and one hepatitis dataset. These datasets were extracted from the public repositories. According to the results, the proposed framework used all types of parameters for

achieving high accuracy. The proposed model was predicting only class 0 or class 1, which was used to depict absence or presence of the disease, which was the main limitation of the proposed framework. However, the limitations present in the traditional models were solved by combining seven heterogeneous classifiers, which increase the complexity and take lot of computation time.

2.2.8.6. Clustering Method

Esteghamati et al. (2014) described the association between homocysteine (Hcy) and metabolic syndrome (MetS). This model uses the clustering method for Hcy with components that were access using the exploratory factor analysis.

2.2.8.7. Age and Gender

Jousilahti et al. (1999) studied the cardiovascular risk factors according to the difference in sex and ages of subjects. The study was done on 14786 men and women whose age ranges between 25 to 64 years. The study was performed on the subjects to determine the various risk factors such as smoking, TC, HDL, BP, BMI, and diabetes. During the study, it was found that CHD appearance in men was 3 times greater than the women whereas mortality was 5 times higher. One-third of the age-related increase in CVD risk between men and women was found to be 50% to 60% in terms of serum TC level, BP, BMI, and diabetes.

2.2.8.8. Family History

Imes et al. (2014) studied the relationship between Family History (FH) of a CVD. They examined the associations between a FH of CVD, perceived CVD risk, and health-promoting behaviors took help of PubMed, CINAHL Plus and PsychINFO. According to the results the relationships between the reported FH of CVD and perceived risk was positive. During this study, 263 articles were studied. The study also included the test interventions that were used to help subjects with FH attribute increased personal risk to them for developing CVD.

2.2.8.9. Diabetes Mellitus

Almdal et al. (2012) studied the effect of type 2 Diabetes Mellitus (DM) on cardiovascular morbidity and mortality. The research was performed on 13105 subjects, which was followed prospectively for 20 years. The proposed research was used for the evaluation of adjusted relative risks of first incident, admission for, or death from ischemic heart disease, acute myocardial infarction, or stroke, as well as total mortality in persons with type 2 DM. According to the results, with the presence of DM as comorbidity, the risk associated with the women for first incident, admission and myocardial infarction was increased from 1.5 to 4.5 and for men 1.5 to 2 respectively. Similarly, the risk associated with women for first incident and admission for stroke was increased from 2 to 6.5 and for men 1.5 to 2 respectively.

2.2.8.10. Smoking

Gupta et al. (2009) conducted the epidemiological study for the identification of risk of cardiovascular disease and the impact of risk factors like tobacco or smoking on it. The proposed study was performed on total 2,051 subjects in which 1,009 were males and 1,042 were females. The age of all the subjects were between 15-39 years. For the correlation of various age and risk factors, logistic regression was used in the study. According to the results, out of whole population, for men and women respectively, the usage of tobacco or smoking was observed in 200 (11.8%) and 18 (1.4%), obesity in 12.4% and 14.3%, high waist-hip ratio in 15% and 32.3%, hypertension in 5.6% and 3.1%, high LDL in 9.4% and 8.9%, low HDL in 16.2% and 49.7%, hypertriglyceridemia in 9.7% and 6%, diabetes in 1.0% and 0.4% and the metabolic syndrome in 3.4% and 3.6%.

Merry et al. (2011) studied the relationship between the lifestyle factors and unstable angina pectoris (UAP). During this study, total 21,148 people were studied whose age range between 20-59 years. During this study, after 16.9 years of follow-ups total AMI and UAP cases were 420 and 274 respectively. According to the results, the risk of disease was increased with smoking and alcohol showed the protective effect to disease, the risks related to AMI were stronger than UAP.

2.2.8.11. BMI

Sangster et al. (2010) proposed a system to determine the efficacy of promoting healthy weight and physical activity using telephone-delivered lifestyle intervention. The system was based on the randomised controlled trial, in these two types of people from urban and rural areas of Australia was used. The training of subjects was based on the, nutrition and physical activity follow-up booster telephone calls, which were supported by the two follow-up booster telephone calls, a pedometer and written materials. The system noted the baseline data for group of subjects for eight weeks and eight months. This system provides two types of outcomes (primary and secondary) that included weight change and physical activity, sedentary time and nutrition habits. The system improves the secondary outcomes.

Wilsgaard et al. (2007) had studied the purpose of the FRS equation for representing the relationships between longitudinal BMI and CHD risk score variation. For this total 10,214 men and women are tested whose ages range from 20 to 61 years at baseline for at least three times. The fixed-effects regression methods for longitudinal data were used for the examination of associations. As the results described the subjects aged between 40 to 49 years with BMI increase of 3kg/m^2 was related to the increase in risk score of 0.45 points (95% confidence interval [CI], 0.29e0.62) in men and 0.66 points (95% CI, 0.52e0.80) in women.

2.2.8.12. Biomarkers

Chang Kim et al. (2010) investigated the use of multiple biomarkers for improving the risks associated with the CHD in prediction in post-menopausal women. The investigation was done on 27,347 post-menopausal women whose ages range between 50 to 79 years. According to the results, the TRF model used in the investigation shows the improvement in C-statistic (0.729 vs. 0.699, $p \leq 0.001$) and net reclassification 6.42% compared with the FRS. The improvement in the accuracy of the additional biomarker model was 0.751 vs. 0.729, $p \leq 0.001$ in C-statistic and for net reclassification improvement was 6.45% noted as compared to TRF model. The system had some limitations in nested control design as compared to a cohort study that was generally

used for assessment of risk prediction. Secondly, the calibration of risk prediction model was not possible in a case control study.

2.2.8.13. HDL

Skretteberg et al. (2012) presented the hypothesis of prognostic effect of HDL for CHD and for CVD and all cause death. During this research, total 1,357 healthy men whose ages between 44– 69 years observed up to 28 years. The bicycle exercise test was used for measuring the physical fitness (PF). Cox proportional survival model was used for the calculation of Hazard ratios for the related parameters. According to the results, the highest HDL quartile was related with the lower risk of CHD. The main limitation of the system was that the outcome found during the study was not generalized to individuals of other ethnicity, age or individuals who have some kind of disease.

Yu et al. (2003) studied the protective effect of HDL on incident CHD. With the incidence of CHD, the relations between the HDL and HDL cholesterol in the 9- years follow up was studied. Total HDL and cholesterol level were measured, during this study, total 2398 middle-aged British male were tested.

2.2.8.14. LDL

Ren et al. (2010) studied the relationship between LDL and the risk associated with the future of CHD events. For this, total 30,378 participants were involved, whose ages were from 35 to 64 years from 11 Chinese provinces. They found 20% of the population had high LDL (equal or more than 30 mg/dL). High LDL increased the CHD risk in people with normal LDL and presenting no other major risk factors.

CHAPTER 3

MATERIALS AND METHOD

Materials and Method

3.1. Introduction

The current study has been carried out since 2009 to 2017 on a cohort of 815 CVD patients aged 20 years and above. During the process of data collection, the personal information of patients were kept confidential. The Quantum Neural Network (QNN) has been used to build knowledge-based CVD prediction for better adoptability and do better classification. In this work, the weightage and priority level of parameters significant to cardiovascular Disease are decided in accordance with the advice of a panel of 3 medical experts.

3.2. Risk Factors

As shown in Table 3.1, the various input parameters for the prediction of CVD.

Table 3.1: Parameters used for the predicting the CVD

S.No.	Type of Risk Factors	Parameters
1.	Physical Risk Factors	Body Mass Index
2.		Age
3.		Gender
4.		Tobacco Smoke
5.		Alcoholism
6.		Family History
7.		Stress
8.	Physiological Risk Factors	Systolic Blood Pressure (SBP)
9.		Diastolic Blood Pressure (DBP)
10.		Total Cholesterol (TC)
11.		C- reactive Protein (hs- CRP)
12.		Brain Natriuretic Peptide
13.		Homocysteine
14.		Sodium

Materials and Method

15.		Potassium
16.		Bicarbonate
17.		Chloride
18.		Calcium
19.		Fibrinogen
20.		Platelet Count
21.		Aspartate transaminase (SGOT)
22.		Creatine Kinase
23.		Creatine Phosphokinase-MB
24.		Troponin
25.	Clinical Risk Factors	Diabetes Mellitus (Random Blood sugar)
26.		Low Density Lipoprotein levels (LDL)
27.		High Density Lipoprotein levels (HDL)
28.		Triglycerides levels

3.2.1. Gender

According to the literature, males are at increased risk of Cardiovascular diseases than pre-menopausal females, whereas post-menopause, risk of CVD equalizes/ matches in both the groups. Whereas, the risk of stroke is almost same in both males and females in all age groups (Lloyd-Jones et.al. 1999, Legato et.al. 1997, Pilote et.al. 1995).

3.2.2. Age

The risk of CVD is proportionate with age. According to various studies, after the age of 65, 80% deaths are attributable to heart diseases. As the age increases, blood vessels become less flexible, non-compliant and fat laden (plaque), hence initiating the cascade of thrombogenic state and increasing the turbulence in blood flow (Lloyd-Jones et.al. 1999).

3.2.3. Total Cholesterol

High level of total blood cholesterol is one of the causes contributing to cardiovascular diseases (Ezzati et.al. 2003). Blood cholesterol levels are directly proportional to risk of

Cardiac events. Table 3.2 shows the various risk levels corresponding to total serum cholesterol levels.

Table 3.2: Risk Levels for Total Cholesterol levels

Total Cholesterol (mg/dl) (mmol/l)	Risk of CVD
Less than 200 mg/dL (5.17 mmol/l)	Desirable
200-239 mg/dL (5.17 – 6.18 mmol/l)	Borderline high risk
240 and over (6.2 mmol/l)	High risk

3.2.4. High Density Lipoprotein (HDL)

Abnormal (lower) levels of High-density lipoprotein may increase the risk of cardiovascular events by speeding up of atherosclerosis, cut off values/ risk levels are shown in Table 3.3. According to various studies, the HDL level is strongly and inversely related to cardiac events, (Assmann et.al. 1996, Turner et.al. 1998, Gordon et.al. 1977).

Table 3.3: Risk Levels for High Density Lipoprotein (HDL)

High Density Lipoprotein (HDL): (mg/dl) (mmol/l)	Risk of CVD
Less than 40 mg/dL (men) (1.03 mmol/l)	increased risk of heart disease
Less than 50 mg/dL (women) (1.29 mmol/l)	increased risk of heart disease
Greater than 60 mg/dL (1.55mmol/l)	some protection against heart disease

3.2.5. Blood pressure

It is one of the most important risk factors contributing towards CVD. All over the world, various epidemiological studies have pointed towards the direct relation between blood pressure elevation and risk of CVD's and stroke (McMahon et.al. 1990, Kannel 1996, Walker et.al. 1992). It has also been noticed that combination of multiple risk factors,

Materials and Method

high blood pressure being the major one, increases the risk of CVD many folds (Anderson et.al. 1991, PROCAM, 1988). The normal age wise Blood pressure range is shown in Table 3.4.

Table 3.4: Age wise Blood Pressure Range (mmHg)

Age Group	Lower Limit*	Average Limit*	Upper Limit*
15-19	105/73	117/77	120/81
20-24	108/75	120/79	132/83
25-29	109/76	121/80	133/84
30-34	110/77	122/81	134/85
35-39	111/78	123/82	135/86
40-44	112/79	125/83	137/87
45-49	115/80	127/84	139/88
50-54	116/81	129/85	142/89
55-59	118/82	131/86	144/90
60-64	121/83	134/87	147/91
65-69	122/84	136/88	149/92
70 and above	123/85	138/89	151/93

(*systolic/diastolic blood pressure)

3.2.6. Diabetes Mellitus

Diabetes is another important factor, which increases the risk of heart diseases (Pyorala et. al. 1987, Krolewski et. al. 1987, Yudkin et. al. 1996, Laakso et. al. 1997, Haffner et. al. 1997, Nathan et. al. 1997, Haffner, 1998). According to Rancho Bernardo Study, which was carried out for 14 years on around 2,400 persons, presence of diabetes increases the

risk of cardiac events up to 1.9 times higher in males and 3.3 times higher in females when compared to non-diabetic individuals (Barrett-Connor et. al. 1991). Moreover, it has been noticed that the relation between diabetes and risk of cardiovascular diseases is not uniform in all races (American Diabetes Association 1993, Head et. al. 1990).

3.2.7. Body Mass Index (BMI)

The BMI is a statistical measurement tool. It is ratio between weight and height (in meters) (Dudina et.al. 2011), as shown in Table 3.5.

Table 3.5: BMI Range

Category	BMI range – kg/m ²
Severely underweight	less than 16.5
Underweight	from 16.5 to 18.5
Normal	from 18.5 to 25
Overweight	from 25 to 30
Obese	Above 30

3.2.8. Smoking

It is one of the most important modifiable factors, which increases the risk of Cardiovascular diseases (Kannel et al., 1990). Globally different studies have confirmed the fact that smoking (Tobacco use) increases the risk of cardiac events and stroke too by many folds (Tverdal et.al. 1993, Doll et.al. 1994, Willett et.al. 1997). Moreover, Smoking contributes to 50% of total deaths, and its one-half occur only due to CVD (Rosengren et.al. 1992, Bartechi et.al. 1994, MacKenzie et.al. 1994 and Manson et.al. 1992).

3.2.9. Alcoholism

Excessive alcohol intake may affect various organ systems, most importantly being cardiovascular system. It increases the blood pressure by different mechanisms, leading to risk of stroke (both infarct and intracranial bleed) and leads to heart failure. It raises serum triglyceride levels and hence a contributing factor for developing obesity, diabetes and CHD. Alcoholism is also associated with various arrhythmias (irregular heartbeats),

cardiomyopathies (diseases of heart muscles) and sudden cardiac death (Luc Djoussé et.al. 2009, Kenneth et.al. 2006 and Roerecke et.al. 2014).

3.2.10. Family History or heredity

Heredity plays a vital role in pathogenesis of heart diseases. Either directly or indirectly by effecting other risk factors too, it increases incidence of CVD many folds. If the heart disease is in a family history of a person, then he/she have higher potential of Heart diseases especially when combined with other modifiable risk factors e.g. eating habits, lifestyle, obesity etc (Valerio et.al. 2016, Bachmann et.al. 2012).

3.2.11. Stress

Stress is an important factor in heart disease. Different studies showed the Correlation between stress and CVD risk in a person's life, their health behaviors and socioeconomic status, which inturn may affect other risk factors too (Dimsdale 2008).

3.2.12. Low Density Lipoprotein (LDL)

LDL cholesterol is commonly known as "bad" cholesterol due to its major contribution in formation of a plaque, which deposits and occludes vessels (atherosclerosis), hence initiating a cascade of CVD's which may end as Heart failure (Brown et.al. 1976, Steinberg et.al. 1989). The spectrum of CVD's may include Stable/ unstable angina (chest pain), myocardial infarction (heart attack), arrhythmias (irregular heartbeats), sudden cardiac death, stroke etc (Brown et.al. 1976, Steinberg et.al. 1989).

3.2.13. C- reactive Protein (hs- CRP)

"High- sensitivity" CRP or hs-CRP is involved in immunological processes that leads to plaque formation and vascular remodelling, which are the main components in etiopathogenesis of atherosclerosis and CVD's (Pfützner et.al. 2006). This test has a high negative predictive value, hence low values of hs- CRP need not be repeated and rules out the risk of CVD whereas higher levels shows high risk of CVD and repeated tests are required to look for the consistency, trends and also de-escalation of various cardiac medications (statins) according to effect of medicines on values of CRP during course of

treatment (Fonseca et.al. 2016). The various levels for C- reactive protein is shown in Table 3.6

Table 3.6: Risk Levels for C- reactive Protein

Range	C- reactive Protein (mg/L)
Lower Limit	<1
Average Limit	1-3
Upper Limit	>3

3.2.14. Brain Natriuretic Peptide (BNP)

These peptide hormones are produced in heart, brain and other organs. The Cardiac Natriuretic Hormones (CNHs) or Natriuretic Peptides are secreted by heart muscle in small/basal amount (both atria and ventricles) for its proper functioning whereas excessive trial and ventricular distension during heart failure stimulates and stretches myocardium to release BNP as a compensatory mechanism much higher above the baseline (Vuolteenaho et al. 2005). Therefore, higher serum levels as shown in Table 3.7 indicate towards presence of impending or frank heart failure (Mahadavan 2014). This test also has a high negative predictive value for CVD's. The increase in CNHs levels apart from heart failure (HF) is also seen in renal failure and liver cirrhosis (Volpe et al. 2014, Federico 2010).

Table 3.7: Risk Levels for Brain Natriuretic Peptide

Range	Brain Natriuretic Peptide (pgm/ ml)
Lower Limit	100
Average Limit	100 - 300
Upper Limit	>300

3.2.15. Homocysteine

It contributes as a major risk factor for CVD's and stroke. With the increase in level of serum Homocysteine (either acquired or congenital), the formation of the structural components of the artery is inhibited as well as degraded, hence leading to endothelial injury and thereafter thrombus formation. It also induces oxidative stress and lead to increase in number of free radicals which further amplify the spiral of CVD's (Ganguly et al. 2015, Kuller et al. 1998). The risk levels for homocysteine is shown in the Figure 3.1

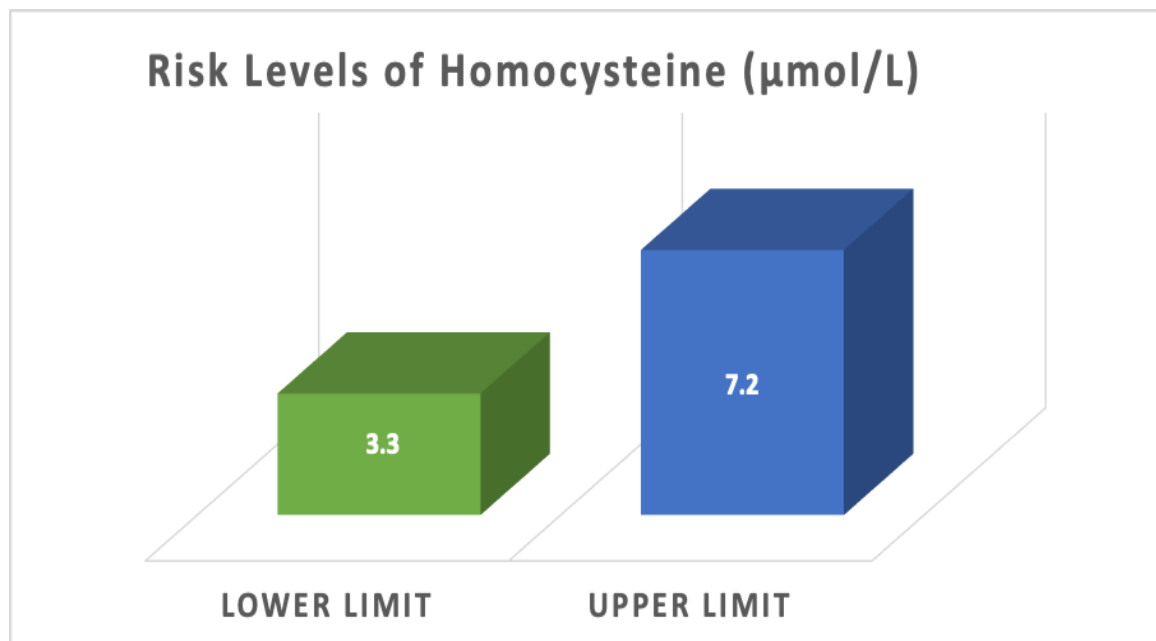


Figure 3.1 Risk Levels for Homocysteine

3.2.16. Electrolytes

Electrolytes are important for the maintenance of resting membrane potentials and also conduction of electric impulses across the cells and cell membranes of various tissues especially heart, nerve and muscle (Watson 1987, Moser et.al. 2003). The risk levels of various electrolytes are shown in Table 3.8

Table 3.8: Risk Levels for Electrolytes

Range	Sodium (mmol/L)	Potassium (mmol/L)	Bicarbonate (mmol/L)	Chloride (mmol/L)	Calcium (mmol/L)
Lower Limit	135	3	18	98	2.1
Upper Limit	147	5	24	107	2.8

3.2.17. Fibrinogen

Fibrinogen is a protein produced in liver, which takes part in coagulation cascade (process which results in formation of blood clot and stoppage of bleeding). Smoking, inactivity, excessive alcohol, hormonal supplements etc. may increase the fibrinogen levels abnormally, hence resulting in a hypercoagulable state (inadvertent blood clotting in absence of any stimulus) which all results in a high probability of blood clots (thrombus), plaque formation, vascular remodelling, embolisation etc which may lead to CVD, stroke, PVD (peripheral vascular disease) etc. To summarize, the risk of heart disease increases with increase in serum fibrinogen level (Stec et al. 2000, Meade 1997 and Ernst et al. 1997). Various levels of fibrinogen is shown in Figure 3.2

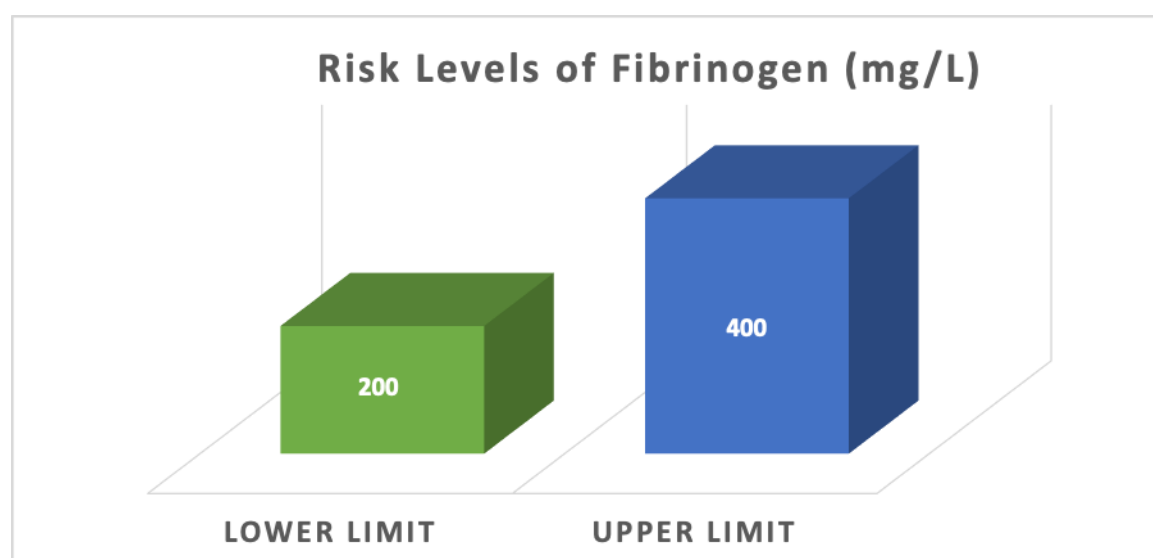


Figure 3.2 Risk Levels for Fibrinogen

3.2.18. Platelet Count

Platelets play a key role in formation of plaque and thrombus especially during thrombotic states (Smoking, inactivity, excessive alcohol, hormonal supplements, pregnancy, dehydration etc) leading to Atherosclerosis and Acute myocardial infarction (AMI) contribute to reperfusion injury (Vinholt et al. 2016, Thaulow et al. 1991, Gregg et al. 2003, Chung et al. 2006). Risk levels of platelet count is shown in Figure 3.3

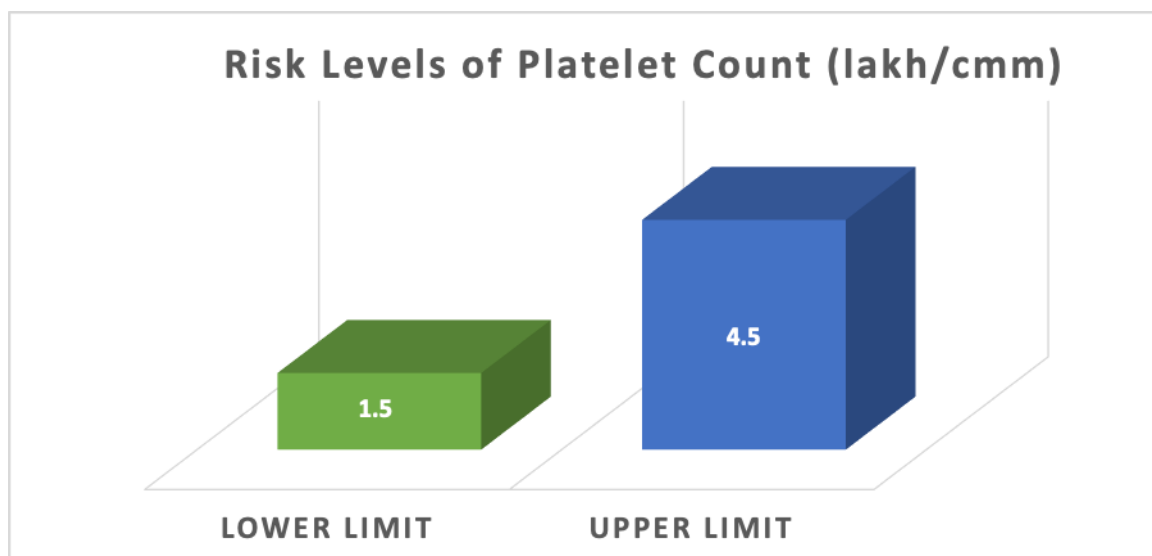


Figure 3.3 Risk Levels for Platelet count

3.2.19. Serum Glutamic Oxaloacetic Transaminase (SGOT)

Serum Glutamic Oxaloacetic Transaminase (SGOT) also called as aspartate aminotransferase (AST). The presence of SGOT/ AST persists in many tissues but highly in heart, liver and skeletal muscle cells. The aspartate aminotransferase to alanine aminotransferase (AST/ALT) ratio, shows the severeness of liver disease. Specifically, in patients of CLD (chronic liver disease) high AST/ALT ratio is related to the increased risk of cardiovascular disease (CVD). (Kunutsor et.al. 2014, Shen et.al. 2015, Killip et al. 1960). The risk levels for Aspartate aminotransferase (SGOT) is shown in Figure 3.4

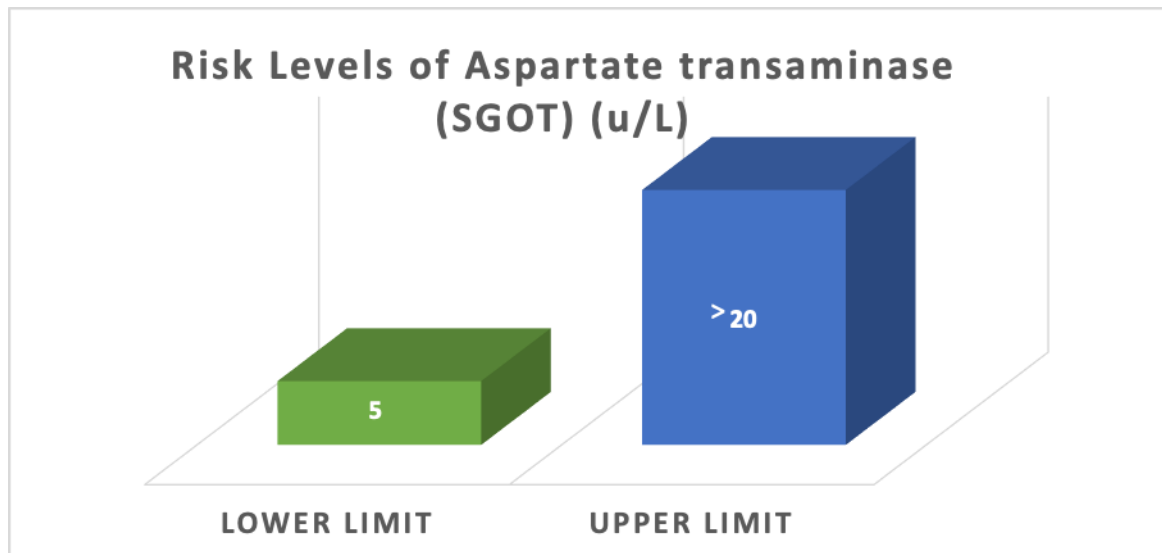


Figure 3.4 Risk Levels for Aspartate transaminase

3.2.20. Creatine Kinase (CK)

The enzyme CK is also known as creatine phosphokinase (CPK) or phosphocreatine kinase and risk levels are shown in Figure 3.5. It belongs to that subset of cardiac enzymes which are known to rise early after a myocardial insult, hence aiding in rapid and timely diagnosis. Early MI can be diagnosed with use of a ratio of CK-MB (Creatine Kinase- Muscle/Brain) to total CK. Elevated ratio or absolute value of CK-MB indicates the level of myocardial injury (Watanabe et al. 2009, Ay et al. 2002).

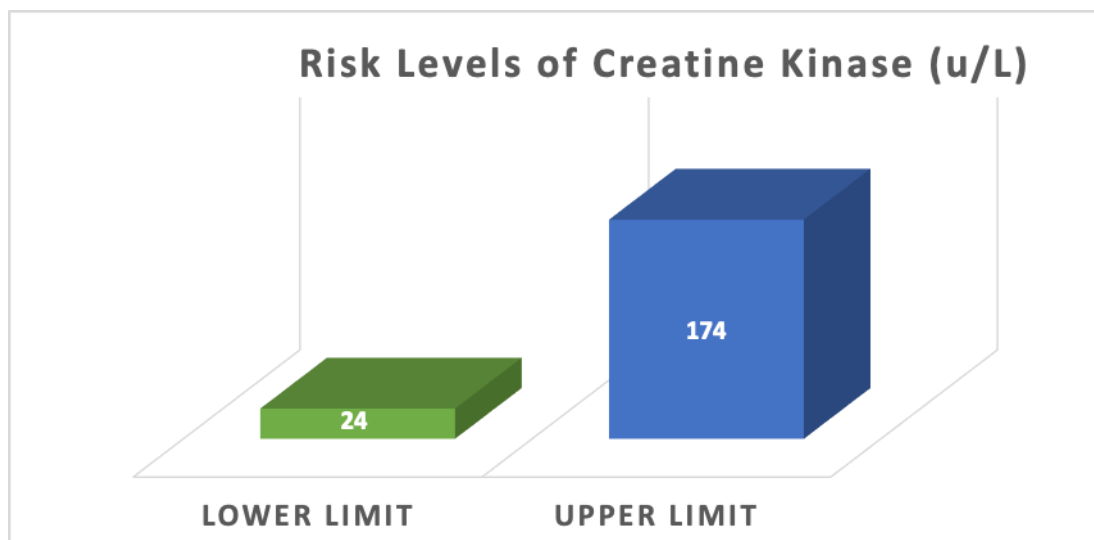


Figure 3.5 Risk Levels for Creatine Kinase

3.2.21. Troponin

Troponin along with Tropomyosin is required for calcium-mediated regulation of cardiac and skeletal muscle contraction. Three isoforms I, T and C are present with tissue specific functions. Two isoforms, namely, troponin I (cTnI) and troponin T (cTnT) are used in assays for cardiac injury diagnosis (Babu et al. 2005). After CVD, the levels of isoforms rise early and stay in serum for a long duration, hence assisting in diagnosis of both acute and subacute myocardial infarction. The various levels for troponin is shown in Table 3.9.

Table 3.9: Risk Levels for Troponin

Range	Troponin (ng/ml)
Lower Limit	0.0
Upper Limit	0.4

3.2.22. Triglycerides

High serum triglyceride levels raise the risk of CVD's and various other risk factors viz. diabetes, obesity etc. The American Heart Association (AHA) recommends that a triglyceride level of 100 mg/dL (1.3 mmol/L) or lower is considered "optimal." The AHA says this optimal level would improve overall health status, including heart health. However, the AHA does not recommend a single drug treatment to target a specific level of triglyceride. Instead, lifestyle modifications such as dietary changes, weight loss and physical activity are rather encouraged (Miller et al. 2011).

3.3. Data Collection

During this study the data of 815 patients has been collected, those have symptoms of heart disease. Furthermore, dataset of 5209 CVD patients from the famous Framingham has been taken from the University of Washington, Seattle, WA, USA ("<http://courses.washington.edu/b513/datasets/datasets.php?class=513>").

The whole dataset is divided into training, validation and testing as shown in Table 4.2

3.3.1. Pre-processing of Data

The steps below given were used for preprocessing the data:

- **Data Cleaning:** In this step, the primary data of 815 patients and secondary data of 5,209 have been checked and cleaned for missing data, noisy data. The outliers have been removed along with the duplicate data.
- **Data Integration:** In this, step, the Primary and secondary data have been combined together to form the consistent data.
- **Data Transformation:** In this step, the whole (primary and secondary) raw data has been converted into structured data for feeding into Machine learning algorithm to create the predictive model. The main steps to achieve this are given below:
 - **Normalization** – In this step, all the data is transformed by scaling up the whole data into specified range between 0 and 1.
 - **Aggregation** – In this step, two or more features have been combined into one.

3.4. Quantum Theory and Quantum Neural Network

The main reason behind using the QNN is its ability to work on overlap data classes. That cannot be handled by traditional machine learning algorithms. As compared to human brain, the QNN also performs well in unseen circumstances. Where classes of information are not closely enclosed or overlapped on each other. Humans think, learns, and update themselves by experiences. Similarly, QNN is able to learn and update its knowledge from sample tagged data automatically due to its adaptive feature. QNN has an inherently fuzzy property due to which it easily handles the overlapped dataset, which comes in different classes. In this situation, QNN assigns the fragmentary overlapped data to all overlapped classes (Purushothaman and Karayiannis, 1994, 1997, 1998). As compared to QNN, the traditional Artificial Neural Network (ANN) cannot handle the overlapped data classes. It works only if the data classes are crisp i.e. tightly enclosed or

not overlapped on each other (Purushothaman and Karayiannis, 1994, 1997, 1998; Kretzschmar et al. 2000).

QNN is very efficient for classifying the uncertain data, due to its inherent fuzzy properties. It encodes the information into distinct classes of certainty or uncertainty. The activation function of QNN expressed as linear superposition of sigmoid function (Zhou et al. 1999) due to which the hidden layer node expresses itself in more than one states. Due to this phenomenon, the QNN overcomes on indeterminacy and increases reliability in pattern recognition (Behrman et al. 2000; Mitsunaga et al. 2006, Daqi et al. 2007).

3.4.1 Quantum theory

Quantum theory explains the fundamental theoretical principles regarding nature and behaviour of matter at sub-atomic level.

According to Quantum theory the basic principles are:

- **The superposition principle:** The superposition principle defines the possible quantum states of a particle in a given system.
- **The measurement principle:** The superposition principle gives the information about all the available state of any particle.
- **Unitary evolution:** - Unitary evolution principle predicts the change in state with time.

3.4.2 Quantum Superposition Principle

According to this principle, at the same time, one particle may persist in two or more quantum state. It explains the rate of apparent effects at subatomic level. According to Quantum superposition principle, the state of any object or particle is unknown, as particle may persist in two or more quantum state at the same time, simultaneously. The concept of Quantum theory is widely applicable and it has been utilized in different fields like quantum physics, quantum chemistry, quantum computing etc., to solve different areas like drug discovery, simulations, complex computation etc. (Kumar et al., 2015, 2017; Bala et al., 1994; Gorecki et al., 2006)

3.4.3 Quantum Neural Network

The main difference between traditional ANN and QNN is in its hidden layer, instead of single state in hidden layer, the QNN's hidden layer have superposition state. In QNN, for achieving the superposition state in hidden layer, in place of using ordinary activation function, a multilevel activation function is used. Here, the summation of sigmoid function shifted by the quantum intervals, was used to achieve the multilevel activation function. Due to micro level intervals of quantum, the QNN has the capability to manage and pass the uncertain data, which improves the accuracy of pattern recognition.

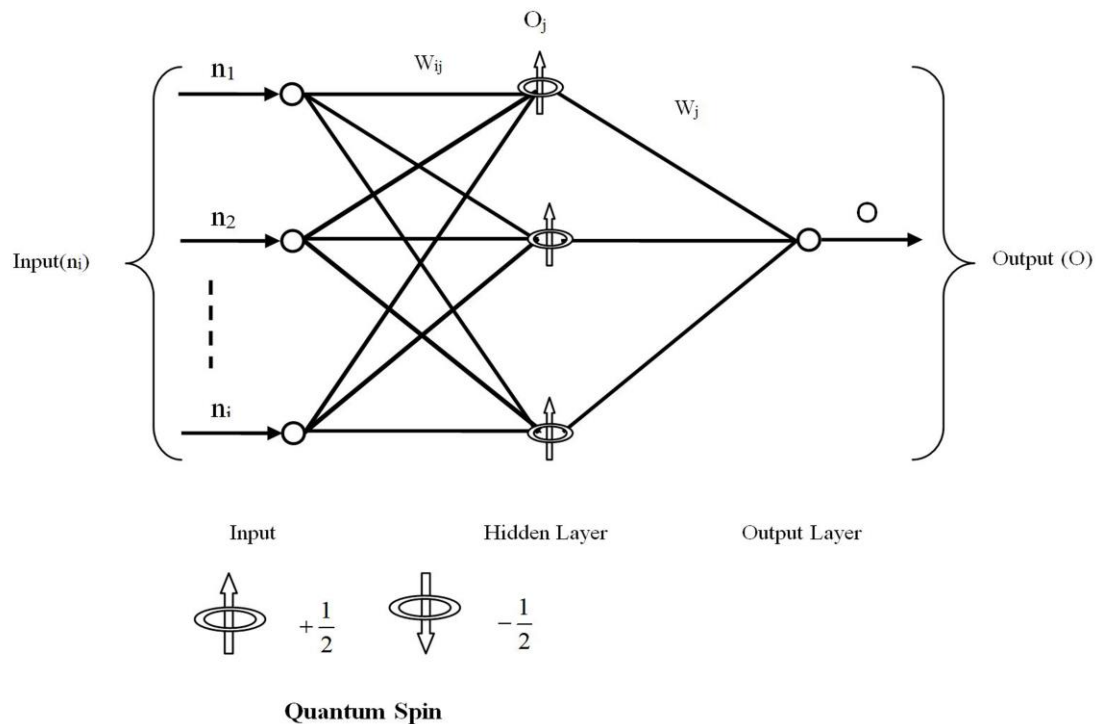


Figure 3.6 Architecture of Quantum Neural Network

Here the three layers QNN has been used, in which one input layer, one hidden layer and one output layer used, as shown in the Figure 3.6. The multilevel hidden layer shifted by Quantum interval has been proposed for the present research work. Here, each neuron of hidden layer contains two excitation states, which virtually represent two more nodes, during learning any one of the nodes may be used. Due to this phenomenon more states can be expressed in a hidden layer neural cell, on other hand as compared to traditional

sigmoid function, only two states can be expressed. Here, the node has been randomly selected by QNN algorithm. (Narayan et. al., 2013, 2014b, 2014c, 2016)

The activation function of hidden layer uses the quantum sigmoid function with various graded levels, is shown below as:

$$sgm(x) = \sum_{r=1}^{n_s} \left(1 / (1 + \exp(-x + \theta^r)) \right) \quad \dots (3.1)$$

In proposed system the total number of quantum levels have used are three *i.e.* One normal with two excited states with the difference of quantum interval θ^r . Here r is quantum level. It is based on the fundamental principal of quantum physics.

Here the delta method has been employed as a training method. The activation function of output layer uses simple sigmoid function, is shown below as:

$$sgm(x) = (1 / (1 + \exp(-x))) \quad \dots (3.2)$$

3.5 Advantages of Quantum Neural Network

The main advantages of Quantum Neural Network are:

- a) In comparison of classical Neural Network, the QNN takes less computational time *i.e.* it is faster than Classical Neural Network (CNN).
- b) In terms of accuracy, the QNN handles uncertain overlapped data more accurately, as it has built-in capability to detect uncertainty based on the knowledge gain during learning on the sample dataset.
- c) Thus, it is more accurate in pattern recognition in compairson to Classical Neural Network.

CHAPTER 4

**PROPOSED PREDICTIVE AND
ADOPTIVE ALGORITHM(S) FOR
ATHEROSCLEROSIS AND
MYOCARDIAL INFARCTION**

Proposed Predictive and Adoptive Algorithm(s) for Atherosclerosis and Myocardial Infarction

4.1 Introduction

In this research work, a novel approach for the prediction of cardiovascular diseases has been presented. The cardiovascular disease is caused due to various factors and to diagnose it on the basis of analysis of several risk factors is a very tedious task for a medical practitioner. So, there is a need of CVD prediction system which would predict accurately and in a timely manner. To make our system effective and reliable, Quantum neural network has been incorporated, which automatically learn and extract significant patterns which are involved in development of heart disease. For this we have created various datasets with the help of data collected from population of patients. The knowledge gained via patterns is applied in developing the automated system for improved diagnosis as well as prognosis of the cardiovascular disease. This research uses techniques drawn from machine learning as well as an understanding of the underlying biological systems. Hence a combination of both aspects is used in present study.

The present study has been conducted during 2009 to 2017. During this study, all the new and also already existing Physical, Physiological and Clinical parameters have been considered. This approach uses Quantum Neural Network (QNN) for better adoptability and classification of different risk factors. The proposed system is able to learn from the patient history and also identifies the current trends. It also updates itself on the basis of newly fed data. Therefore, this system would serve as a highly reliable, flexible and rationalized prediction model.

4.2 Methodology used to implement proposed Cardiovascular Disease Prediction System

To implement proposed Cardiovascular Disease Prediction System, following methodology has been adopted:

Proposed Predictive and Adoptive Algorithm(s) for Atherosclerosis and Myocardial Infarction

- Data Description - Patients Data: Primary and Secondary Data Collection to construct the final dataset, based on different parametric Category.
- Review the existing approaches and Algorithm(s) applied for achieving the highly accurate Cardiovascular Disease Prediction System.
- Creating up the Quantum Neural Network which is a Predictive and Adoptive Algorithm(s) for Cardiovascular Disease Prediction System.

4.2.1 Data Description - Patients Data

During this study the data of 815 patients was collected, who were already having symptoms of different categories of heart diseases. Furthermore, dataset of 5,209 cardiovascular disease patients from the famous FRS study has been extracted from the University of Washington, Seattle, WA, USA ("<http://courses.washington.edu/b513/datasets/datasets.php?class=513>"). In the present study, the weightage and priority levels of parameters significant to cardiovascular disease is decided in accordance with the advice of medical panel comprising three experts. Table 4.1 show the various input parameters for the prediction of CVD.

Table 4.1: Parameters used for the prediction of heart disease

S.No.	Type of Risk Factors	Parameters
1.	Physical Risk Factors	Body Mass Index
2.		Age
3.		Gender
4.		Tobacco Smoke
5.		Alcoholism
6.		Family History
7.		Stress
8.		Systolic Blood Pressure (SBP)
9.		Diastolic Blood Pressure (DBP)
10.		Total Cholesterol (TC)
11.		C- reactive Protein (hs- CRP)
12.		Brain Natriuretic Peptide

Proposed Predictive and Adoptive Algorithm(s) for Atherosclerosis and Myocardial Infarction

S.No.	Type of Risk Factors	Parameters
13.	Physiological Risk Factors	Homocysteine
14.		Sodium
15.		Potassium
16.		Bicarbonate
17.		Chloride
18.		Calcium
19.		Fibrinogen
20.		Platelet Count
21.		Aspartate transaminase (SGOT)
22.		Creatine Kinase
23.		Creatine Phosphokinase-MB
24.		Troponin
25.	Clinical Risk Factors	Diabetes Mellitus (Random Blood sugar)
26.		Low Density Lipoprotein levels (LDL)
27.		High Density Lipoprotein levels (HDL)
28.		Triglycerides levels

The proposed system has been trained and was evaluated using the data by incorporating the advice of Cardiologist/ Physician regarding the clinical diagnosis for various corresponding dataset. During development of this CVD prediction system, the dataset has been divided into training, validation and testing as shown in Figure 4.1.

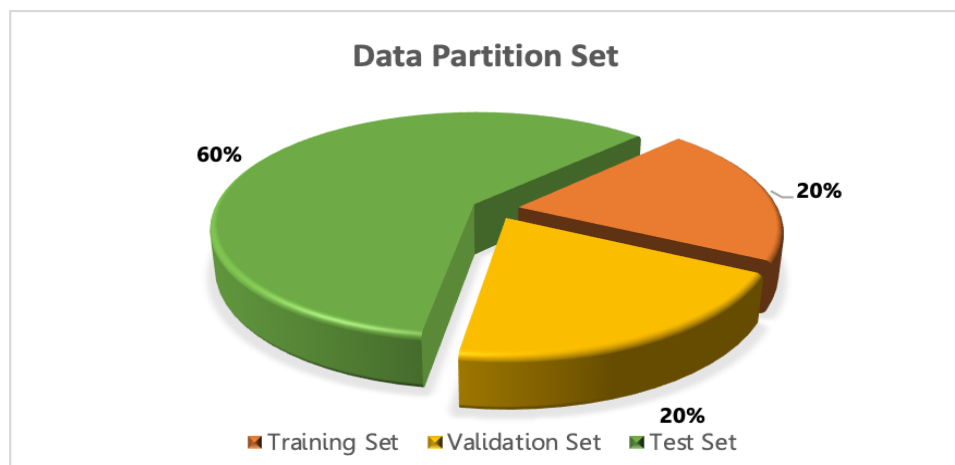


Figure 4.1 Data Partition Set

Proposed Predictive and Adoptive Algorithm(s) for Atherosclerosis and Myocardial Infarction

During our study, we have consulted several physicians and identified four groups of CVD risk categories according to percentage level as "normal", "low risk range", "intermediate risk range" and "highest risk range". The percentage for each level is shown in Table 4.2

Table 4.2 Percentage level of CVD risk categories

S.No.	Risk category	Percentage
1.	Normal (under baseline risk)	<26
2.	Low risk range	26- 51
3.	Intermediate risk range	51- 76
4.	Highest risk range	≥ 76

During data analysis, percentage wise patient distribution has been categorised into Lower, Average and High Level. As shown in Figure 4.2, out of 815 patients, 63% of total number of patients had High Blood pressure while 29% patients showed normal blood pressure values and the remaining 8% had low blood pressure readings. The total cholesterol content was high in 82% patients, while 12% exhibited normal values of cholesterol and 6% exhibited low cholesterol. The 31% patients have high C - reactive protein (hs CRP), 50% patients have normal C - reactive protein (hs CRP) and 19% patients have low C - reactive protein (hs CRP). All the 100% patients exhibited low Brain Natriuretic Peptide (BNP) while 63% patients have high homocysteine. The 75% patients have high sodium and 89% patients have high Potassium. Similarly, 94% patients have high Bicarbonate and 63% patients have high Chloride. The fibrinogen content is high in 50% patients while 22% patients have high platelet count. Only 10% patients exhibited high Aspartate transaminase (SGOT). Creatine kinase was reported to the high in 14% of the cohort while 53% patients reported to have high Troponin. Beside the above discussed parameters, the 85.50 % Patients were on Medication.

Proposed Predictive and Adoptive Algorithm(s) for Atherosclerosis and Myocardial Infarction

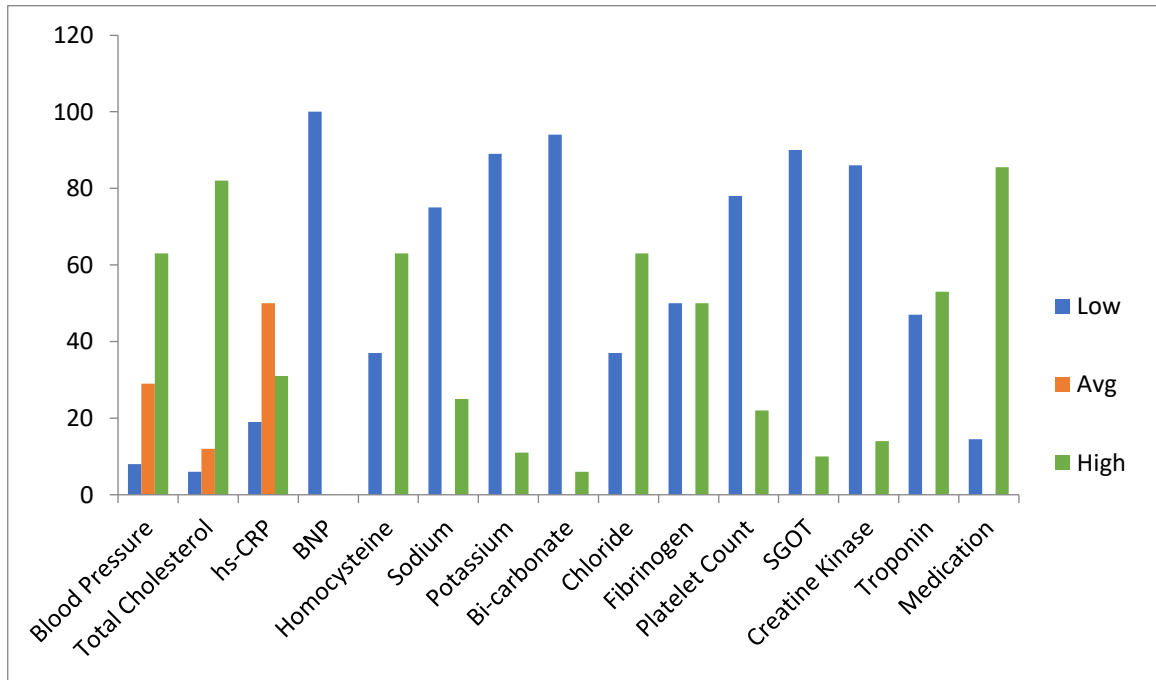


Figure 4.2 Bar Diagram showing percentage of patients with various parameters

4.3 Architecture of Proposed Cardiovascular Disease Prediction System

As discussed above, the QNN based CVD prediction system is a potential solution for the development of highly accurate and quick CVD prediction system. The QNN has the ability to learn the patterns, in present study the patterns of the heart diseases mainly from historical data and hence leading to CVD prediction after meticulous analysis of different patterns. The Architecture diagram of the proposed system is shown below in Figure 4.3 and the conceptual activity diagram of the proposed CVD prediction system is shown in Figure 4.4.

Proposed Predictive and Adoptive Algorithm(s) for Atherosclerosis and Myocardial Infarction

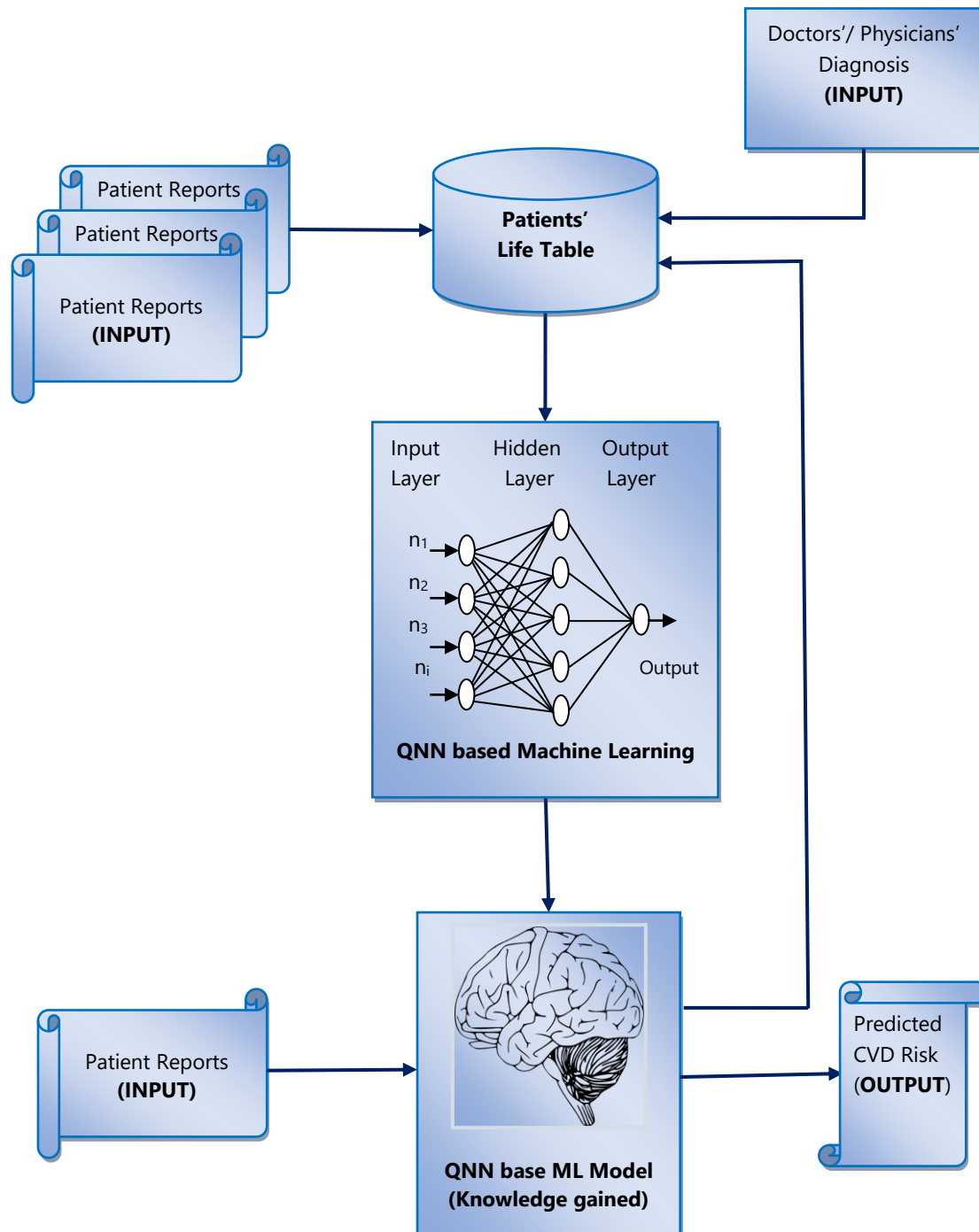


Figure 4.3 Architecture diagram of proposed system

Proposed Predictive and Adoptive Algorithm(s) for Atherosclerosis and Myocardial Infarction

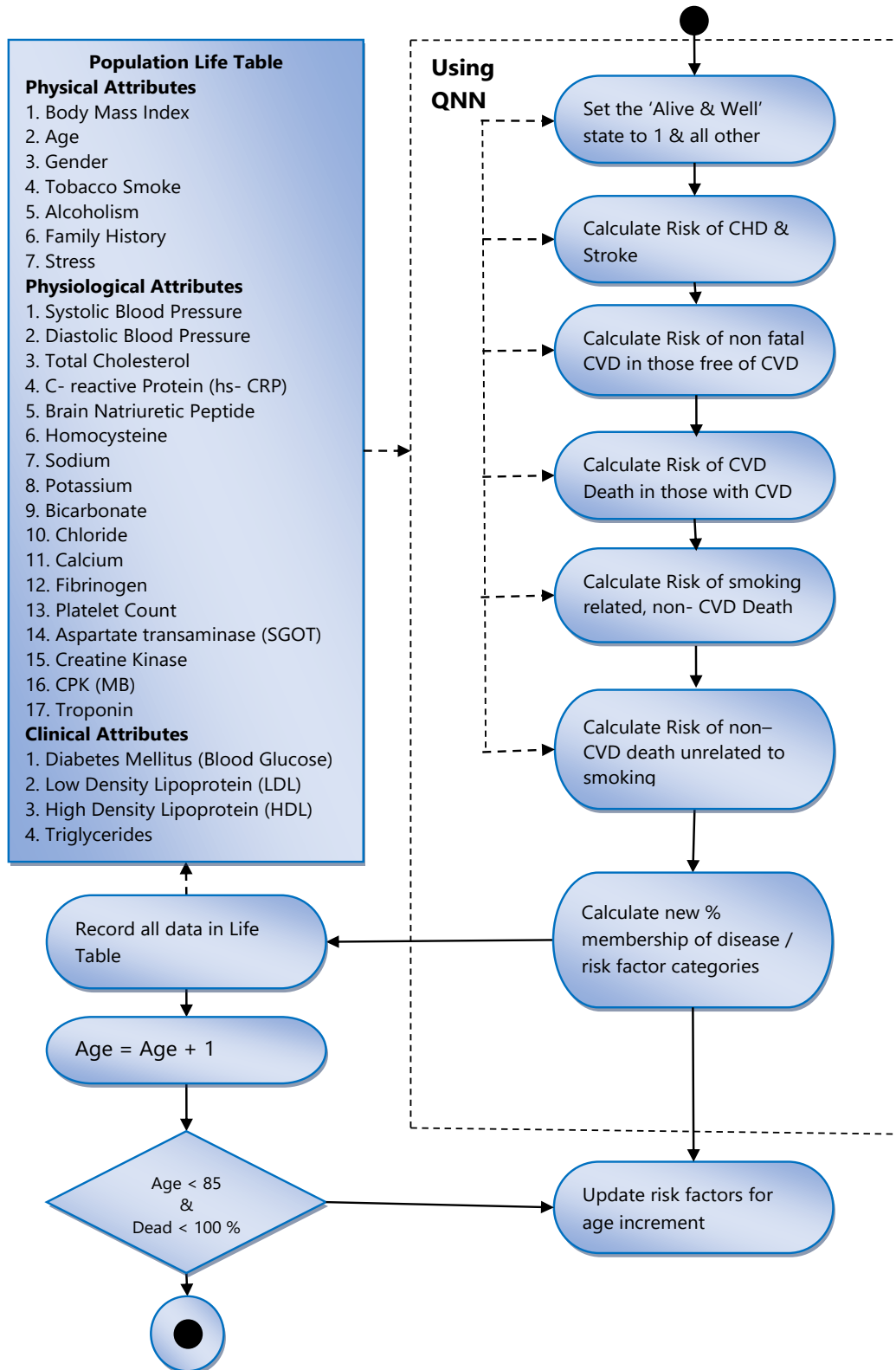


Figure 4.4 Activity diagram of proposed system

Proposed Predictive and Adoptive Algorithm(s) for Atherosclerosis and Myocardial Infarction

4.3.1 Proposed Algorithm of QNN based CVD Prediction System

Quantum Neural Network (QNN) based on multi-level transfer function was introduced by Karayiannis et al. (2006 a, b). In comparison to Artificial Neural Network (ANN), the hidden units of QNN uses the nonlinear activation functions instead of linear as in ANN. Here the nonlinear activation function consists of linear superposition of multi-sigmoid functions. Hence, more states can be expressed in hidden layer as compared to traditional sigmoid function in which only two states may be expressed. Recently in few years/ past, various models of QNN have been used in other research areas like, weather prediction, disease diagnosis, voice recognition, machine translation etc (Zhou et al. 1999, Karayiannis et al. 2006 a, b, Narayan et al. 2013, Narayan et al. 2014, Jeswal et al. 2018).

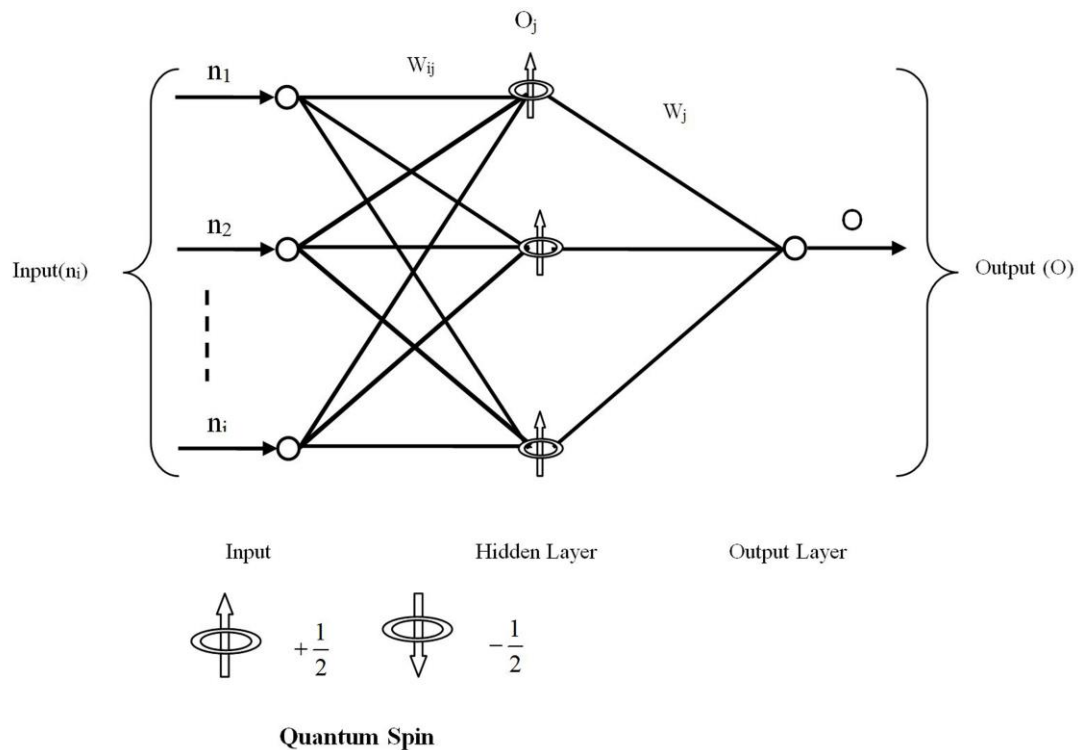


Figure.4.5 Architecture of Quantum Neural Network

Here the three layers QNN has been used, in which one input layer, one hidden layer and one output layer used, as shown in the Figure 4.5. Here, each neuron of hidden layer

Proposed Predictive and Adoptive Algorithm(s) for Atherosclerosis and Myocardial Infarction

contains two excitation states, which virtually represent two more nodes in itself, with the difference of quantum interval θ^r with quantum level r . The n_s denotes the number of excitation levels, η is learning rate, which should be a small random value, the error rate of output layer is denoted by δ_k and the error rate of hidden layer is δ_j . The input to the input layer is denoted by n_i . The output of hidden is denoted by O_j . Similarly, the output of output layer is denoted by O_k , respectively. W_{ij} denotes the weights between input and hidden layers and W_{kj} denotes the weights between hidden and output layers. During the learning, the small random numbers must be assigned to initial weights and t denotes target value (Narayan et al. 2014 a, b, c, 2016).

Given R training pairs $\{n_1, t_1; n_2, t_2; \dots; n_R, t_R\}$

Where n_i ($J \times 1$) is input and t_i ($K \times 1$) is target values for given inputs, as shown in the figure 4.4.

The error signal term of the output (δ_k) and hidden layers (δ_j) are written as,

$$\delta_k = (t_k - O_k) O_k (1 - O_k) \quad \{k=1, 2, 3, \dots, K\} \quad \dots \quad (4.1)$$

$$\delta_j = O_j (1 - O_j) \sum_k W_{kj}^{New} \delta_k \quad \{j=1, 2, 3, \dots, J\} \text{ and } \{k=1, 2, 3, \dots, K\} \quad \dots \quad (4.2)$$

Consequently, output layer weights (W_{kj}^{New}) and hidden layer weights (W_{ji}^{New}) are adjusted as.

$$W_{kj}^{New} = W_{kj}^{old} + \eta \delta_k O_j \quad \{j=1, 2, 3, \dots, J\} \text{ and } \{k=1, 2, 3, \dots, K\} \quad \dots \quad (4.3)$$

$$W_{ji}^{New} = W_{ji}^{old} + \eta \delta_j O_i \quad \{i=1, 2, 3, \dots, I\} \text{ and } \{j=1, 2, 3, \dots, J\} \quad \dots \quad (4.4)$$

Proposed Predictive and Adoptive Algorithm(s) for Atherosclerosis and Myocardial Infarction

The simple sigmoid function has been used as the activation function from input to hidden layer and is expressed as:

$$sgm(x) = (1/(1 + \exp(-x))) \quad \dots\dots\dots (4.5)$$

The sigmoid function with various graded levels has been used as the activation function for each hidden neuron i.e. from hidden to output layer is expressed as follows:

$$sgm(x) = \sum_{r=1}^{n_s} (1/(1 + \exp(-x + \theta^r))) \quad \dots\dots\dots (4.6)$$

Here the initial value of θ^r is initialized by zero, then apply Eq1 to 6 and iterate this process several times upto i iterations until it reaches the lowest possible error and then apply Eq 4.7.

After i iterations, when minimum possible error is obtained, then increase quantum interval by very small interval $\Delta\theta$

$$\theta^r = \theta^r + \Delta\theta \quad \dots\dots\dots (4.7)$$

Repeat this step until δ_k is decreased to an acceptable accuracy.

The Step by Step details of the above discussed algorithm is expressed by the flow chart as shown in Figure 4.6

Proposed Predictive and Adoptive Algorithm(s) for Atherosclerosis and Myocardial Infarction

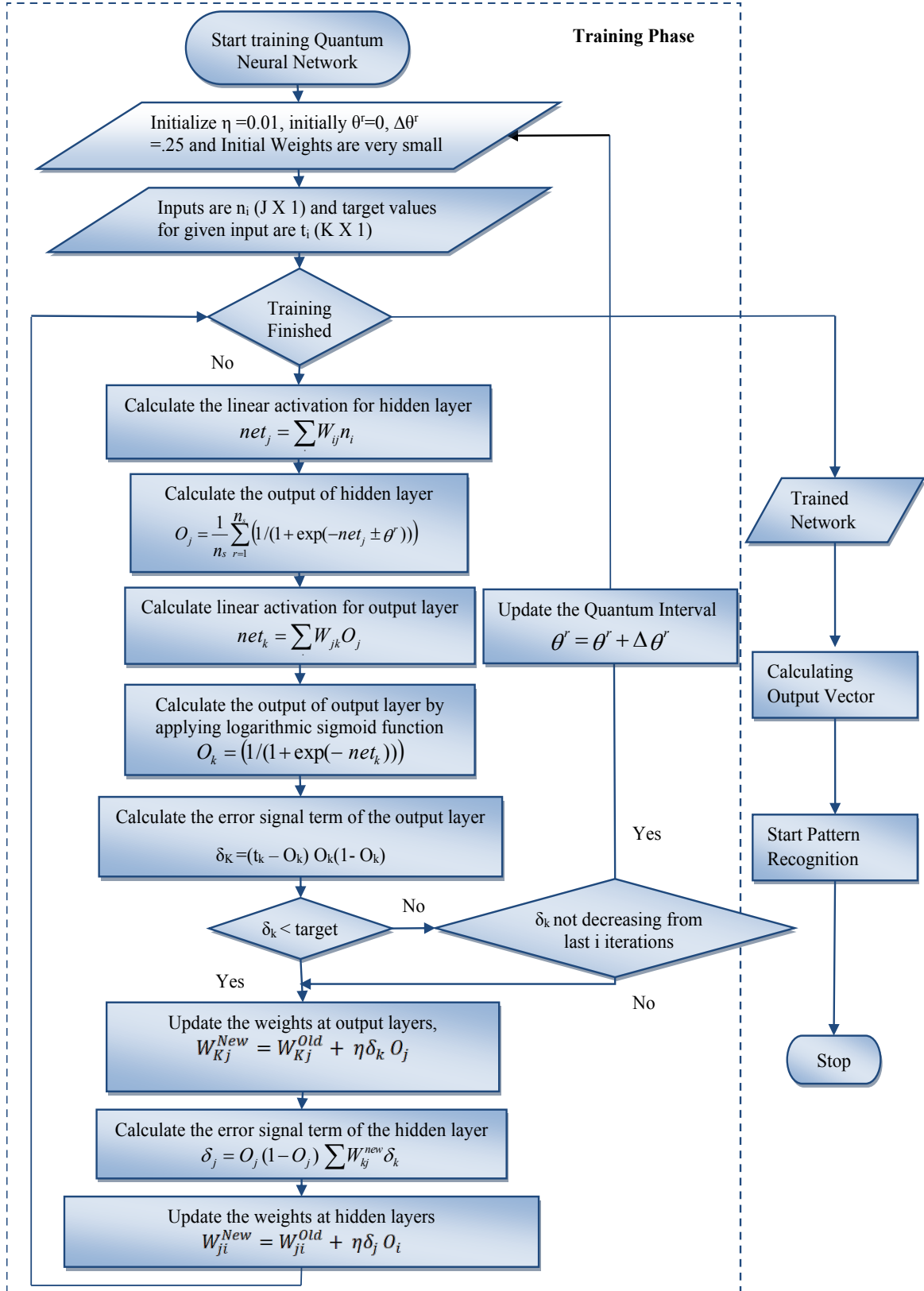


Figure 4.6: Flow chart of Quantum Neural Network for Heart Disease Prediction System

Proposed Predictive and Adoptive Algorithm(s) for Atherosclerosis and Myocardial Infarction

4.4 Conclusion

In this chapter, the architecture of proposed Cardiovascular Disease Prediction System (CDPS) along with the algorithm has been discussed in detail. Quantum neural network (QNN) which is the foundation of current study has also been extensively reviewed here with the thorough discussion going from basics of QNN with its properties inherited from Quantum physics and traditional Artificial neural network, to its functioning and implementation. Furthermore, the complete flow charts of Activity diagram and of principal Quantum Neural network algorithm has been incorporated with all the details. Using the QNN algorithm/ flow chart, the proposed system has been implemented, which is further discussed in chapter 5 in detail.

CHAPTER 5

RESULTS AND DISCUSSIONS

Results and Discussions

5.1. Introduction

In this chapter, the proposed Cardiovascular Disease (CVD) prediction system has been analyzed and tested. To test and validate the proposed model, the data set of 815 patient records has been utilized along with the dataset of 5209 Cardiovascular disease patients from the famous Framingham study has been extracted from the University of Washington, Seattle, WA, USA ("<http://courses.washington.edu/b513/datasets/datasets.php?class=513>"). To find out the optimal/ ideal/ prime learning rule for the proposed CVD prediction system various tests have been conducted/ performed with the aim to identify optimum/ ideal number of hidden nodes and Quantum interval. The details of the tests which have been performed are as follows.

5.2. Experiment and Results

To evaluate the proposed system's performance, different experiments have been conducted, using the primary data set of 815 patients and the secondary dataset of 5209 patients. As shown in the figure 5.1. It is identified that the optimal number of hidden nodes for the proposed CVD prediction system is 85.

The Quantum Interval (θ^r) for Equation 4.6 has been optimized for achieving the optimal learning rule. The same experiment has been performed several times for the distinct values of Quantum Interval (θ^r), to find out the optimal result/ value. This experiment is performed with the random data sets for training, validation and testing from the primary and secondary data set of patients. The results appended below in Table 5.1 are the average taken after 50 times of calculations. In Table 5.1, Quantum Interval (θ^r) 0.25 shows the best performance respecting all parameters i.e. Epoch or iterations required to train the network, the training, validation and test performance.

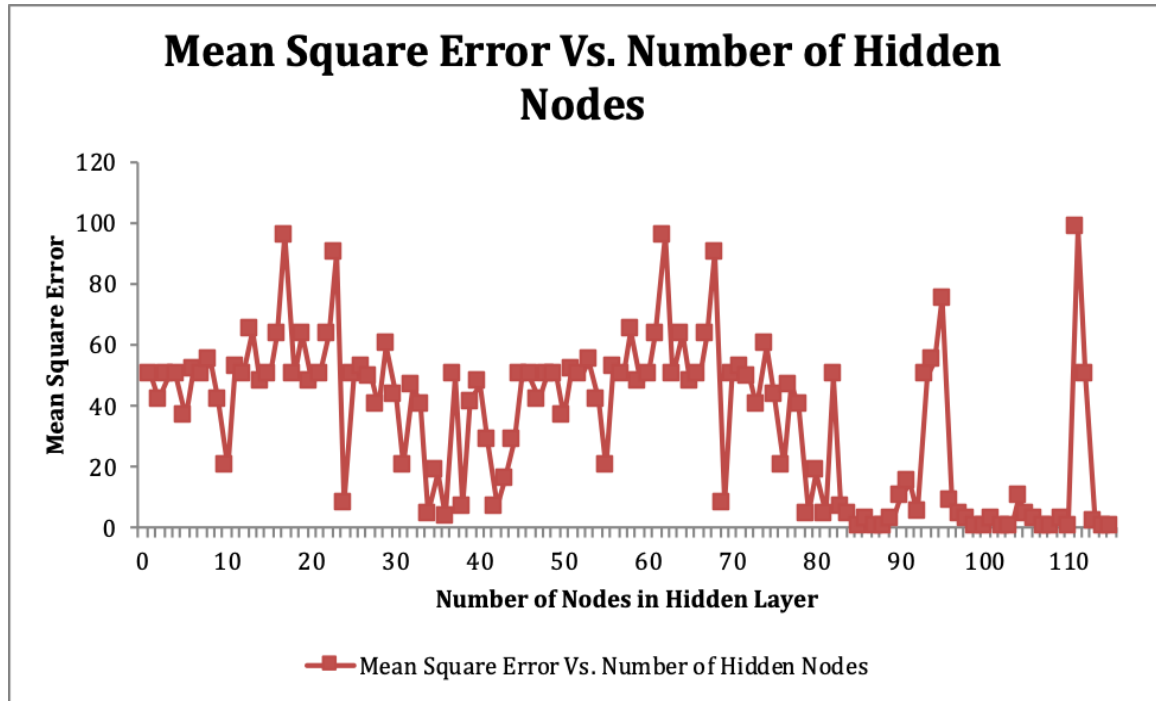


Figure 5.1: Mean square Error versus number of hidden nodes in QNN for proposed CVD prediction System

The first row of table i.e. at $\theta = 0$ represents the training value of ANN model and rest of the rows shows the training performances of QNN model with different value of θ^r .

Table 5.1: Results of performance parameter for QNN and ANN

S.No.	Quantum Interval (θ)	Epoch (average Iterations)	Training performance (MSE)	Validation performance (MSE)	Test performance (MSE)
1	$\theta=0 \Rightarrow$ ANN	237.6966	0.003152	0.003147	0.003137
2	0.25	178.8762	0.000042	0.000042	0.000042
3	0.5	180.0399	0.000351	0.000352	0.000351
4	0.75	178.3053	0.000126	0.000215	0.000225
5	1	178.4371	0.000471	0.000472	0.000478
6	1.25	174.968	0.000376	0.000385	0.000384
7	1.5	179.4031	0.000251	0.000257	0.00025
8	1.75	182.6746	0.000392	0.000395	0.000386
9	2	185.0459	0.000128	0.000125	0.000127

Results and Discussions

S.No.	Quantum Interval (θ)	Epoch (average Iterations)	Training performance (MSE)	Validation performance (MSE)	Test performance (MSE)
10	2.25	187.1536	0.000258	0.000257	0.000256
11	2.5	185.1117	0.000124	0.000124	0.000126
12	2.75	191.2814	0.000204	0.000204	0.000204
13	3	199.6906	0.000105	0.000105	0.000105

5.2.1. Experiments in context to proposed Cardiovascular Disease Prediction

System using Physical parameters

The proposed Cardiovascular Disease Prediction System (CDPS) based on QNN has been tested using the physical parameters like Age, Smoking, Body Mass Index (BMI), Stress, Alcohol, Gender, Family History along with other physiological parameters like Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP) etc, Using the above mentioned parameters the proposed system showed an average accuracy of 98% in predicting CVD risk category, as shown in the Table 5.2.

Table 5.2 (Part-1) Random Testing of different experimental values using Physical parameters

S.No.	Age (1)	Smoking (2)	SBP (3)	DBP (4)	BMI (5)	Stress (6)	Alcohol (7)
1	30	No	113	78	26	No	No
2	35	No	117	80	27	No	No
3	40	No	112	79	25	No	No
4	40	Yes	136	86	39	Yes	Yes
5	45	No	127	84	30	No	No
6	50	No	118	81	26	No	No
7	55	No	125	84	27	No	No
8	60	No	121	83	25	No	No
9	65	Yes	138	88	32	Yes	Yes
10	70	Yes	140	89	32	Yes	Yes

Results and Discussions

S.No.	Age (1)	Smoking (2)	SBP (3)	DBP (4)	BMI (5)	Stress (6)	Alcohol (7)
11	20	No	109	75	26	No	No
12	25	No	121	80	30	No	No
13	40	No	115	80	26	No	No
14	40	Yes	128	84	34	Yes	Yes
15	45	No	123	82	28	No	No
16	50	No	121	82	27	No	No
17	55	No	128	85	28	No	No
18	60	Yes	136	87	32	Yes	Yes
19	65	Yes	141	89	35	Yes	Yes
20	70	No	134	88	28	No	No
21	25	No	115	78	27	No	No
22	25	Yes	129	82	37	Yes	Yes
23	35	No	114	79	26	No	No
24	35	Yes	128	83	35	Yes	Yes
25	40	Yes	133	85	37	Yes	Yes
26	45	Yes	139	88	40	Yes	Yes
27	50	No	129	85	30	No	No
28	65	No	122	84	25	No	No
29	65	No	136	88	30	No	No
30	70	No	136	88	29	No	No
31	20	No	113	76	27	No	No
32	25	Yes	133	84	40	Yes	Yes
33	35	Yes	132	85	38	Yes	Yes
34	40	Yes	131	85	36	Yes	Yes
35	45	Yes	130	85	34	Yes	Yes
36	45	Yes	138	87	39	Yes	Yes
37	55	Yes	139	88	37	Yes	Yes
38	60	No	124	84	26	No	No
39	65	No	127	85	27	No	No

Results and Discussions

S.No.	Age (1)	Smoking (2)	SBP (3)	DBP (4)	BMI (5)	Stress (6)	Alcohol (7)
40	70	No	131	87	27	No	No
41	20	No	116	77	28	No	No
42	25	No	117	78	28	No	No
43	30	No	118	79	28	No	No
44	35	Yes	135	86	40	Yes	Yes
45	45	Yes	132	85	35	Yes	Yes
46	50	Yes	134	86	35	Yes	Yes
47	55	Yes	141	89	38	Yes	Yes
48	60	Yes	137	88	34	Yes	Yes
49	65	No	129	86	27	No	No
50	70	No	132	87	28	No	No
51	20	Yes	122	79	32	Yes	Yes
52	30	No	112	77	26	No	No
53	35	No	122	81	29	No	No
54	40	No	117	80	27	No	No
55	45	No	117	80	26	No	No
56	50	No	123	83	27	No	No
57	55	No	129	85	29	No	No
58	60	Yes	142	89	37	Yes	Yes
59	70	No	127	86	26	No	No
60	70	Yes	145	91	36	Yes	Yes
61	25	No	109	76	25	No	No
62	25	Yes	124	81	34	Yes	Yes
63	30	Yes	128	83	36	Yes	Yes
64	40	No	119	81	27	No	No
65	45	Yes	133	86	36	Yes	Yes
66	50	Yes	136	87	36	Yes	Yes
67	55	Yes	142	89	39	Yes	Yes
68	60	Yes	139	88	35	Yes	Yes

Results and Discussions

S.No.	Age (1)	Smoking (2)	SBP (3)	DBP (4)	BMI (5)	Stress (6)	Alcohol (7)
69	65	No	131	86	28	No	No
70	70	No	129	86	27	No	No
71	25	No	114	77	27	No	No
72	25	Yes	126	81	35	Yes	Yes
73	30	Yes	130	83	37	Yes	Yes
74	40	No	120	81	28	No	No
75	45	Yes	135	86	37	Yes	Yes
76	50	Yes	137	87	37	Yes	Yes
77	55	Yes	144	90	40	Yes	Yes
78	60	No	129	85	28	No	No
79	65	No	133	87	28	No	No
80	70	Yes	147	92	38	Yes	Yes
81	20	No	120	79	30	No	No
82	20	Yes	131	82	39	Yes	Yes
83	25	Yes	132	83	39	Yes	Yes
84	30	Yes	127	82	35	Yes	Yes
85	35	No	111	78	25	No	No
86	35	Yes	131	84	37	Yes	Yes
87	40	Yes	127	83	32	Yes	Yes
88	50	Yes	132	86	34	Yes	Yes
89	55	No	118	82	25	No	No
90	65	No	126	85	26	No	No
91	20	No	111	76	26	No	No
92	25	No	112	77	26	No	No
93	30	No	116	79	27	No	No
94	35	No	113	78	26	No	No
95	35	No	120	81	28	No	No
96	45	No	115	80	25	No	No
97	55	No	120	82	26	No	No

Results and Discussions

S.No.	Age (1)	Smoking (2)	SBP (3)	DBP (4)	BMI (5)	Stress (6)	Alcohol (7)
98	60	Yes	141	89	36	Yes	Yes
99	70	No	125	85	26	No	No
100	70	Yes	143	90	35	Yes	Yes
101	25	Yes	127	82	36	Yes	Yes
102	30	Yes	131	84	38	Yes	Yes
103	40	No	122	82	28	No	No
104	45	No	121	82	27	No	No
105	50	No	127	84	29	No	No
106	55	Yes	134	87	34	Yes	Yes
107	60	Yes	147	91	40	Yes	Yes
108	65	Yes	147	91	39	Yes	Yes
109	70	Yes	149	92	39	Yes	Yes
110	20	No	119	78	29	No	No
111	20	Yes	132	83	40	Yes	Yes
112	30	No	115	78	27	No	No
113	40	No	114	79	26	No	No
114	40	Yes	137	87	40	Yes	Yes
115	50	No	120	82	26	No	No
116	55	No	126	84	28	No	No
117	60	No	123	83	26	No	No
118	65	Yes	139	89	34	Yes	Yes
119	70	Yes	142	90	34	Yes	Yes
120	20	No	114	77	27	No	No
121	30	No	110	77	25	No	No
122	35	Yes	134	85	39	Yes	Yes
123	40	No	123	82	29	No	No
124	45	No	124	83	28	No	No
125	55	No	121	83	26	No	No
126	55	Yes	136	87	35	Yes	Yes

Results and Discussions

S.No.	Age (1)	Smoking (2)	SBP (3)	DBP (4)	BMI (5)	Stress (6)	Alcohol (7)
127	65	Yes	149	92	40	Yes	Yes
128	70	Yes	151	93	40	Yes	Yes
129	20	Yes	129	82	38	Yes	Yes
130	25	Yes	130	83	38	Yes	Yes
131	30	Yes	125	82	34	Yes	Yes
132	35	Yes	129	84	36	Yes	Yes
133	40	No	125	83	30	No	No
134	50	Yes	142	89	40	Yes	Yes
135	55	Yes	138	88	36	Yes	Yes
136	60	No	134	87	30	No	No
137	65	No	124	84	26	No	No
138	70	No	123	85	25	No	No
139	20	Yes	128	81	37	Yes	Yes
140	25	Yes	123	80	32	Yes	Yes
141	35	No	116	79	27	No	No
142	40	Yes	134	86	38	Yes	Yes
143	45	No	126	83	29	No	No
144	50	No	116	81	25	No	No
145	55	No	123	83	27	No	No
146	60	No	126	84	27	No	No
147	65	Yes	143	90	36	Yes	Yes
148	70	No	138	89	30	No	No
149	20	No	117	78	28	No	No
150	20	Yes	125	80	35	Yes	Yes
151	25	No	120	79	29	No	No
152	30	Yes	124	81	32	Yes	Yes
153	30	Yes	134	85	40	Yes	Yes
154	35	No	119	80	28	No	No
155	45	Yes	129	84	32	Yes	Yes

Results and Discussions

S.No.	Age (1)	Smoking (2)	SBP (3)	DBP (4)	BMI (5)	Stress (6)	Alcohol (7)
156	50	Yes	131	85	32	Yes	Yes
157	60	No	132	86	29	No	No
158	20	Yes	123	80	34	Yes	Yes
159	25	No	118	79	28	No	No
160	30	No	119	80	28	No	No
161	35	No	123	82	30	No	No
162	45	No	118	81	26	No	No
163	50	No	124	83	28	No	No
164	50	Yes	140	88	39	Yes	Yes
165	55	No	131	86	30	No	No
166	60	No	128	85	27	No	No
167	60	Yes	144	90	38	Yes	Yes
168	25	No	111	76	26	No	No
169	30	No	121	80	29	No	No
170	35	Yes	125	82	32	Yes	Yes
171	45	No	120	81	27	No	No
172	50	No	126	84	28	No	No
173	55	Yes	133	86	32	Yes	Yes
174	60	Yes	145	90	39	Yes	Yes
175	65	Yes	144	90	37	Yes	Yes
176	65	Yes	146	91	38	Yes	Yes
177	70	Yes	146	91	37	Yes	Yes
178	20	No	108	75	25	No	No
179	20	Yes	126	81	36	Yes	Yes
180	30	No	122	81	30	No	No
181	30	Yes	133	84	39	Yes	Yes
182	35	Yes	126	83	34	Yes	Yes
183	40	Yes	130	84	35	Yes	Yes
184	45	Yes	136	87	38	Yes	Yes

Results and Discussions

S.No.	Age (1)	Smoking (2)	SBP (3)	DBP (4)	BMI (5)	Stress (6)	Alcohol (7)
185	50	Yes	139	88	38	Yes	Yes
186	60	No	131	86	28	No	No
187	65	No	134	87	29	No	No

Table 5.2 (Part-2) Random Testing of different experimental values using Physical parameters

S.No.	Gender (8)	Family History (9)	Desired Result based on Doctor's Diagnosis	Predicted Result by Proposed System	
			Risk Category	Result QNN	Risk Category
1	Female	No	Normal	12	Normal
2	Female	No	Normal	25	Normal
3	Female	No	Normal	0	Normal
4	Male	Yes	Highest	93	Highest
5	Female	No	Low	50	Low
6	Female	No	Normal	6	Normal
7	Female	No	Normal	25	Normal
8	Female	No	Normal	0	Normal
9	Male	Yes	Intermediate	56	Intermediate
10	Male	Yes	Intermediate	56	Intermediate
11	Female	No	Normal	6	Normal
12	Female	No	Low	50	Low
13	Female	No	Normal	12	Normal
14	Male	Yes	Intermediate	62	Intermediate
15	Female	No	Low	31	Low
16	Female	No	Normal	18	Normal
17	Female	No	Low	37	Low
18	Male	Yes	Intermediate	56.2	Intermediate
19	Male	Yes	Intermediate	68.7	Intermediate
20	Male	No	Low	37.5	Low
21	Female	No	Normal	25	Normal

Results and Discussions

S.No.	Gender (8)	Family History (9)	Desired Result based on Doctor's Diagnosis	Predicted Result by Proposed System	
			Risk Category	Result QNN	Risk Category
22	Male	Yes	Highest	81.2	Highest
23	Female	No	Normal	12.5	Normal
24	Male	Yes	Intermediate	68.7	Intermediate
25	Male	Yes	Highest	81.2	Highest
26	Male	Yes	Highest	100	Highest
27	Female	No	Low	50	Low
28	Male	No	Normal	0	Normal
29	Male	No	Low	50	Low
30	Male	No	Low	43.7	Low
31	Female	No	Normal	18.7	Normal
32	Male	Yes	Highest	100	Highest
33	Male	Yes	Highest	87.5	Highest
34	Male	Yes	Intermediate	75	Intermediate
35	Male	Yes	Intermediate	62.5	Intermediate
36	Male	Yes	Highest	93.7	Highest
37	Male	Yes	Highest	81.2	Highest
38	Female	No	Normal	12.5	Normal
39	Male	No	Normal	18.7	Normal
40	Male	No	Normal	25	Normal
41	Female	No	Low	31.2	Low
42	Female	No	Low	31.2	Low
43	Female	No	Low	31.2	Low
44	Male	Yes	Highest	100	Highest
45	Male	Yes	Intermediate	68.7	Intermediate
46	Male	Yes	Intermediate	68.7	Intermediate
47	Male	Yes	Highest	87.5	Highest
48	Male	Yes	Intermediate	62.5	Intermediate
49	Male	No	Normal	25	Normal

Results and Discussions

S.No.	Gender (8)	Family History (9)	Desired Result based on Doctor's Diagnosis	Predicted Result by Proposed System	
			Risk Category	Result QNN	Risk Category
50	Male	No	Low	31.2	Low
51	Male	Yes	Intermediate	56.2	Intermediate
52	Female	No	Normal	6.2	Normal
53	Female	No	Low	43.7	Low
54	Female	No	Normal	18.7	Normal
55	Female	No	Normal	6.2	Normal
56	Female	No	Normal	25	Normal
57	Female	No	Low	43.7	Low
58	Male	Yes	Highest	81.2	Highest
59	Male	No	Normal	12.5	Normal
60	Male	Yes	Intermediate	75	Intermediate
61	Female	No	Normal	0	Normal
62	Male	Yes	Intermediate	62	Intermediate
63	Male	Yes	Intermediate	75	Intermediate
64	Female	No	Normal	25	Normal
65	Male	Yes	Intermediate	75	Intermediate
66	Male	Yes	Intermediate	75	Intermediate
67	Male	Yes	Highest	93.7	Highest
68	Male	Yes	Intermediate	68	Intermediate
69	Male	No	Low	31.2	Low
70	Male	No	Normal	18.5	Normal
71	Female	No	Normal	18.5	Normal
72	Male	Yes	Intermediate	68.7	Intermediate
73	Male	Yes	Highest	81.2	Highest
74	Female	No	Low	31.3	Low
75	Male	Yes	Highest	81.5	Highest
76	Male	Yes	Highest	81.2	Highest
77	Male	Yes	Highest	100	Highest

Results and Discussions

S.No.	Gender (8)	Family History (9)	Desired Result based on Doctor's Diagnosis	Predicted Result by Proposed System	
			Risk Category	Result QNN	Risk Category
78	Female	No	Low	31.5	Low
79	Male	No	Low	37.5	Low
80	Male	Yes	Highest	87.5	Highest
81	Female	No	Low	50	Low
82	Male	Yes	Highest	93.7	Highest
83	Male	Yes	Highest	93.5	Highest
84	Male	Yes	Intermediate	68.7	Intermediate
85	Female	No	Normal	0	Normal
86	Male	Yes	Highest	81.2	Highest
87	Male	Yes	Intermediate	56.2	Intermediate
88	Male	Yes	Intermediate	62.5	Intermediate
89	Female	No	Normal	0	Normal
90	Male	No	Normal	12.5	Normal
91	Female	No	Normal	12.5	Normal
92	Female	No	Normal	12.5	Normal
93	Female	No	Normal	25	Normal
94	Female	No	Normal	6.2	Normal
95	Female	No	Low	37.5	Low
96	Female	No	Normal	0	Normal
97	Female	No	Normal	6.2	Normal
98	Male	Yes	Intermediate	75	Intermediate
99	Male	No	Normal	6.2	Normal
100	Male	Yes	Intermediate	68.7	Intermediate
101	Male	Yes	Intermediate	75	Intermediate
102	Male	Yes	Highest	87.5	Highest
103	Female	No	Low	37.5	Low
104	Female	No	Normal	25	Normal
105	Female	No	Low	43.2	Low

Results and Discussions

S.No.	Gender (8)	Family History (9)	Desired Result based on Doctor's Diagnosis	Predicted Result by Proposed System	
			Risk Category	Result QNN	Risk Category
106	Male	Yes	Intermediate	62.5	Intermediate
107	Male	Yes	Highest	100	Highest
108	Male	Yes	Highest	93.5	Highest
109	Male	Yes	Highest	93.7	Highest
110	Female	No	Low	43.2	Low
111	Male	Yes	Highest	100	Highest
112	Female	No	Normal	18.5	Normal
113	Female	No	Normal	6.23	Normal
114	Male	Yes	Highest	100	Highest
115	Female	No	Normal	12.5	Normal
116	Female	No	Low	31.2	Low
117	Female	No	Normal	6.3	Normal
118	Male	Yes	Intermediate	62.5	Intermediate
119	Male	Yes	Intermediate	62.5	Intermediate
120	Female	No	Normal	25	Normal
121	Female	No	Normal	0	Normal
122	Male	Yes	Highest	93.7	Highest
123	Female	No	Low	43.2	Low
124	Female	No	Low	37.5	Low
125	Female	No	Normal	12.5	Normal
126	Male	Yes	Intermediate	68.3	Intermediate
127	Male	Yes	Highest	100	Highest
128	Male	Yes	Highest	100	Highest
129	Male	Yes	Highest	87.5	Highest
130	Male	Yes	Highest	87.5	Highest
131	Male	Yes	Intermediate	62.5	Intermediate
132	Male	Yes	Intermediate	75	Intermediate
133	Female	No	Low	50	Low

Results and Discussions

S.No.	Gender (8)	Family History (9)	Desired Result based on Doctor's Diagnosis	Predicted Result by Proposed System	
			Risk Category	Result QNN	Risk Category
134	Male	Yes	Highest	100	Highest
135	Male	Yes	Intermediate	75	Intermediate
136	Female	No	Low	50	Low
137	Male	No	Normal	6.3	Normal
138	Male	No	Normal	0	Normal
139	Male	Yes	Highest	81.5	Highest
140	Male	Yes	Intermediate	56.2	Intermediate
141	Female	No	Normal	18.5	Normal
142	Male	Yes	Highest	87.5	Highest
143	Female	No	Low	43.7	Low
144	Female	No	Normal	0	Normal
145	Female	No	Normal	18.5	Normal
146	Female	No	Normal	18.7	Normal
147	Male	Yes	Intermediate	75	Intermediate
148	Male	No	Low	50	Low
149	Female	No	Low	37.5	Low
150	Male	Yes	Intermediate	68.7	Intermediate
151	Female	No	Low	43.2	Low
152	Male	Yes	Intermediate	56.5	Intermediate
153	Male	Yes	Highest	100	Highest
154	Female	No	Low	31.2	Low
155	Male	Yes	Intermediate	56.5	Intermediate
156	Male	Yes	Intermediate	56.3	Intermediate
157	Female	No	Low	43.7	Low
158	Male	Yes	Intermediate	62.5	Intermediate
159	Female	No	Low	37.5	Low
160	Female	No	Low	37.5	Low
161	Female	No	Low	50	Low

Results and Discussions

S.No.	Gender (8)	Family History (9)	Desired Result based on Doctor's Diagnosis	Predicted Result by Proposed System	
			Risk Category	Result QNN	Risk Category
162	Female	No	Normal	12.5	Normal
163	Female	No	Low	31.2	Low
164	Male	Yes	Highest	93.7	Highest
165	Female	No	Low	50	Low
166	Female	No	Normal	25	Normal
167	Male	Yes	Highest	87.5	Highest
168	Female	No	Normal	6.2	Normal
169	Female	No	Low	43.7	Low
170	Male	Yes	Intermediate	56.5	Intermediate
171	Female	No	Normal	18.7	Normal
172	Female	No	Low	37.5	Low
173	Male	Yes	Intermediate	56.2	Intermediate
174	Male	Yes	Highest	93.7	Highest
175	Male	Yes	Highest	81.5	Highest
176	Male	Yes	Highest	87.5	Highest
177	Male	Yes	Highest	81.2	Highest
178	Female	No	Normal	0	Normal
179	Male	Yes	Intermediate	75	Intermediate
180	Female	No	Low	50	Low
181	Male	Yes	Highest	93.5	Highest
182	Male	Yes	Intermediate	62.5	Intermediate
183	Male	Yes	Intermediate	68.3	Intermediate
184	Male	Yes	Highest	87.5	Highest
185	Male	Yes	Highest	87.5	Highest
186	Female	No	Low	37.5	Low
187	Male	No	Low	43.2	Low

Results and Discussions

To identify the Impact of different CVD parameters on classification, the correlation analysis has been done with the data set of the CVD patents. The Impact wise ranking of CVD parameters affecting classification most are as follows in Table 5.3.

Table 5.3 Impact ranking of CVD parameters affecting classification

S.No.	Parameters	Rank
1.	Age	1
2.	BMI	
3.	Smoking	2
4.	Stress	
5.	Alcohol	
6.	Family History	
7.	SBP	3
8.	Gender	4
9.	DBP	5

5.2.2. Experiment with proposed Cardiovascular Disease Prediction System using Physiological parameters solely.

The experiment using the Physiological parameters has been conducted with the proposed CDPS. The physiological parameters which were taken into consideration during experimentation are Age, Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), Brain Natriuretic Peptide (BNP), Homocysteine, Total Cholesterol, Sodium, Potassium, Bicarbonate, Chloride, Calcium, High-sensitivity C-Reactive Protein (hs-CRP), Fibrinogen, Platelet Count, Serum Glutamic Oxaloacetic Transaminase (SGOT), Creatine Kinase, Troponin levels.

Using the above-mentioned physiological parameters, the CDPS showed an average accuracy of 99.53% in predicting CVD risk category, as shown in Table 5.4.

Results and Discussions

Table 5.4 (Part-1) Random Testing of different experimental values using

Physiological parameters

S.No.	Age (1)	SBP (2)	DBP (3)	TC (4)	hs-CRP (5)	BNP (6)	Homocysteine (7)	Sodium (8)	Potassium (9)	Bicarbonate (10)	Chloride (11)
1	65	217	166	345	4	275	6.7	148.1	4.8	22.3	103.5
2	25	121	80	200	1	125	3.9	136.8	3.6	18.6	96.5
3	65	191	141	315	3	238	6	145.3	4.5	21.4	101.8
4	30	134	85	250	2	150	4.2	138.7	3.8	19.2	97.5
5	50	160	110	285	2	200	5.5	142.5	4.2	20.5	100
6	45	133	86	225	1	138	4	137.8	3.7	18.9	96.7
7	65	149	92	250	2	150	4.2	138.7	3.8	19.2	97.5
8	15	160	110	315	3	238	6	145.3	4.5	21.4	101.7
9	45	200	150	345	4	275	6.7	148.1	4.8	22.3	103.5
10	20	182	133	345	4	275	6.8	148.1	4.8	22.7	103.7
11	25	200	150	355	4	288	6.9	149.0	4.9	22.9	104.3
12	60	226	174	365	4	300	7.2	150	5	23	105
13	35	191	141	335	4	263	6.4	147.9	4.7	22.6	103.2
14	35	200	150	345	4	275	6.7	148.3	4.1	22.7	103.5
15	15	111	75	140	1	113	3.5	135.9	3.9	18.2	95.6
16	45	139	88	250	2	150	4.2	138.6	3.7	19.5	97.5
17	55	144	90	250	2	150	4.2	138.7	3.8	19.2	97.5
18	20	138	85	265	2	175	4.7	140.6	4.0	19.8	98.7
19	50	226	174	365	4	300	7.2	150	5	23	105
20	45	165	117	305	3	225	5.7	144.3	4.4	21.1	101.5
21	55	151	100	265	2	175	4.6	140.3	4.0	19.8	98.7
22	60	147	91	250	2	150	4.7	138.6	3.8	19.2	97.5
23	15	147	95	285	2	200	5.5	142.1	4.2	20.5	100

Results and Discussions

S.No.	Age (1)	SBP (2)	DBP (3)	TC (4)	hs-CRP (5)	BNP (6)	Homocysteine (7)	Sodium (8)	Potassium (9)	Bicarbonate (10)	Chloride (11)
24	55	182	133	315	3	238	6	145.3	4.5	21.4	101.7
25	35	156	105	285	2	200	5.5	142.5	4.2	20.5	100
26	65	208	158	335	4	263	6.6	147.1	4.7	22.6	103.2
27	55	226	174	365	4	300	7.2	150	5	23	105
28	70	156	105	265	2	175	4.6	140.6	4.0	19.7	98.5
29	30	138	85	255	2	163	4.1	139.9	3.9	19.6	98.2
30	70	145	91	225	1	138	4.3	137.2	3.7	18.4	96.7
31	60	128	85	140	1	113	3.4	135.7	3.5	18.2	95.2

Table 5.4 (Part-2) Random Testing of different experimental values using

Physiological parameters

S.No.	Calcium (12)	Fibrinogen (13)	Platelet Count (14)	SGOT (15)	Creatine Kinase (16)	Troponin (17)	Desired Result based on Doctor's Diagnosis	Predicted Result by Proposed System	
							Risk category wise	Percentage wise	Risk category wise
1	2.7	375	4.1	18.2	155.2	0.8	Highest	87.5	Highest
2	2.1	225	1.7	6.7	42.5	0.2	Normal	12.5	Normal
3	2.5	337.5	3.6	15.1	127.2	0.8	Intermediate	68.7	Intermediate
4	2.2	250	2.2	8.5	61.5	0.2	Normal	25	Low
5	2.4	300	3	12.5	99	0.5	Low	50	Low
6	2.2	237.5	2.6	7.1	52.2	0.1	Normal	18.5	Normal

Results and Discussions

S.No.	Calcium (12)	Fibrinogen (13)	Platelet Count (14)	SGOT (15)	Creatine Kinase (16)	Troponin (17)	Desired Result based on Doctor's Diagnosis	Predicted Result by Proposed System	
							Risk category wise	Percentage wise	Risk category wise
7	2.2	250	2.5	8.5	61.5	0.2	Normal	25	Low
8	2.5	337.5	3.5	15.3	127.1	0.7	Intermediate	68.7	Intermediate
9	2.7	375	4.2	18.2	155.2	0.8	Highest	87.5	Highest
10	2.7	375	4.1	18.2	155.5	0.9	Highest	87.5	Highest
11	2.7	387.5	4.3	19.6	164.2	0.9	Highest	93.7	Highest
12	2.8	400	4.5	20	174	1	Highest	100	Highest
13	2.6	362.5	3.9	17.2	145.8	0.8	Highest	81.2	Highest
14	2.7	375	4.1	18.1	155.2	0.9	Highest	87.5	Highest
15	2.1	212.5	1.6	5.9	33.3	0.1	Normal	6.5	Normal
16	2.2	250	2.2	8.7	61.5	0.2	Normal	25	Normal
17	2.2	250	2.5	8.7	61.5	0.2	Normal	25	Low
18	2.3	275	2.6	10.6	80.2	0.3	Low	37.5	Low
19	2.8	400	4.5	20	174	1	Highest	100	Highest
20	2.5	325	3.3	14.7	117.7	0.6	Intermediate	62.5	Intermediate
21	2.3	275	2.6	10.2	80.2	0.4	Low	37.5	Low
22	2.2	250	2.2	8.75	61.5	0.5	Normal	25	Normal
23	2.4	300	3	12.5	99	0.5	Low	50	Low
24	2.5	337.5	3.5	15.1	127.1	0.6	Intermediate	68.7	Intermediate
25	2.4	300	3	12.5	99	0.5	Low	50	Low
26	2.6	362.5	3.9	17.8	145.8	0.8	Highest	81.5	Highest
27	2.8	400	4.5	20	174	1	Highest	100	Highest

Results and Discussions

S.No.	Calcium (12)	Fibrinogen (13)	Platelet Count (14)	SGOT (15)	Creatine Kinase (16)	Troponin (17)	Desired Result based on Doctor's Diagnosis	Predicted Result by Proposed System	
							Risk category wise	Percentage wise	Risk category wise
28	2.3	275	2.6	10.2	80.2	0.4	Low	37.5	Low
29	2.3	262.5	2.4	9.8	70.8	0.3	Low	31.3	Low
30	2.2	237.5	2.6	7.1	52.1	0.2	Normal	18.5	Normal
31	2.1	212.5	1.6	5.9	33.3	0.1	Normal	6.2	Normal

To identify the Impact of different CVD parameters on classification, the correlation analysis has been done with the data set of the CVD patents. The Impact wise ranking of CVD parameters affecting classification most are as follows in Table 5.5.

Table 5.5: Impact ranking of CVD parameters affecting classification

S.No.	Parameters	Rank
1.	BNP	Equal weightage/ impact in classification
2.	Fibrinogen	
3.	Creatine Kinase	
4.	SGOT	
5.	Troponin	
6.	Homocysteine	
7.	Total Cholesterol	
8.	hs-CRP	
9.	Chloride	
10.	Sodium	
11.	Bicarbonate	
12.	Potassium	

S.No.	Parameters	Rank
13.	Calcium	
14.	Platelet Count	

5.2.3. Experiment with proposed Cardiovascular Disease Prediction System using Clinical parameters

The clinical parameters have also been used for testing the proposed Cardiovascular Disease Prediction System (CDPS). The clinical parameters used were Diabetes, levels of High-Density Lipoprotein (HDL), Low density Lipoprotein (LDL), and Triglycerides. The proposed system showed an average accuracy of 97.97% with the Clinical parameters in predicting CVD risk category, as shown in Table 5.6.

Table 5.6 Random Testing of different experimental values using Clinical parameters

Diabetes	LDL	HDL	Triglycerides	Equated/ Calculated Desired Result		Desired Result based on Doctor's Diagnosis	Result and Accuracy of Proposed QNN based System	
				Out put	Risk Category		Out put	Risk Category
151	135.6	52.5	255.2	59.3	Intermediate	Intermediate	59.3	Intermediate
121	111.2	72.5	183.5	34.8	Low	Low	34.7	Low
188	154.3	27.5	442.7	90.6	Highest	Highest	90.6	Highest
95	85	90	161.5	12.5	Normal	Normal	12.5	Normal
80	70	100	149	0	Normal	Normal	0	Normal
201	160	20	499	100	Highest	Highest	100	Highest
117	107.5	75	180.5	31.5	Low	Low	31.5	Low

Results and Discussions

Diabetes	LDL	HDL	Triglycerides	Equated/ Calculated Desired Result		Desired Result based on Doctor's Diagnosis	Result and Accuracy of Proposed QNN based System	
				Out put	Risk Category		Out put	Risk Category
143	131.7	57.5	217.7	53.3	Intermediate	Intermediate	53.3	Intermediate
113	103.7	77.5	177.5	28.1	Low	Low	28.2	Low
91	81.2	92.5	158.3	9.3	Normal	Normal	9.3	Normal
196	158.1	22.5	480.5	96.8	Highest	Highest	96.8	Highest
139	130	60	199	50.5	Low	Low	50.5	Intermediate
162	141.2	45	311.5	68.5	Intermediate	Intermediate	68.5	Intermediate
83	73.7	97.5	152.1	3.1	Normal	Normal	3.1	Normal
192	156.2	25	461.5	93.7	Highest	Highest	93.7	Highest
106	96.2	82.5	170.8	21.8	Normal	Normal	21.8	Normal
128	118.7	67.5	189.6	40.6	Low	Low	40.6	Low
166	143.1	42.5	330.2	71.8	Intermediate	Intermediate	71.8	Intermediate
181	150.6	32.5	405.5	84.3	Highest	Highest	84.4	Highest
87	77.5	95	155.2	6.2	Normal	Normal	6.2	Normal
185	152.5	30	424	87.5	Highest	Highest	87.5	Highest
155	137.5	50	274	62.5	Intermediate	Intermediate	62.5	Intermediate
132	122.5	65	192.5	43.5	Low	Low	43.6	Low
125	115	70	186.5	37.5	Low	Low	37.5	Low
177	148.7	35	386.5	81.2	Highest	Highest	81.2	Highest
170	145	40	349	75	Intermediate	Intermediate	75	Intermediate
98	88.5	87.5	164.6	15.6	Normal	Normal	15.6	Normal
136	126.2	62.5	195.8	46.8	Low	Low	46.8	Low
110	100	80	174	25	Normal	Normal	25	Normal
158	139.3	47.5	292.7	65.6	Intermediate	Intermediate	65.6	Intermediate
173	146.5	37.5	367.5	78.1	Highest	Highest	78.1	Highest

Results and Discussions

Diabetes	LDL	HDL	Triglycerides	Equated/ Calculated Desired Result		Desired Result based on Doctor's Diagnosis	Result and Accuracy of Proposed QNN based System	
				Out put	Risk Category		Out put	Risk Category
102	92.5	85	167.8	18.7	Normal	Normal	18.7	Normal
147	133.5	55	236.5	56.5	Intermediate	Intermediate	56.5	Intermediate

To identify the Impact of different CVD parameters on classification, the correlation analysis has been done with the data set of the CVD patents. The Impact wise ranking of CVD parameters affecting classification most are as follows in Table 5.7.

Table 5.7: Impact ranking of CVD parameters affecting classification

S.No.	Parameters	Rank
1.	Diabetes	Equal weightage/ impact in classification
2.	LDL	
3.	Triglycerides	
4.	HDL	

5.2.4. Experiment with proposed Cardiovascular Disease Prediction System using Physical, Physiological and Clinical parameters

The proposed Cardiovascular Disease Prediction System (CDPS) based on QNN has been tested using the physical parameters like Age, Smoking, Body Mass Index (BMI), Stress, Alcohol, Gender, Family History, HDL, LDL, Triglycerides, TC, hs-CRP, BNP, Homocysteine, Sodium, Potassium, Bi-carbonate, Chloride, Calcium along with Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP) both parameters belong to physiological

Results and Discussions

parameters. Using the above mentioned parameters, the proposed system exhibits the average accuracy of 98.5% with Physical, Physiological and Clinical parameters in predicting CVD risk category, as shown in the Table 5.8.

Table 5.8 (Part-1) Random Testing of different experimental values using all parameters (Physical, Physiological and Clinical parameters)

S.No.	Age (1)	Gender (2)	SBP (3)	DBP (4)	Smoking (5)	BMI (6)	Stress (7)	Alcohol (8)	Family History (9)	DM (10)	LDL (11)	HDL (12)
1	25	Female	121	80	No	27	No	No	No	121	111.2	72.5
2	65	Female	191	141	No	25	No	No	No	188	154.5	27.5
3	30	Male	134	85	Yes	39	Yes	Yes	Yes	95	85	90
4	50	Female	160	110	No	30	No	No	No	80	70	100
5	45	Female	133	86	No	26	No	No	No	201	160	20
6	65	Female	149	92	No	27	No	No	No	117	107.5	75
7	15	Female	160	110	No	25	No	No	No	143	131.7	57.5
8	45	Male	200	150	Yes	32	Yes	Yes	Yes	113	103.5	77.5
9	20	Male	182	133	Yes	32	Yes	Yes	Yes	91	81.2	92.5
10	25	Female	200	150	No	26	No	No	No	196	158.1	22.5
11	60	Female	226	174	No	30	No	No	No	139	130	60
12	35	Female	191	141	No	26	No	No	No	162	141.2	45
13	35	Male	200	150	Yes	34	Yes	Yes	Yes	83	73.7	97.5
14	15	Female	111	75	No	28	No	No	No	192	156.5	25
15	45	Female	139	88	No	27	No	No	No	106	96.2	82.5
16	55	Female	144	90	No	28	No	No	No	128	118.7	67.5
17	20	Male	138	85	Yes	32	Yes	Yes	Yes	166	143.1	42.5
18	50	Male	226	174	Yes	35	Yes	Yes	Yes	181	150.6	32.5

Results and Discussions

S.No.	Age (1)	Gender (2)	SBP (3)	DBP (4)	Smoking (5)	BMI (6)	Stress (7)	Alcohol (8)	Family History (9)	DM (10)	LDL (11)	HDL (12)
19	45	Male	165	117	No	28	No	No	No	87	77.5	95
20	55	Female	151	100	No	27	No	No	No	185	152.5	30
21	60	Male	147	91	Yes	37	Yes	Yes	Yes	155	137.5	50
22	15	Female	147	95	No	26	No	No	No	132	122.5	65
23	55	Male	182	133	Yes	35	Yes	Yes	Yes	125	115	70
24	35	Male	156	105	Yes	37	Yes	Yes	Yes	177	148.7	35
25	65	Male	208	158	Yes	40	Yes	Yes	Yes	170	145	40
26	55	Female	226	174	No	30	No	No	No	98	88.5	87.5
27	70	Male	156	105	No	25	No	No	No	136	126.2	62.5
28	30	Male	138	85	No	30	No	No	No	110	100	80
29	70	Male	145	91	No	29	No	No	No	158	139.3	47.5
30	60	Female	128	85	No	27	No	No	No	173	146.8	37.5

Table 5.8 (Part-2) Random Testing of different experimental values using all parameters (Physical, Physiological and Clinical parameters)

S.No.	Triglycerides (13)	TC (14)	hs-CRP (15)	BNP (16)	Homocysteine (17)	Sodium (18)	Potassium (19)	Bicarbonate (20)	Chloride (21)	Calcium (22)
1	183.3	200	1	125	3.7	136.8	3.6	18.6	96.2	2.1
2	442.7	315	3	238	6	145.3	4.3	21.4	101.8	2.5
3	161.5	250	2	150	4.2	138.7	3.7	19.5	97.5	2.2
4	149	285	2	200	5.2	142.5	4.5	20.5	100	2.4
5	499	225	1	138	4.0	137.8	3.8	18.9	96.8	2.2
6	180.2	250	2	150	4.2	138.7	3.7	19.2	97.5	2.2

Results and Discussions

S.No.	Triglycerides (13)	TC (14)	hs-CRP (15)	BNP (16)	Homocysteine (17)	Sodium (18)	Potassium (19)	Bicarbonate (20)	Chloride (21)	Calcium (22)
7	217.7	315	3	238	6	145.3	4.3	21.4	101.7	2.5
8	177.1	345	4	275	6.7	148.1	4.1	22.3	103.5	2.7
9	158.3	345	4	275	6.7	148.1	4.1	22.3	103.5	2.7
10	480.2	355	4	288	6.9	149.0	4.9	22.6	104.7	2.7
11	199	365	4	300	7.2	150	5	23	105	2.8
12	311.5	335	4	263	6.4	147.1	4.7	22.6	103.2	2.6
13	152.1	345	4	275	6.7	148.1	4.8	22.7	103.5	2.7
14	461.5	140	1	113	3.5	135.9	3.5	18.2	95.6	2.1
15	170.8	250	2	150	4.2	138.6	3.8	19.5	97.5	2.2
16	189.6	250	2	150	4.2	138.7	3.8	19.2	97.5	2.2
17	330.2	265	2	175	4.7	140.3	4.6	19.7	98.5	2.3
18	405.2	365	4	300	7.2	150	5	23	105	2.8
19	155.5	305	3	225	5.7	144.8	4.3	21.2	101.5	2.5
20	424	265	2	175	4.6	140.6	4.6	19.7	98.5	2.3
21	274	250	2	150	4.2	138.7	3.7	19.5	97.5	2.2
22	192.7	285	2	200	5.5	142.5	4.5	20.5	100	2.4
23	186.5	315	3	238	6	145.3	4.3	21.4	101.7	2.5
24	386.6	285	2	200	5.2	142.5	4.5	20.5	100	2.4
25	349	335	4	263	6.6	147.1	4.1	22.6	103.2	2.6
26	164.6	365	4	300	7.2	150	5	23	105	2.8
27	195.8	265	2	175	4.6	140.6	4.6	19.7	98.7	2.3
28	174	255	2	163	4.5	139.9	3.6	19.6	98.1	2.3
29	292.7	225	1	138	4.3	137.8	3.8	18.4	96.8	2.2
30	367.7	140	1	113	3.5	135.9	3.9	18.2	95.6	2.1

Results and Discussions

Table 5.8 (Part-3) Random Testing of different experimental values using all parameters (Physical, Physiological and Clinical parameters)

S.No.	Fibrinogen (23)	Platelet Count (24)	SGOT (25)	Creatine Kinase (26)	Troponin (27)	Desired Result based on Doctor's Diagnosis	Predicted Result by Proposed System	
						Risk category wise	Percentage wise	Risk category wise
1	225	1.8	6.8	42.7	0.1	Normal	12.5	Normal
2	337.5	3.6	15.3	127.2	0.6	Intermediate	68.7	Intermediate
3	250	2.5	8.7	61.5	0.2	Normal	25	Low
4	300	3	12.5	99	0.5	Low	50	Low
5	237.5	2.6	7.8	52.1	0.1	Normal	18.5	Normal
6	250	2.5	8.7	61.5	0.2	Normal	25	Low
7	337.5	3.6	15.3	127.2	0.6	Intermediate	68.7	Intermediate
8	375	4.2	18.1	155.5	0.8	Highest	87.5	Highest
9	375	4.2	18.2	155.2	0.8	Highest	87.5	Highest
10	387.5	4.1	19.6	164.6	0.9	Highest	93.7	Highest
11	400	4.5	20	174	1	Highest	100	Highest
12	362.5	3.3	17.8	145.8	0.8	Highest	81.2	Highest
13	375	4.2	18.2	155.2	0.8	Highest	87.5	Highest
14	212.5	1.8	5.9	33.3	0.1	Normal	6.2	Normal
15	250	2.5	8.7	61.5	0.2	Normal	25	Normal
16	250	2.5	8.5	61.5	0.2	Normal	25	Low
17	275	2.2	10.6	80.2	0.3	Low	37.5	Low
18	400	4.5	20	174	1	Highest	100	Highest
19	325	3.3	14.7	117.7	0.6	Intermediate	62.5	Intermediate

Results and Discussions

S.No.	Fibrinogen (23)	Platelet Count (24)	SGOT (25)	Creatine Kinase (26)	Troponin (27)	Desired Result based on Doctor's Diagnosis	Predicted Result by Proposed System	
						Risk category wise	Percentage wise	Risk category wise
20	275	2.6	10.2	80.2	0.3	Low	37.5	Low
21	250	2.2	8.7	61.5	0.2	Normal	25	Normal
22	300	3	12.5	99	0.5	Low	50	Low
23	337.5	3.6	15.3	127.1	0.6	Intermediate	68.7	Intermediate
24	300	3	12.5	99	0.5	Low	50	Low
25	362.5	3.9	17.1	145.8	0.8	Highest	81.2	Highest
26	400	4.5	20	174	1	Highest	100	Highest
27	275	2.6	10.6	80.2	0.3	Low	37.5	Low
28	262.5	2.4	9.6	70.8	0.3	Low	31.5	Low
29	237.5	2.6	7.8	52.1	0.1	Normal	18.7	Normal
30	212.5	1.7	5.9	33.3	0.1	Normal	6.5	Normal

5.2.5. Experiment with proposed Cardiovascular Disease Prediction System using the parameters used in Framingham Risk Score

The parameters used in Framingham Risk Score were also considered for testing the proposed Cardiovascular Disease Prediction System (CDPS). The parameters used are Age, Gender, Smoking, Systolic Blood Pressure (SBP), HDL, Total Cholesterol (TC) and Smoker. The above mentioned parameters were used to evaluate and compare with FRS. During test, the proposed CVD prediction system exhibited the best results with accuracy on the dataset. Data used in the prediction of accuracy is done by the random testing on different experimental values are shown in Table 5.9. In predicting the CVD risk category, an overall accuracy of 98.57% has been achieved by the proposed system, whereas in FRS, the highest level of accuracy was 19.22%. This system's accuracy is significantly

Results and Discussions

higher as compared to other existing approaches.

Table 5.9 (Part-1) Random Testing of different experimental values using the parameters used in Framingham Risk Score

S.No.	Parameters Used for Comparison of Proposed System and FRS (Patient Record)							Manually Calculated Result	Equated/ Desired Result
	Age	Gender	TC	HDL	Smoker	SBP	Medication	Output	Risk Category
1	40	Male	4.2	2.2	No	117	Yes	18.8	Normal
2	35	Female	5.0	1.8	No	121	Yes	37.5	Low Risk
3	55	Female	5.3	1.7	No	132	Yes	43.8	Low Risk
4	45	Male	5.3	1.7	Yes	126	Yes	43.8	Low Risk
5	70	Male	5.8	1.4	No	140	Yes	56.3	Intermediate
6	65	Female	4.7	1.9	No	135	Yes	31.3	Low Risk
7	45	Male	7.2	0.8	No	190	Yes	87.5	Highest Risk
8	60	Male	5.3	1.7	No	132	Yes	43.8	Low Risk
9	50	Male	6.4	1.2	Yes	134	Yes	68.8	Intermediate
10	60	Female	4.2	2.2	No	130	Yes	18.8	Normal
11	25	Male	6.4	1.2	Yes	126	Yes	68.8	Intermediate
12	60	Male	4.2	2.2	No	126	Yes	18.8	Normal
13	20	Female	7.7	0.7	No	160	No	93.8	Highest Risk
14	35	Male	6.6	1.0	Yes	129	Yes	75.0	Intermediate
15	60	Male	8.3	0.5	No	190	Yes	100.0	Highest Risk
16	60	Female	7.7	0.7	No	180	Yes	93.8	Highest Risk
17	25	Female	8.3	0.5	No	165	Yes	100.0	Highest Risk
18	25	Male	5.8	1.4	Yes	123	Yes	56.3	Intermediate
19	45	Male	5.8	1.4	No	129	No	56.3	Intermediate

Results and Discussions

S.No.	Parameters Used for Comparison of Proposed System and FRS (Patient Record)							Manually Calculated Result	Equated/ Desired Result
	Age	Gender	TC	HDL	Smoker	SBP	Medication	Output	Risk Category
20	65	Female	7.7	0.7	No	170	Yes	93.8	Highest Risk
21	70	Male	5.3	1.7	Yes	136	Yes	43.8	low Risk
22	40	Female	4.7	1.9	No	125	Yes	31.3	Low Risk
23	65	Female	6.4	1.2	No	145	Yes	68.8	Intermediate
24	45	Female	5.0	1.8	No	122	Yes	37.5	Low Risk
25	25	Male	5.3	1.7	No	120	Yes	43.8	Low Risk
26	40	Female	6.1	1.3	No	125	Yes	62.5	Intermediate
27	40	Male	7.7	0.7	Yes	160	No	93.8	Highest Risk
28	65	Male	5.3	1.7	No	134	Yes	43.8	Low Risk
29	65	Male	3.7	2.5	Yes	124	Yes	6.3	Normal
30	55	Male	6.1	1.3	Yes	134	Yes	62.5	Intermediate
31	65	Female	5.5	1.6	No	140	Yes	50.0	Low Risk
32	55	Male	4.2	2.2	No	123	No	18.8	Normal
33	55	Male	6.9	0.9	No	139	Yes	81.3	Highest Risk
34	45	Male	8.3	0.5	Yes	170	No	100.0	Highest Risk
35	20	Male	8.3	0.5	No	160	Yes	100.0	Highest Risk
36	35	Male	8.3	0.5	Yes	175	Yes	100.0	Highest Risk
37	55	Female	5.8	1.4	No	135	Yes	56.3	Intermediate
38	30	Female	4.7	1.9	No	115	No	31.3	Low Risk
39	50	Male	7.7	0.7	Yes	175	No	93.8	Highest Risk
40	35	Female	5.3	1.7	No	120	Yes	43.8	Low Risk
41	60	Female	4.5	2.1	No	130	Yes	25.0	Normal
42	55	Female	3.9	2.3	No	125	Yes	12.5	Normal

Results and Discussions

S.No.	Parameters Used for Comparison of Proposed System and FRS (Patient Record)							Manually Calculated Result	Equated/ Desired Result
	Age	Gender	TC	HDL	Smoker	SBP	Medication	Output	Risk Category
43	60	Male	6.9	0.9	Yes	160	Yes	81.3	Highest Risk
44	40	Male	4.7	1.9	No	120	Yes	31.3	Low Risk
45	50	Male	4.5	2.1	No	123	Yes	25.0	Normal
46	35	Female	3.7	2.5	No	110	Yes	6.3	Normal
47	25	Female	4.2	2.2	No	110	Yes	18.8	Normal
48	20	Female	4.2	2.2	No	110	Yes	18.8	Normal
49	50	Female	4.5	2.1	No	125	Yes	25.0	Normal
50	40	Female	6.4	1.2	No	129	Yes	68.8	Intermediate
51	65	Male	6.9	0.9	No	175	No	81.3	Highest Risk
52	55	Male	3.9	2.3	No	121	Yes	12.5	Normal
53	25	Female	6.4	1.2	No	125	Yes	68.8	Intermediate
54	40	Female	4.5	2.1	No	115	Yes	25.0	Normal
55	30	Female	5.8	1.4	No	120	Yes	56.3	Intermediate
56	60	Male	5.5	1.6	No	134	Yes	50.0	Low Risk
57	50	Male	6.9	0.9	Yes	190	Yes	81.3	Highest Risk
58	50	Female	6.6	1.0	No	180	Yes	75.0	Intermediate
59	35	Male	4.2	2.2	No	116	Yes	18.8	Normal
60	55	Female	4.5	2.1	No	124	Yes	25.0	Normal
61	35	Female	6.6	1.0	No	160	Yes	75.0	Intermediate
62	20	Male	5.8	1.4	Yes	122	Yes	56.3	Intermediate
63	55	Female	6.6	1.0	No	140	No	75.0	Intermediate
64	70	Male	6.9	0.9	Yes	180	Yes	81.3	Highest Risk
65	40	Female	5.5	1.6	No	126	Yes	50.0	Low Risk

Results and Discussions

S.No.	Parameters Used for Comparison of Proposed System and FRS (Patient Record)							Manually Calculated Result	Equated/ Desired Result
	Age	Gender	TC	HDL	Smoker	SBP	Medication	Output	Risk Category
66	45	Female	4.7	1.9	No	120	Yes	31.3	Low Risk
67	25	Female	4.7	1.9	No	115	Yes	31.3	Low Risk
68	30	Male	3.7	2.5	Yes	112	No	6.3	Normal

Table 5.9 (Part-2) Random Testing of different experimental values using the parameters used in Framingham Risk Score

S.No	Desired Result based on Doctor's Diagnosis	Result and Accuracy of Proposed QNN based System		% Accuracy QNN	Accuracy Group-wise	Result FRS	% Accuracy FRS	Accuracy Group-wise
		Output	Risk Category					
1	Normal	18.7	Normal	99.7	99.9	0.9	5.2	19.6
2	Low Risk	37.5	Low Risk	100		0.7	2.0	
3	Low Risk	43.8	Low Risk	100		6.3	14.4	
4	Low Risk	43.7	Low Risk	99.8		9.4	21.6	
5	Intermediate	56.2	Intermediate	99.9		27.7	49.2	
6	Low Risk	31.2	Low Risk	99.9		8.3	26.7	
7	Highest Risk	87.5	Highest Risk	100		15.6	17.8	
8	Low Risk	43.7	Low Risk	99.8	99.9	13.4	30.8	20.4
9	Intermediate	68.8	Intermediate	100		23.3	33.9	
10	Normal	18.8	Normal	100		4.5	24	
11	Intermediate	68.8	Intermediate	100		1.6	2.4	
12	Normal	18.7	Normal	99.7		7.4	39.6	

Results and Discussions

S.No	Desired Result based on Doctor's Diagnosis	Result and Accuracy of Proposed QNN based System		% Accuracy QNN	Accuracy Group-wise	Result FRS	% Accuracy FRS	Accuracy Group-wise
		Output	Risk Category					
13	Highest Risk	93.8	Highest Risk	100		0.9	0.9	
14	Intermediate	75.0	Intermediate	100		8.6	11.5	
15	Highest Risk	100.0	Highest Risk	100	99.9	46.5	46.5	30.5
16	Highest Risk	93.8	Highest Risk	100		31.2	33.3	
17	Highest Risk	100.0	Highest Risk	100		3.6	3.6	
18	Intermediate	56.2	Intermediate	99.9		1	1.7	
19	Intermediate	56.2	Intermediate	99.9		6.2	11.1	
20	Highest Risk	93.8	Highest Risk	100		35.4	37.8	
21	low Risk	43.7	low Risk	99.8		34.8	79.7	
22	Low Risk	31.3	Low Risk	100	99.9	1.2	4.0	15.7
23	Intermediate	68.8	Intermediate	100		20.0	29.2	
24	Low Risk	37.5	Low Risk	100		2.1	5.7	
25	Low Risk	43.7	Low Risk	99.8		0.1	0.3	
26	Intermediate	62.5	Intermediate	100		3.0	4.9	
27	Highest Risk	93.7	Highest Risk	99.9		24.3	25.9	
28	Low Risk	43.7	Low Risk	99.8		17.5	40.1	
29	Normal	6.3	Normal	100	99.9	14.5	23.2	31
30	Intermediate	62.5	Intermediate	100		26.3	42.1	
31	Low Risk	49.8	Low Risk	99.6		13.1	26.2	
32	Normal	18.7	Normal	99.7		5	26.6	
33	Highest Risk	81.3	Highest Risk	100		23.8	29.4	
34	Highest Risk	100.0	Highest Risk	100		38.3	38.3	
35	Highest Risk	100.0	Highest Risk	100	98.3	1.1	1.1	13.9
36	Highest Risk	100.0	Highest Risk	99.9		21.3	21.3	
37	Intermediate	50.8	Low Risk	90.2		8.8	15.7	

Results and Discussions

S.No	Desired Result based on Doctor's Diagnosis	Result and Accuracy of Proposed QNN based System		% Accuracy QNN	Accuracy Group-wise	Result FRS	% Accuracy FRS	Accuracy Group-wise
		Output	Risk Category					
38	Low Risk	31.3	Low Risk	100		0.2	0.7	
39	Highest Risk	93.7	Highest Risk	99.9		40.0	42.7	
40	Low Risk	43.8	Low Risk	100		0.9	2.0	
41	Normal	25.0	Normal	100	99.9	5.1	20.6	16.9
42	Normal	12.5	Normal	100		2.5	20.1	
43	Highest Risk	81.3	Highest Risk	100		44.3	54.5	
44	Low Risk	31.2	Low Risk	99.8		1.5	5.1	
45	Normal	25.0	Normal	100		3.9	15.7	
46	Normal	6.3	Normal	100		0.1	2.7	
47	Normal	18.8	Normal	100		0.1	0.2	
48	Normal	18.8	Normal	100	99.2	0.1	0.1	13.8
49	Normal	24.0	Normal	96.0		2.4	9.9	
50	Intermediate	68.8	Intermediate	100		4.1	5.9	
51	Highest Risk	81.3	Highest Risk	100		36.1	44.3	
52	Normal	12.5	Normal	100		4.1	32.7	
53	Intermediate	68.0	Intermediate	98.9		0.6	0.8	
54	Normal	25.0	Normal	100		0.7	2.9	
55	Intermediate	56.3	Intermediate	100	99.9	0.7	1.2	14.9
56	Low Risk	50.0	Low Risk	100		15.5	31.1	
57	Highest Risk	81.3	Highest Risk	100		30.2	37.2	
58	Intermediate	75.0	Intermediate	100		11.2	14.9	
59	Normal	18.7	Normal	99.7		0.4	2.4	
60	Normal	25.0	Normal	100		3.3	13.4	
61	Intermediate	75.0	Intermediate	100		2.8	3.8	
62	Intermediate	56.2	Intermediate	99.9	99.9	0.2	0.4	15.0

Results and Discussions

S.No	Desired Result based on Doctor's Diagnosis	Result and Accuracy of Proposed QNN based System		% Accuracy QNN	Accuracy Group-wise	Result FRS	% Accuracy FRS	Accuracy Group-wise
		Output	Risk Category					
63	Intermediate	75.0	Intermediate	100		14.6	19.5	
64	Highest Risk	81.3	Highest Risk	100		56.3	69.3	
65	Low Risk	50.0	Low Risk	100		2.2	4.4	
66	Low Risk	31.2	Low Risk	99.9		1.7	5.5	
67	Low Risk	31.3	Low Risk	100		0.1	0.3	
68	Normal	6.3	Normal	100		0.3	5.7	

5.3. Methods used for Evaluation and Validation of proposed QNN based Cardiovascular Disease Prediction System

The Framingham Risk Score and European heart score are commonly used tools, which are considered most effective methods for CVD risk prediction in present time. However, these methods have some major drawbacks such as being less accurate and based on a very old study conducted in the 1960s wherein the parameters taken into consideration were also limited. The validity of the Framingham equation is undetermined as the lifestyle of the population has changed completely. The newly available risk parameters are not included. On the other hand, these systems are based on conventional statistical methods and are not recently updated. The major drawbacks of these studies are their ineffectiveness with regard to value ranges, i.e., the FRS is only applicable for patients aged between 20 to 100 years and the maximum threshold limit of CVD risk is 30%. Presently, the accuracy of the FRS is only 19.22%, as evaluated and compared with the proposed system using the same data, as shown in Figure 5.2

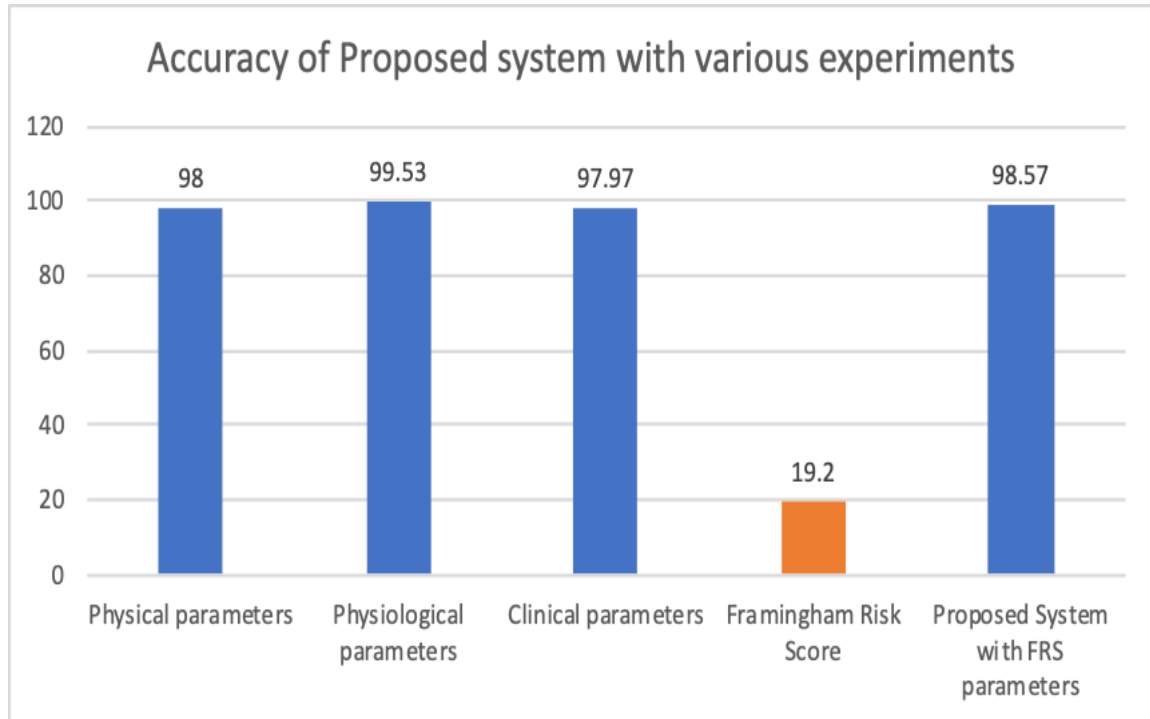


Figure 5.2 Accuracy of proposed system with various experiments

As shown in figure 5.2, the accuracy achieved in predicting the CVD risk category with different classes of parameters is 98% with Physical, 99.53% with Physiological, 97.97% with Clinical and 98.57% with FRS parameters.

On the other hand, in European heart score, if blood pressure is over 180 mmHg or cholesterol over 8 mmol/L, patients are automatically classified as high risk without considering the medication factor (European Society of Cardiology, 2016). In many situations, with proper and effective medication, a patient with risk parameters such as high blood pressure or high cholesterol levels may survive for a very long time.

Other than the FRS and European heart score, the fuzzy-evidential based CVD prediction system has been proposed by Khatibi and Montazer in 2010, but they have achieved only 91.5% accuracy (Khatibi et al. 2010). Akay proposed the neural network based coronary artery disease prediction system in 1992, which achieved an accuracy of 84%, they used the records of 100 patients for testing, but only clinical parameters were considered/ included in their study (Akay 1992). The Bayesian classification and neural network based ischemic heart disease prediction system was proposed by Kukar et al. in 1999, but in

Results and Discussions

their study, ECG data was used as input parameter for prediction (Kukar 1999). The probability theory (logistic regression) angiographic coronary disease based system was proposed by Detrano et al. in 1989, they collected the data of 425 patients and achieved only 77% accuracy. In their study, only clinical and non-invasive test data were considered as input parameters for their system (Detrano et al. 1989).

In the current study, we have introduced a new machine learning based system which uses the QNN for learning and for its knowledge buildup. On the basis of knowledge gained from various data patterns, the proposed system predicts the CVD risk with very high accuracy. Over a period of time, as the data increase, it will update and correct itself on the basis of the current trends and data patterns.

The proposed CVD system has been tested and validated using Indian and American patients' data, and achieved the highest accuracy among existing systems

5.4. Conclusion

As compared and evaluated during experiments, the proposed CVD prediction system shows the best results on the dataset used, when compared to the FRS. This system has been evaluated and validated with the data of patients and the doctors' diagnoses (predictions). This system can serve as an aid to medical practitioners in early diagnosis and planning better medication strategies. Data showing prediction accuracy of random testing on different experimental values are shown in Table 5.6. In predicting the CVD risk category, an overall accuracy of 98.5% has been achieved by the proposed system, whereas with the FRS the highest level of accuracy was 19.22%. This system's accuracy is considerably higher as compared to other existing approaches. Thus, this system must be used instead of well-known FRS or other existent tools. Hence, the proposed system will serve as a significant tool for doctors/practitioners.

CHAPTER 6

CONCLUSIONS AND FUTURE SCOPE

Conclusions and Future Scope

6.1 Conclusion

In this thesis, machine learning based intelligent Cardiovascular Disease (CVD) prediction system has been used for predicting the risk of CVD. The architecture of the proposed Cardiovascular Disease Prediction System (CDPS) has been discussed in detail. Moreover, the method used to implement the machine learning as QNN is also included. Several experiments have been performed to identify the optimal learning rules, used in QNN for proposed Cardiovascular Disease Prediction System (CDPS). After optimizing the value of Quantum Interval (θ^r) viz. Equation 4.6 with the different values of Quantum Interval (θ^r), the optimal learning rule has been achieved. The same experiment has been performed several times to find out the optimal value of Quantum Interval (θ^r). During experimentations, the results of different experiments have also been compared with conventional ANN and quantum neural networks.

The accuracy of this Cardiovascular Disease Prediction System (CDPS) has been enhanced significantly by incorporating QNN.

During testing, patients' data alongwith the doctors' diagnosis (predictions) were used for evaluation and validation. The CVD risk predictions done with the proposed system, using the dataset of the Framingham study, confirmed the possible risk of death, deaths which actually took place and has been listed as due to myocardial infarction (MI) and coronary heart disease (CHD) in the dataset of the Framingham study.

The proposed system attained 98.5% accuracy in predicting the CVD risk category. The performance of proposed system has also been evaluated together with other existing Cardiovascular Disease Prediction system/ methods like Framingham. The accuracy of the proposed system is considerably higher than the FRS and other existing approaches.

Conclusions and Future Scope

The proposed system will serve as an excellent tool for a medical practitioner in predicting the risk of CVD as well as in planning better medication and treatment strategies. The use of this system may effectively help in making an early diagnosis. The main key unique aspects of this research are given below:

- An early diagnosis may be effectively made by using this system, as in this research newly available CVD parameters have been used, which are not used in combination in any of the existing research. Due to use of various advanced CVD parameters the efficacy of this system has been increased tremendously.
- This system uses QNN as machine learning algorithm for classification of CVD, which gives it a built-in capability to learn new data over the time i.e. system can update itself based on the new dataset. The most important aspect of using QNN is, it works very well with the ambiguous data.

Thus, the accuracy is considerably higher compared to the other existing approaches. Thus, this system must be used instead of the well-known FRS.

6.2 Future Scope

Although the proposed Cardiovascular Disease Prediction System (CDPS) is showing higher accuracy using the Quantum Neural Network based machine learning approach, but still there is some possibility for enhancement.

The future directions are given below:

Upgraded Models:

Even though using huge data, enhanced models may perform better, in terms of predicting the Cardiovascular Disease (CDPS).

BIBLIOGRAPHY

Bibliography:

- [1]. "CDC Diabetes Cost-effectiveness Group, Cost-effectiveness of intensive glycemic control, intensified hypertension control, and serum cholesterol level reduction for type 2 diabetes", (2002), JAMA, Vol. 287(2542-51).
- [2]. "The Prospective Cardiovascular Munster Study (PROCAM), Prevalence of hyperlipidemia in persons with hypertension and/or diabetes mellitus and the relationship to coronary heart disease", (1988), Am Heart J, Vol. 116, pp. 1713-24.
- [3]. Agency for Healthcare Research and Quality. 2006 National Healthcare Quality Report. Rockville, MD: U.S. Department of Health and Human Services, Agency for Healthcare Research and Quality, 2006. AHRQ Pub. No. 07-0013.
- [4]. Akay M., (1992), "Noninvasive diagnosis of coronary artery disease using a neural network algorithm", Biol Cybern, Vol. 67(4), pp. 361-367.
- [5]. Akkala A., Bhatt D., Devabhaktuni V., Kumar A., (2013), "Knowledge-based neural network approaches for modeling and estimating radon concentrations", Environmental Progress and Sustainable Energy, Vol. 32(2), pp. 355-364.
- [6]. Almdal T., Scharling H., Jensen J.S., Vestergaard H., (2012), "The Independent Effect of Type 2 Diabetes Mellitus on Ischemic Heart Disease, Stroke, and Death", Arch Intern Med, Vol. 164, pp. 1422-1426.
- [7]. American Diabetes Association. Detection and management of lipid disorders in diabetes. Diabetes Care, 1993, Vol. 16 (5), pp.828-834.
- [8]. Anderson K.M., Wilson P.W., Odell P.M., Kannel W.B., (1991), "An updated coronary risk profile", A statement for health professionals. Circulation, Vol. 83 (1), pp. 356-362.
- [9]. Anooj P.K., (2012), "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules", Journal of King Saud University – Computer and Information Sciences, Vol. 24(1), pp. 27-40.

- [10]. Artigao-Rodenas L.M., Carbayo-Herencia J.A., Divisón-Garrote J.A., Gil-Guillén V.F., Massó-Orozco J., (2013), "Framingham Risk Score for Prediction of Cardiovascular Diseases: A Population-Based Study from Southern Europe", PLoS ONE, Vol. 8(9): e73529.
- [11]. Assmann G., Schulte H., von Eckardstein A., Huang Y., (1996), "High-density lipoprotein cholesterol as a predictor of coronary heart disease risk. The PROCAM experience and patho-physiological implications for reverse cholesterol transport.", *Atherosclerosis*, Vol. 124 (Suppl), pp. S11–S20.
- [12]. Ay H., Arsava E.M., Saribas O., Creatine Kinase-M.B., (2002) "Elevation After Stroke is not Cardiac in Origin Comparison with Troponin T Levels", *Stroke.*, Vol.33(1), pp. 286-9.
- [13]. Babuin L., Jaffe A.S., (2005), "Troponin: the biomarker of choice for the detection of cardiac injury", *CMAJ*, Vol. 173 (10), pp. 178- 185.
- [14]. Bachmann J.M., Willis B.L., Ayers C.R., Khera A., Berry J.D., (2012), "Association between Family History and Coronary Heart Disease Death across Long-Term Follow-Up in Men: The Cooper Center Longitudinal Study", *Circulation*. 2012, Vol. 125, pp. 3092-3098.
- [15]. Bala P., Lesyng B., McCammon J.A., (1994), "Applications of quantum-classical and quantum-stochastic molecular dynamics simulations for proton transfer processes", *Chemical Physics*, Vol. 180(2-3), pp. 271-285.
- [16]. Barrett-Connor E.L., Cohn B.A., Wingard D.L., Edelstein S.L., (1991), "Why Is diabetes mellitus a stronger risk factor for fatal ischemic heart disease in women than in men? The Rancho Bernardo Study", *JAMA*, Vol. 265(5), pp. 627–631.
- [17]. Bartechi C.E., MacKenzie T.D., Schrier R.W., (1994), "The human costs of tobacco use (first of two parts)", *N Engl J Med*, Vol. 330, pp. 907–12.
- [18]. Bashir S., Qamar U., Khan F.H., Naseem L., (2016), "HMF: A medical decision support framework using multi-layer classifiers for disease prediction", *Journal of Computational Science*, Vol. 13, pp. 10-25.

- [19]. Belder A. J. de, Jewitt D.E., Wainwright R.J., Thomas M.R., (2001), "Development and validation of a Bayesian index for predicting major adverse cardiac events with percutaneous transluminal coronary angioplasty", *Heart*, Vol. 85, pp.69-72.
- [20]. Bergoderi M.C., Thomas R.J., Albuquerque F.N., Batsis J.A., Burdiat G., Perez-Terzic C., Trejo-Gutierrez J., and Lopez-Jimenez F., (2012), "Validity of cardiovascular risk prediction models in Latin America and among Hispanics in the United States of America: a systematic review", *Panam Salud Publica*, Vol. 32(2), pp. 131-139.
- [21]. Bhopal R., Fischbacher C., Vartiainen E., Unwin N., White M. and Alberti G., (2005), "Predicted and Observed Cardiovascular Disease in South Asians: Application of FINRISK, Framingham and SCORE Models to Newcastle Heart Project data", *J Public Health*, Vol. 27(1), pp.93–100.
- [22]. Brown M.S., Goldstein J.L., (1976), "Receptor-mediated control of cholesterol metabolism", *Science*, Vol. 191, pp.150–154.
- [23]. Capewell S., Ford E.S., Croft J.B., Critchley J.A., Greenlund K.J., Labarthe D.R., (2010), "Cardiovascular risk factor trends and potential for reducing coronary heart disease mortality in the United States of America", *Bull World Health Organ*, Vol.88(2), pp. 120–130.
- [24]. Cardiovascular disease risk factors: new areas for research. Report of a WHO Scientific Group. WHO Technical Report Series 841. Geneva: World Health Organization, 1994.
- [25]. Chang Kim H., Greenland P., Rossouw J.E., Manson J.E., Cochrane B., Lasser N.L., Limacher M.C., Lloyd-Jones D.M., Margolis K.L., Robinson J.G., (2010), "Multimarker Prediction of Coronary Heart Disease Risk", *Journal of American College of Cardiology*, Vol. 55 (19), pp. 2080-2091.
- [26]. Chang S.W., Kareem S.A., Merican A.F., Zain R.B., (2013), "Oral cancer prognosis based on clinicopathologic and genomic markers using a hybrid of feature selection and machine learning methods", *BMC Bioinformatics*, Vol. 14(170), pp. 1-15.

- [27]. Chen M., Han J., and Yu P. S., (1996), "Data Mining: An Overview from Database Perspective", IEEE Trans. Knowl. Dat. Eng., Vol. 8(6), pp. 866-883.
- [28]. Cole C.R., Blackstone E.H., Pashkow F.J., (1999), "Heart rate recovery immediately after exercise as a predictor of mortality", N Engl J Med, Vol.341(1351 -7).
- [29]. Conroya R.M., Pyorola K., Fitzgerald A.P., Sansc S., Menottid A., De Backere G., De Bacquere D., Ducimetie`ref P., Jousilahtig P., Keilh U., Njølstadi I., Oganovj R.G., Thomsenk T., Tunstall-Pedoel H., Tverdalm A., Wedeln H., Whincupo P., Wilhelmsenn L., Grahama I.M., (2003), "Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project", European Heart Journal, Vol. 24, pp. 987-1003.
- [30]. D'Agostino R.B., Vasan R.S., Pencina M.J., (2008), "General cardiovascular risk profile for use in primary care: the Framingham Heart Study", Circulation, 2008, Vol.117 (6), pp. 743-753.
- [31]. Delaney Brendan C., Fitzmaurice David A., Amjid C., Richard H. (1999), "Can computerized decision support systems deliver improved quality in primary care?", BMJ, Vol. 319, pp. 1281 – 1284.
- [32]. Detrano R., Janosi A., Steinbrunn W., (1989), "International application of a new probability algorithm for the diagnosis of coronary artery disease", Am J Cardiol, Vol. 64(5), pp. 304–310.
- [33]. Dewey F.E., B.A, Freeman J.V., Engel G., Oviedo R., Abrol N., Ahmed N., Myers J., Froelicher V.F., (2007), "Novel predictor of prognosis from exercise stress testing: Heart rate variability response to the exercise treadmill test", Am Heart J, Vol. 153, pp. 281 -285.
- [34]. Dimsdale J.E., (2008), "Psychological Stress and Cardiovascular Disease", J Am Coll Cardiol, Vol. 51(13), pp. 1237–1246.
- [35]. Djoussé L., Lee M., Buring J.E., Gaziano M.J., (2009), "Alcohol Consumption and Risk of Cardiovascular Disease and Mortality in Women: Potential Mediating Mechanisms", Circulation. 2009, Vol. 120(3), pp. 237–244.

- [36]. Doll R., Peto R., Wheatley K., Gray R., Sutherland I., (1994), "Mortality in relation to smoking: 40 years' observations on male British doctors", *BMJ*, Vol. 309(6959), pp. 901–911.
- [37]. Dudina A., Cooney M.T., Bacquer D.D., (2011), "Relationships between body mass index, cardiovascular mortality, and risk factors: a report from the SCORE investigators", *Eur J Cardiovasc Prev Rehabil*, Vol.18(5), pp. 731–742.
- [38]. Dutta K., Prakash N., Kaushik S., (2010), "Probabilistic neural network approach to the classification of demonstrative pronouns for indirect anaphora in Hindi", *Expert Systems with Applications*, Vol. 37(8), pp. 5607-5613.
- [39]. Ellestad M.H., Wan M.K.C., (1975), "Predictive implications of stress testing: Follow-up of 2,700 subjects after maximum exercise treadmill testing", *Circulation* 51, Vol. 363(9).
- [40]. Ernst E., Koenig W., (1997), "Fibrinogen and cardiovascular risk", *Vascular Medicine*, Vol. 2, pp. 115-125
- [41]. Esteghamati A., Hafezi- Nejad N., Zandieh A., Sheikhbahaei S., Ebadi M., Nakhjavani M., (2014), "Homocysteine and metabolic syndrome: From clustering to additional utility in prediction of coronary heart disease", *Journal of Cardiology*, Vol. 64, pp. 290-296.
- [42]. European Society of Cardiology [homepage on the Internet] HeartScore; [Accessed May 29, 2016]. Available from: <https://escol.escardio.org/heartscore/download.aspx>.
- [43]. Ezzati M., Hoorn S.V., Rodgers A., (2003), "Estimates of global and regional potential health gains from reducing multiple selected major risk factors", *Lancet*, Vol. 362(9380), pp. 271–280.
- [44]. Federico C., (2010), "Natriuretic Peptide System and Cardiovascular Disease", *Heart Views*, Vol. 11(1), pp. 10–15.

- [45]. Fonseca F.A.H., Izar M., (2016), "High-Sensitivity C-Reactive Protein and Cardiovascular Disease across Countries and Ethnicities", *CLINICS*, Vol. 71(4), pp. 235-242.
- [46]. Frank L., Mueller J.A., (2003), "Medical data analysis using self-organizing data mining technologies," *Systems Analysis Modelling Simulation*, Vol. 43(10), pp. 1399 - 1408.
- [47]. Frawley, Shapiro P., (1996), "Knowledge Discovery in Databases: An Overview", The AAAI/MIT Press.
- [48]. Ganguly P., Alam S.F., (2015), "Role of homocysteine in the development of cardiovascular disease", *Nutrition Journal*, Vol. 14(6), pp. 1-10.
- [49]. Gauba H., Partha P.K., Roy P., Singh P., Dogra D.P., Balasubramanian R., (2017), "Prediction of advertisement preference by fusing EEG response and sentiment analysis", *Neural Networks*, Vol. 92, pp. 77-88.
- [50]. Gautam A., Balasubramanian R., Raghuvanshi S., (2018), "A hybrid approach for the delineation of brain lesion from CT images", *Biocybernetics and Biomedical Engineering*, Vol. 38(3), pp. 504-518.
- [51]. Gerszten R.E., Wang T.J., (2008), "The search for new cardiovascular biomarkers", *Nature*, Vol. 451, pp. 949-952.
- [52]. Goodman S., (2005), "Heart Disease, Genetic Testing and the Development of New Predictive Tools", Associated content, Health & wellness.
- [53]. Gordon T., Castelli W.P., Hjortland M.C., Kannel W.B., Dawber T.R., (1977), "High density lipoprotein as a protective factor against coronary heart disease. The Framingham Study", *Am J Med*, Vol. 62(5), pp. 707-714.
- [54]. Gorecki A., Bala P., Lesyng B., (2006), "Parallelization of the quantum dynamics code for cluster architecture and its applications to the Gross-Pitaevskii equation", *Proceedings from the Fifth European Conference on Computational Chemistry*, Vol. 106(3), pp. 664-669.

- [55]. Gregg D., Goldschmidt-Clermont P.J., (2003), "Platelets and Cardiovascular Disease", *Circulation*. 2003, Vol. 108, pp. e88-e90.
- [56]. Guha S., Pal S.K., Chatterjee N., Ghosh A., PK Deb P.K., (2008), "How Predictive is the Framingham's Risk Prediction Algorithm in Indian Perspective? A Retrospective Case–Control Study from Kolkata", *Indian Heart J*, Vol. 60, pp. 330–332.
- [57]. Gupta I., (2000), "ANNBasedEstimatorforDistillation-InferentialControl", Proceedings of National Conference TIET.COM at Thaper Institute of Engineering and Technology Patiala.
- [58]. Gupta P., Kaur B., (2014), "Accuracy Enhancement of Heart Disease Diagnosis System Using Neural Network and Genetic Algorithm", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 4 (8), pp. 160-166.
- [59]. Gupta R., Misra A., Vikram N.K., Kondal D., Gupta S.S., Agrawal A., Pandey R.M., (2009), "Younger age of escalation of cardiovascular risk factors in Asian Indian subjects", *BMC Cardiovascular Disorders*, Vol. 9(28), pp. 1-12.
- [60]. Haffner S.M., "Management of dyslipidemia in adults with diabetes", *Diabetes Care*, Vol. 21, pp. 160–78.
- [61]. Haffner S.M., Miettinen H., (1997), "Insulin resistance implications for type II diabetes mellitus and coronary heart disease", *Am J Med*, Vol. 103, pp. 152 – 162.
- [62]. Han J., Kamber, M., (2006), "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers.
- [63]. Head J., Fuller J.H., (1990), "International variations in mortality among diabetic patients: the WHO Multinational Study of Vascular Disease in Diabetics", *Diabetologia*, Vol. 33(8), pp. 477–481.
- [64]. <http://who.int/en/> [homepage on the Internet]. WHO Report, March 2013, [updated 2015 January]. Available from: <http://www.who.int/mediacentre/factsheets/fs317/en/index.html>

- [65]. Imes C.C., Lewis F.M., (2014), "Family history of cardiovascular disease (CVD), perceived CVD risk, and health-related behavior: A review of the literature", National Institute of Health, Vol. 29(2), pp. 1-26.
- [66]. Indrayan, (2005), "A Forecasting Vascular Disease cases and associated mortality in India. NCMH Background Papers- Burden of disease in India", Division of Biostatistics and Medical Informatics.
- [67]. Jensen, P.A., Bard, J.F., (2002), "Mathematics of discrete-time Markov chains. Operations Research Models and Methods", John Wiley & Sons, pp. 466-492.
- [68]. Jeswal S.K., Chakraverty S., (2018), "Recent Developments and Applications in Quantum Neural Network: A Review", Archives of Computational Methods in Engineering, pp. 1-15.
- [69]. Jolley M., Stinstra J., Pieper S., Macleod R., Brooks D.H., Cecchin F., Triedman J.K., (2008), "A computer modeling tool for comparing novel ICD electrode orientations in children and adults", Heart Rhythm, Vol. 5(4), pp. 573-4.
- [70]. Jousilahti P., Vartiainen E., Tuomilehto J., Puska P., (1999), "Sex, Age, Cardiovascular Risk Factors, and Coronary Heart Disease", Circulation. 1999, Vol. 99, pp. 1165-1172.
- [71]. Jouven X., Empana J.P., Schwartz P.J., (2005), "Heart rate profile during exercise as a predictor of sudden death", N Engl J Med, Vol. 352, pp. 1951 - 1958.
- [72]. Kadiyala A., Kumar A., (2017), "Vector Time Series-Based Radial Basis Function Neural Network Modeling of Air Quality Inside a Public Transportation Bus Using Available Software", Environmental Progress and Sustainable Energy, Vol. 36(1), pp. 4-10.
- [73]. Kannel W.B., (1996), "Blood pressure as a cardiovascular risk factor: prevention and treatment", JAMA, Vol. 275(20), pp. 1571–1576.
- [74]. Kannel, W.B., Higgins, M., (1990), "Smoking and hypertension as predictors of cardiovascular risk in population studies", J. Hypertens, Vol. 8(3).

- [75]. Karayiannis N.B., Mukherjee A., Glover J.R., (2006), "An evaluation of quantum neural networks in the detection of epileptic seizures in the neonatal electroencephalogram", *Soft Comput.*, Vol. 10(4), pp. 382–396.
- [76]. Karayiannis N.B., Mukherjee A., Glover J.R., (2006), "Detection of pseudosinusoidal epileptic segments in the neonatal EEG by cascading a rule-based algorithm with a neural network", *IEEE Trans Biomed Eng*, Vol.53(4), pp. 633–641.
- [77]. Karthikeyan T., Kanimozhi V.A., (2017), "Deep Learning Approach for Prediction of Heart Disease Using Data mining Classification Algorithm Deep Belief Network", *International Journal of Advanced Research in Science, Engineering and Technology*, Vol.4(1), pp. 3194-3201.
- [78]. Kauhanen J., Kaplan G.A., Goldberg D.E., Salonen J.T., (1997), "Beer binging and mortality: results from the Kuopio ischaemic heart disease risk factor study, a prospective population based study", *Br Med J*, Vol. 315, pp. 846–51.
- [79]. Keil U., Chambless L.E., Doring A., Filipiak B., Stieber J., (1997), "The relation of alcohol intake to coronary heart disease and all cause mortality in a beer-drinking population", *Epidemiology*, Vol. 8, pp. 150–6.
- [80]. Khatibi V., Montazer G.A., (2010), "A fuzzy-evidential hybrid inference engine for coronary heart disease risk assessment", *Expert Systems with Applications*, Vol, 37, pp. 8536-8542.
- [81]. Khatibi V., Montazer G.A., (2010), "A fuzzy-evidential hybrid inference engine for coronary heart disease risk assessment", *Expert Syst Appl.*, Vol. 37(12), pp. 8536–8542.
- [82]. Killip T III, Payne M.A., (1960), "High Serum Transaminase Activity in Heart Disease Circulatory Failure and Hepatic Necrosis", *Circulation* 1960, Vol. 21, pp. 646-60.
- [83]. Klatsky A.L. (1996), "Alcohol, coronary heart disease, and hypertension", *Ann Rev Med*, Vol. 47, pp. 149–60.

- [84]. Krishnaiah V., Narsimha G., Chandra N.S., (2014), "Heart Disease Prediction System Using Data Mining Technique by Fuzzy K-NN Approach", *Emerging ICT for Bridging the Future - Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1. Advances in Intelligent Systems and Computing*, vol.337.
- [85]. Krolewski A.S., Kosinski E.J., Warram J.H., (1987), "Magnitude and determinants of coronary artery disease in juvenileonset, insulin-dependent diabetes mellitus", *Am J Cardiol*, Vol. 59, pp. 750–5.
- [86]. Kukar M., Kononenko I., Groselj C., Kralj K., Fettich J., (1999), "Analysing and improving the diagnosis of ischaemic heart disease with machine learning", *Artif Intell Med.*, Vol. 16(1), pp. 25–50.
- [87]. Kuller L.H., Evans R.W., (1998), "Homocysteine, Vitamins and Cardiovascular Disease", *Circulation*. 1998, Vol. 98, pp. 196-199.
- [88]. Kumar N., Goel N., Yadav T.C., Pruthi V., (2017), "Quantum chemical, ADMET and molecular docking studies of ferulic acid amide derivatives with a novel anticancer drug target", *Medicinal Chemistry Research*, Vol. 26(8), pp. 1822-1834.
- [89]. Kumar N., Pruthi V., Goel N., (2015), "Structural, thermal and quantum chemical studies of p-coumaric and caffeic acids", *Journal of Molecular Structure*, Vol. 1085, pp. 242-248.
- [90]. Kunutsor S.K., Apekey T.A., Khan H., (2014), "Liver enzymes and risk of cardiovascular disease in the general population: a meta-analysis of prospective cohort studies", *Atherosclerosis*, Vol. 236(1), pp. 7-17.
- [91]. Laakso M., Lehto S., (1997), "Epidemiology of macrovascular disease in diabetes", *Diabetes Rev*, Vol. 5, pp. 294–315.
- [92]. Latha P., Subramanian R., (2008), "Intelligent Heart Disease Prediction System Using CANFIS and Genetic Algorithm", *International Journal of Biological, Biomedical and Medical Sciences*, Vol. 3(3), pp. 157- 160.

- [93]. Lauer M., Okin P., Martin G., (1998), "Impaired heart rate response to graded exercise: prognostic implications of chronotropic incompetence in the Framingham Heart Study", *Circulation*, Vol. 93, pp. 1520 -6.
- [94]. Lauraitis A., Maskeliūnas R., Damaševičius R., (2018), "ANN and Fuzzy Logic Based Model to Evaluate Huntington Disease Symptoms", *Journal of Healthcare Engineering*, Vol. 2018, pp. 1-10.
- [95]. Law M.R., Morris J.K., Wald N.J., (1997), "Environmental tobacco smoke exposure and ischaemic heart disease: an evaluation of the evidence", *Br Med J*, Vol. 315, pp. 973–82.
- [96]. Legato M.J., Padus E., Slaughter E., (1997), "Women's perceptions of their general health with special reference to their risk of coronary artery disease: results of a national telephone survey", *J Womens Health*, Vol. 6(2), pp.189–198.
- [97]. Lewsey J.D., Lawson K.D., Ford I., Fox K.A.A., Ritchie L.D., Tunstall-Pedoe H., Watt G.C.M., Woodward M., Kent S., Neilson M., Briggs A.H., (2014), "A cardiovascular disease policy model that predicts life expectancy taking into account socioeconomic deprivation", *Heart*, Vol. 0, pp. 1–8.
- [98]. Liam T.H., He Y., (1997), "Passive smoking and coronary heart disease: a brief review", *Clin Exp Pharmacol Physiol*, Vol. 24, pp. 993–6.
- [99]. Lipinski M.J., Vectrovec G.W., Froelicher V.F., (2004), "Importance of the first two minutes after exercise treadmill testing in predicting mortality and the presence of coronary artery disease in men", *Am J Cardiol*, Vol. 93, pp. 445 -9.
- [100]. Liu J., Hong Y., D'Agostino R.B. Sr., Wu Z., Wang W., Sun J., Wilson P.W., Kannel W.B., Zhao D., (2004), "Predictive value for the Chinese population of the Framingham CHD risk assessment tool compared with the Chinese Multi-Provincial Cohort Study". *JAMA*, Vol. 291, pp. 2591-9.
- [101]. Lloyd-Jones D.M., Larson M.G., Beiser A., Levy D., (1999), "Lifetime risk of developing coronary heart disease" *Lancet*, Vol. 353(9147), pp. 89–92.

- [102]. Long N.C., Meesad P., Unger H., (2015), "A highly accurate firefly based algorithm for heart disease prediction", *Expert Systems with Applications*, Vol. 42, pp. 8221-8231.
- [103]. Lown B., Verrier R.L., (1976), "Neural activity and ventricular fibrillation", *N Engl J Med*, Vol. 294, pp. 1165 – 70.
- [104]. MacKenzie T.D., Bartechi C.E., Schrier R.W., (1994), "The human costs of tobacco use (second of two parts)", *N Engl J Med*, Vol. 330, pp. 975–80.
- [105]. MacKenzie T.D., Bartechi C.E., Schrier R.W., (1994), "The human costs of tobacco use", *N Engl J Med*, Vol. 330, pp. 975–980.
- [106]. Mahadavan G., Nguyen T.H., Horowitz J.D., (2014), "Brain natriuretic peptide: a biomarker for all cardiac disease?", *Curr Opin Cardiol*, Vol. 29(2), pp.160-6.
- [107]. Mainous A.G., Koopman R.J., Diaz V.A., Everett C.J., Wilson P.W.F., Tilley B.C., (2007), "A Coronary Heart Disease Risk Score Based on Patient Reported Information", *Am J Cardiology*, Vol. 99(9), pp. 1236–1241.
- [108]. Malpas S.C., (2002), "Neural influences on cardiovascular variability: possibilities and pitfalls", *Am J Physiol Heart Circ Physiol*, Vol. 282, pp. 116 – 120.
- [109]. Manson J.E., Tosteson H., Ridker P.M., (1992), "The primary prevention of myocardial infarction", *N Engl J Med*, Vol. 326, pp. 1406–16.
- [110]. Manson J.E., Tosteson H., Ridker P.M., (1992), "The primary prevention of myocardial infarction", *N Engl J Med*, Vol. 326 (21), pp. 1406–1416.
- [111]. Mansourvar M., Asaemi A., Raj R.G., Kareem S.A., (2017), "Antony CD, Idris N, Baba MS. A Fuzzy Inference System for Skeletal Age Assessment in Living Individual", *International Journal of Fuzzy Systems*, Vol. 19(3), pp. 838-848.
- [112]. Martin C., Mark V., Joyce F., (2004), "Description and validation of a Markov model of survival for individuals free of cardiovascular disease that uses Framingham risk factors", *BMC Medical Informatics and Decision Making*, Vol. 4(6), pp. 1 - 15.

- [113]. Massimo Volpe M., Rubattu S., and Burnett J. Jr3, (2014), "Natriuretic peptides in cardiovascular diseases: current use and perspectives", *European Heart Journal*, Vol. 35, pp. 419–425.
- [114]. Mastoi Q.U.A., Wah T.Y., Raj R.G., Iqbal U., (2018), "Automated Diagnosis of Coronary Artery Disease: A Review and Workflow", *Cardiology Research and Practice*, Vol. 2018, pp. 1-10.
- [115]. Mayani M., Mohanty B., Singh R.P., (2007), "A multikinetik model approach to predict gluconic acid production in an airlift bioreactor", *Biotechnology journal*, Vol. 2(5), pp. 631-639.
- [116]. McMahon R., Peto R., Cuttler J., (1990), "Blood pressure, stroke and coronary heart disease. Part 1, Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias". *Lancet*, Vol. 335(8692), pp. 765–774.
- [117]. Meade T.W., (1997), "Fibrinogen and cardiovascular disease", *J Clin Pathol*, Vol. 50, pp. 13-15.
- [118]. Mehta R., Lauer M., Pashkow F., (1998), "Association of chronotropic incompetence with myocardial ischemia and prognosis", *J Am Coll Cardiol*, Vol. 32, pp. 1280 -6.
- [119]. Mendis S., Puska P., Norrving B., (2011), "Global Atlas on cardiovascular disease prevention and control, Health Organization in collaboration with the World Heart Federation and the World Stroke Organization." Available from: http://www.world-heart-federation.org/fileadmin/user_upload/images/CVD_Health/Global_CVD_Atlas.pdf
- [120]. Merry A.H.H., Boer J.M.A., Schouten L.J., Feskens E.J.M., Verschuren W.M.M., Gorgels A.P.M. Van den Brandt P.A., (2011), "Smoking, alcohol consumption, physical activity, and family history and the risks of acute myocardial infarction and unstable angina pectoris: a prospective cohort study", *BMC Cardiovascular Disorders*, Vol. 11(13), pp. 1-14.

- [121]. Miller M., Stone N.J., Ballantyne C., Bittner V., Criqui M.H., Ginsberg H.N., Goldberg A.C., Howard W.J., Jacobson M.S., Kris-Etherton P.M., Lennie T.A., Levi M., Mazzone T., Pennathur S., (2011), "Triglycerides and Cardiovascular Disease: A Scientific Statement From the American Heart Association", *Circulation*. 2011, Vol. 123, pp. 2292-2333.
- [122]. Minino M.A., Kung C.H., Kochanek K.D., Xu J and Murphy S.L., (2011), "Deaths: Final Data for 2009", Division of Vital Statistics (Atlanta, GA: Centers for Disease Control), Vol. 60 (3).
- [123]. Montgomery A.A., Fahey T., Ben Shlomo Y., Harding J., (2003), "The influence of absolute cardiovascular risk, patient utilities, and costs on the decision to treat hypertension: a Markov decision analysis". *J Hypertens*, Vol. 21, pp. 1753-1759.
- [124]. Montgomery A.A., Fahey T., MacKintosh C., (2000), "Estimation of cardiovascular risk in hypertensive patients in primary care" *Br J Gen Pract.*, Vol. 50, pp. 127–8.
- [125]. Moser M., Basile J., Sica D., (2003), "Electrolytes in Cardiovascular Disease and Hypertension", *Journal of clinical hypertension*, Vol. 5(6), pp. 402-407.
- [126]. Mukamal K.J., (2006), "The Effects Of Smoking And Drinking On Cardiovascular Disease And Risk Factors", *Alcohol Research & Health*, Vol. 29(3), pp. 199-202.
- [127]. Narayan R., Chakraverty S., Singh V.P., (2013), "Machine translation using quantum neural network for simple sentences", *International Journal of Information and Computation Technology*, Vol.3(7), pp. 683–690.
- [128]. Narayan R., Chakraverty S., Singh V.P., (2014a), "Neural Network based Parts of Speech Tagger for Hindi", *Proceedings of 3rd International conference, Advances and control and Optimisation of dynamical systems*, Kanpur, India, Kanpur, India: proceedings of IFAC-Elsevier.
- [129]. Narayan R., Singh V.P., Chakraverty S., (2014b), "Quantum neural network based machine translator for Hindi to English", *ScientificWorldJournal*, Vol. 2014(485737).

- [130]. Narayan R., Singh V.P., Chakraverty S., (2014c), "Quantum Neural Network based Parts of Speech Tagger for Hindi", *International Journal of Advancements in Technology*, Vol. 5(2), pp. 137–152.
- [131]. Narayan R., Chakraverty S., Singh V.P., (2016), "Quantum neural network based machine translator for English to Hindi", *Applied Soft Computing*, Vol. 38, pp. 1060–1075.
- [132]. Nathan D.M., Meigs J., Singer D.E., (1997), "The epidemiology of cardiovascular disease in type 2 diabetes mellitus: how sweet it is . . . or is it?", *Lancet*, Vol. 350(I), pp. 4–9.
- [133]. Norhayati M.N., Sharifah Amirah S.A.H., Husniati Y.L., (2013), "Cardiovascular risk: associated factors, assessment and agreement between WHO/ISH risk prediction chart and Framingham Scoring System among Primary Care Patients in Kelantan, Malaysia", *International Journal of Collaborative Research on Internal Medicine & Public Health*, Vol. 5(12), pp. 652-663.
- [134]. Norris, J.R., (1997), "Markov Chains", Cambridge University Press, Cambridge.
- [135]. Palaniappan S., Awang R., (2008), "Intelligent Heart Disease Prediction System Using Data Mining Techniques", *IJCSNS International Journal of Computer Science and Network Security*, Vol. 8 (8), pp. 108 – 115.
- [136]. Panigrahi S.K., Chakraverty S., Mishra B.K., (2009), "Vibration based damage detection in a uniform strength beam using genetic algorithm", *Meccanica*, Vol. 44(697),
- [137]. Patil B.S., Kumaraswamy Y.S., (2009), "Extraction of Significant Patterns from Heart DiseaseWarehouses for Heart Attack Prediction", *International Journal of Computer Science and Network Security*, Vol. 9 (2), pp. 228-235.
- [138]. Pavlopoulos S.A., Delopoulos A.N., (1999), "Designing and implementing the transition to a fully digital hospital", *IEEE Transactions on Information Technology in Biomedicine*, Vol. 3(1), pp. 6-19.

- [139]. Perini R., Veicsteinas A., (2003), "Heart rate variability and autonomic activity at rest and during exercise in various physiological conditions", *Eur J Appl Physiol*, Vol. 90, pp. 317 – 25.
- [140]. Pfützner A., Forst T., (2006), "High-Sensitivity C-Reactive Protein as Cardiovascular Risk Marker in Patients with Diabetes Mellitus", *Diabetes Technology & Therapeutics*, Vol. 8(1), pp. 28-36
- [141]. Pilote L., Hlatky M.A., (1995), "Attitudes of women toward hormone therapy and prevention of heart disease", *Am Heart J*, Vol. 129(6), pp.1237–1238.
- [142]. Pyorala K., Laakso M., Uusitupa M., (1987), "Diabetes and atherosclerosis: an epidemiologic view", *Diabetes/Metab Rev*, Vol. 3, pp. 463–524.
- [143]. Ramazan D., Sehmus O., Davut A., Fatma C., Orhan Y., (2008), "Predictive Factors of Cardiovascular Disease in Patients on Maintenance Hemodialysis", *Dialysis & Transplantation*, pp. 1-6.
- [144]. Rehm J.T., Bondy S.J., Sempos C.T., Vuong C.V., (1997), "Alcohol consumption and coronary heart disease morbidity and mortality", *Am J Epidemiol*, Vol. 146, pp.495–501.
- [145]. Ren J., Grundy S.M., Liu J., Wang W., Wang M., Sun J., Liu J., Li Y., Wu Z., Zhao D., (2010), "Long-term coronary heart disease risk associated with very-low-density lipoprotein cholesterol in Chinese: The results of a 15-Year Chinese Multi-Provincial Cohort Study (CMCS)", *Atherosclerosis*, Vol. 211(1), pp. 327-32.
- [146]. Rimm E.B., Klatsky A., Grobbee D., Stampfer M.J., (1996), "Review of moderate alcohol consumption and reduced risk of coronary heart disease: is the effect due to beer, wine, or spirits?", *Br Med J*, Vol. 312, pp. 731–6.
- [147]. Robin H., Eberhardt J.S., Gaertner R., Kam J., Armstrong M., (2006), "Interpreting diagnostic assays by means of statistical modeling", *IVD Technology*.

- [148]. Roerecke M., Rehm J., (2014), "Chronic heavy drinking and ischaemic heart disease: a systematic review and meta-analysis", *J. Open Heart*, Vol. 1:e00135.
- [149]. Rosengren A., Wilhelmsen L., Wedel H., (1992), "Coronary heart disease, cancer and mortality in male middle-aged light smokers", *J Intern Med.*, Vol. 231(4), pp. 357–362.
- [150]. Sahoo D.M., Chakraverty S., (2017), "Functional link neural network approach to solve structural system identification problems", *Neural Computing and Applications*, pp. 1-12.
- [151]. Sangster J., Furber S., Allman-Farinelli M., Haas M., Phongsavan P., Mark A., Bauman A., (2010), "A population-based lifestyle intervention to promote healthy weight and physical activity in people with cardiac disease: The PANACHE (Physical Activity, Nutrition And Cardiac Health) study protocol", *BMC Cardiovascular Disorders*, Vol. 10 (17), pp. 1-8.
- [152]. Schwartz P.J., La Rovere M.T., Vanoli E., "Autonomic nervous system and sudden cardiac death: experimental basis and clinical observations for post–myocardial infarction risk stratification", *Circulation*, Vol. 85, pp. 177-91.
- [153]. Schwartz P.J., Vanoli E., Stramba-Badiale M., (1988), "Autonomic mechanisms and sudden death: new insights from analysis of baroreceptor reflexes in conscious dogs with and without a myocardial infarction", *Circulation*, Vol. 78, pp. 969- 79.
- [154]. Shaper A.G., Wannamethee G., Walker M. (1994), "Alcohol and coronary heart disease: a perspective from the British Regional Heart Study", *Int J Epidemiol*, Vol. 23, pp. 482–94.
- [155]. Shen J., Zhang J., Wen J., Ming Q., Zhang J., Xu Y., (2015), "Correlation of serum alanine aminotransferase and aspartate aminotransferase with coronary heart disease", *Int J Clin Exp Med*, Vol. 8(3), pp. 4399-4404.
- [156]. Shrimali V., Anand R.S., Kumar V., Srivastav R.K., (2008), "Medical feature based qualitative evaluation of denoising techniques for ultrasound liver images", *International Journal of Signal and Imaging Systems Engineering*, Vol. 1(2), pp. 135-144.

- [157]. Shrimali V., Anand R.S., Kumar V., Srivastav R.K., (2009), "Medical feature based evaluation of structuring elements for morphological enhancement of ultrasonic images", *Journal of Medical Engineering & Technology*, Vol. 33(2), pp. 158-169.
- [158]. Skretteberg P.T., Grundvolda I., Kjeldsen S.E., Erikssen J.E., Sandvik L., Liestøl K., Erikssen G., Pedersen T.F., Bodegard J., (2012), "HDL-cholesterol and prediction of coronary heart disease: Modified by physical fitness? A 28-year follow-up of apparently healthy men", *Atherosclerosis*, Vol. 220, pp. 250- 256.
- [159]. Sonnenberg F.A., Beck J.R., (1993), "Markov models in medical decision making: a practical guide", *Med Decis Making*, Vol. 13, pp. 322-338.
- [160]. Sood P.S., Nwabeze S.N., Mbarika V.W.A., Prakash N., Chatterjee S., Ray P., Mishra S., (2008), "Electronic Medical Records: A Review Comparing the Challenges in Developed and Developing Countries", *Hawaii International Conference on System Sciences, Proceedings of the 41st Annual*, pp. 1530-1605.
- [161]. Soofi M.A., Youssef M.A., (2015), "Prediction of 10-year risk of hard coronary events among Saudi adults based on prevalence of heart disease risk factors", *J. Saudi Heart Association*, Vol. 127, pp. 152-159.
- [162]. Stec J.J., Silbershatz H., Tofler G.H., Matheney T.H., Sutherland P., Lipinska I., Massaro J.M., Wilson P., Muller J.E., D'Agostino R. B. Sr, (2000), "Association of Fibrinogen With Cardiovascular Risk Factors and Cardiovascular Disease in the Framingham Offspring Population", *Circulation*, Vol. 102, pp. 1634-1638.
- [163]. Steinberg D., Parthasarathy S., Carew S., Khoo J.C., Witztum J.L., (1989), "Beyond cholesterol. Modifications of low-density lipoprotein that increase its atherogenicity", *NEngl J Med.*, Vol. 320, pp. 915-924.
- [164]. Stovring H., Harmsen C.G., Wisloff T., Jarbol D.E., Nexoe J., Nielsen J.B. Kristiansen I.S., (2013), "Competing Risk Approach for the European Heart SCORE Model based on Cause-specific and All-cause Mortality", *Euro J of Prev Cardiol.*, Vol.20(5), pp. 827-836.

- [165]. Thaulow E., Erikssen J., Sandvik L., Stormorken H., Cohn P.F., (1991), "Blood platelet count and function are related to total and cardiovascular death in apparently healthy men", *Circulation*, Vol. 84(2), pp. 613-7.
- [166]. Tian S., (1997), "Cystatin C Measurement and Its Practical Use in Patients with Various Renal Diseases," *Clinical Nephrology*, Vol. 48(2), pp. 104–108.
- [167]. Tsevat J., Weinstein M.C., Williams L.W., Tosteson A.N., Goldman L., (1991), "Expected gains in life expectancy from various coronary heart disease risk factor modifications", *Circulation*, Vol. 83, pp. 1194-1201.
- [168]. Turner R.C., Millns H., Neil H.A., (1998), "Risk factors for coronary artery disease in non-insulin dependent diabetes mellitus: United Kingdom Prospective Diabetes Study (UKPDS: 23)", *BMJ*, Vol. 316(7134), pp. 823–828.
- [169]. Tverdal A., Thelle D., Stensvold I., Leren P., Bjartveit K., (1993), "Mortality in relation to smoking history: 13 years' follow-up of 68,000 Norwegian men and women 35–49 years", *J Clin Epidemiol*, Vol.46(5), pp.475–487.
- [170]. Valerio L., Peters R.J., Zwinderman A.H., Pinto-Sietsma S.J., (2016), "Association of Family History With Cardiovascular Disease in Hypertensive Individuals in a Multiethnic Population", *J Am Heart Assoc*, Vol. 5:e004260.
- [171]. Vinholt P.J., Hvas A.M., Frederiksen H., Bathum L., Jørgensen M.K., Nybo M., (2016), "Platelet count is associated with cardiovascular disease, cancer and mortality: A population-based cohort study", *Thromb Res*, Vol. 148, pp. 136-142.
- [172]. Vuolteenaho O., Ala-Kopsala M., Ruskoaho H., (2005), "BNP as a biomarker in heart disease", *Adv Clin Chem*, Vol. 40, pp. 1-36.
- [173]. Wagh R., Paygude S.S., (2016), "CDSS for Heart Disease Prediction Using Risk Factors", *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 4(6), pp. 12082-12089.

- [174]. Walker W.G., Neaton J.D., Cutler J.A., Neuwirth, Cohen J.D., (1992), "Renal function change in hypertensive members of the Multiple Risk Factor Intervention Trial. Racial and treatment effects", The MRFIT Research Group. JAMA, Vol. 268(21), pp. 3085–3091.
- [175]. Wannamethee S.G., Shaper A.G., Lennon L., Morris R.W., (2005), "Metabolic Syndrome vs Framingham Risk Score for Prediction of Coronary Heart Disease, Stroke, and Type 2 Diabetes Mellitus", Arch Intern Med, Vol. 165, pp. 2644-2650.
- [176]. Watanabe M., Okamura T., Kokubo Y., Higashiyama A., Okayama A., (2009), "Elevated serum creatine kinase predicts first-ever myocardial infarction: a 12-year population-based cohort study in Japan, the Suita study", International Journal of Epidemiology, Vol. 38, pp.1571–1579.
- [177]. Watson J.E., (1987), "Fluid and electrolyte disorders in cardiovascular patients", Nurs Clin North Am., Vol. 22(4), pp.797-803.
- [178]. Weinstein M.C., Coxson P.G., Williams L.W., Pass T.M., Stason W.B., Goldman L., (1987), "Forecasting coronary heart disease incidence, mortality, and cost: the Coronary Heart Disease Policy Model", Am J Public Health, Vol.77, pp. 1417-1426.
- [179]. Weinstein M.C., Coxson P.G., Williams L.W., Pass T.M., Stason W.B., Goldman L., (1987), "Forecasting Coronary Heart Disease Incidence, Mortality and Cost: the Coronary Heart Disease Policy Model", Am J Public Health, Vol. 77, pp.1417–1426.
- [180]. WHO Report, October 2008.
- [181]. Willett W.C., Green A., Stampfer M.J., (1987), "Relative and absolute excess risks of coronary heart disease among women who smoke cigarettes", N Engl J Med, Vol. 317(21), pp. 1303–1309.
- [182]. Wilsgaard T., Arnesen E., (2007), "Body Mass Index and Coronary Heart Disease Risk Score: The Tromsø Study, 1979 to 2001", Elsevier, Vol. 17(2), pp. 100-105.

- [183]. Wilson Peter W.F., Bozeman S.R., Burton T.M., Hoaglin D.C., Ben-Joseph R., Pashos C.L., (2008), "Prediction of First Events of Coronary Heart Disease and Stroke With Consideration of Adiposity", *Circulation*, Vol. 118, pp. 124-130.
- [184]. Wu R., Peters W., Morgan M.W., (2002), "The next generation clinical decision support: linking evidence to best practice", *J Healthc Inf Manag*, Vol. 16(4), pp.50–55.
- [185]. Yang T., Lin H.C., Chen M.L., (2006), "Meta modeling approach in solving the machine parameters optimization problem using neural network and genetic algorithms: a case study", *Rob. Comp-Int. Manuf*, Vol. 22(4), pp. 322– 331.
- [186]. Yu S., Yarnell J.W.G., Sweetnam P., Bolton C.H., (2003), "High density lipoprotein subfractions and the risk of coronary heart disease: 9-years follow-up in the Caerphilly Study", *Atherosclerosis*, Vol. 166(2), pp.331-338.
- [187]. Yudkin J.S., Blauth C., Drury P. (1996), "Prevention and management of cardiovascular disease in patients with diabetes mellitus: an evidence base", *Diabet Med*, Vol. 13, pp. 101–21.
- [188]. Zhou J., Gan Q., Krzyzak A., (1999), "Recognition of handwritten numerals by quantum neural network with fuzzy features", *Intl. J. on document analysis and recognition*, Vol. 2, pp. 30-36.

LIST OF PUBLICATIONS

Following are the papers that have been published/ communicated from the present research:

Published

[1]. Narain R, Saxena S and Goyal AK (2016) Cardiovascular Risk Prediction: A Comparative study of Framingham and Quantum Neural Network based approach. Patient Preference and Adherence. 10, 1259-1270. **SCIE impact factor 1.718 (5 year impact 1.831), first volume published in 2007**

[2]. Narain R, Saxena S and Goyal AK (2014) Cardiovascular Disease Prediction based on Physical factors using Quantum Neural Network. IJIP. 8(3), 83-94 - **Index: Google Scholar, arXiv, Indian Citation Index, getCITED), first volume published in 2007**

[3]. Narain R, Saxena S and Goyal AK (2017) A Novel Heart Disease Prediction System based on Quantum Neural Network using Clinical Parameters. Annual Research & Review in Biology. 14(2): 1-10. **Index: SCOPUS (from Elsevier), PubMed, Index Medicus, ProQuest, Google Scholar, arXiv, CrossRef, getCITED) etc., first volume published in 2011.**

Under Communication

[1]. Narain R, Saxena S and Goyal AK. Intelligent heart disease prognostic system based on quantum neural network. BioMed Research International Patient Preference and Adherence. — Under Communication. **SCIE impact factor 1.718 (5 year impact 1.831), first volume published in 2007**

Cardiovascular risk prediction: a comparative study of Framingham and quantum neural network based approach

This article was published in the following Dove Press journal:

Patient Preference and Adherence

19 July 2016

[Number of times this article has been viewed](#)

Renu Narain¹
Sanjai Saxena¹
Achal Kumar Goyal²

¹Department of Biotechnology,
Thapar University, Punjab, India;

²University Computer Center,
Gurukul Kangri University, Haridwar,
Uttarakhand, India

Purpose: Currently cardiovascular diseases (CVDs) are the main cause of death worldwide. Disease risk estimates can be used as prognostic information and support for treating CVDs. The commonly used Framingham risk score (FRS) for CVD prediction is outdated for the modern population, so FRS may not be accurate enough. In this paper, a novel CVD prediction system based on machine learning is proposed.

Methods: This study has been conducted with the data of 689 patients showing symptoms of CVD. Furthermore, the dataset of 5,209 CVD patients of the famous Framingham study has been used for validation purposes. Each patient's parameters have been analyzed by physicians in order to make a diagnosis. The proposed system uses the quantum neural network for machine learning. This system learns and recognizes the pattern of CVD. The proposed system has been experimentally evaluated and compared with FRS.

Results: During testing, patients' data in combination with the doctors' diagnosis (predictions) are used for evaluation and validation. The proposed system achieved 98.57% accuracy in predicting the CVD risk. The CVD risk predictions by the proposed system, using the dataset of the Framingham study, confirmed the potential risk of death, deaths which actually occurred and had been recorded as due to myocardial infarction and coronary heart disease in the dataset of the Framingham study. The accuracy of the proposed system is significantly higher than FRS and other existing approaches.

Conclusion: The proposed system will serve as an excellent tool for a medical practitioner in predicting the risk of CVD. This system will be serving as an aid to medical practitioners for planning better medication and treatment strategies. An early diagnosis may be effectively made by using this system. An overall accuracy of 98.57% has been achieved in predicting the risk level. The accuracy is considerably higher compared to the other existing approaches. Thus, this system must be used instead of the well-known FRS.

Keywords: myocardial infarction, atherosclerosis, Framingham risk score, cardiovascular disease

Introduction

As per the World Health Organization (WHO), the main cause of mortality worldwide is due to cardiovascular diseases (CVDs). According to the WHO report, in the year 2008 approximately 17.3 million people died due to CVDs, which is 30% of all global deaths and by 2030, the total number of deaths will increase up to 23.6 million, mainly from heart disease and stroke.¹

At present, the deaths occur due to lack of early medical diagnosis of CVDs, thereby posing a big challenge to health care organizations. Correct and timely diagnosis of

Correspondence: Sanjai Saxena
Department of Biotechnology, Thapar
University, Bhadson Road, Patiala,
Punjab 147004, India
Email sanjaibitech@yahoo.com

patients is required for effective treatment and for quality service. The decisions are currently based on the experience and understanding of the doctor, instead of intelligent tools. To overcome this problem, the approach of combining clinical decision making supported by computer based patient records has been proposed by Wu et al.² The main aim of the present approach was to minimize the possibilities of error in doctors' inference and decisions, which may also be safer for patients.²

These days, some sort of information systems are being used by many hospitals to manage patients' information, but these systems are not used for clinical decision making. Many researchers have proposed CVD prediction systems based on different approaches. Khatibi and Montazer proposed a fuzzy-evidential based CVD prediction system in the year 2010, with which they have achieved accuracy of 91.58%.³ Akay proposed the neural network based coronary artery disease prediction system in 1992. They used 100 patient records for testing and achieved accuracy of 84% by using the data of clinical tests.⁴ The Bayesian classification and neural network based ischemic heart disease prediction system have been proposed by Kukar et al in 1999. They have used electrocardiogram (ECG) data as input for prediction.⁵ The probability theory (logistic regression) based angiographic coronary disease system has been proposed by Detrano et al in 1989. They have collected the data of 425 patients and achieved accuracy of 77%. The clinical and noninvasive test data have been used in the system for prediction.⁶

To overcome the above gaps, an intelligent heart disease prediction system is proposed here, which uses the heart disease databases with quantum neural network (QNN) to make intelligent clinical decisions which cannot be done in traditional decision support systems.⁷ In this paper, the proposed system based on QNN has been compared with the Framingham risk score (FRS) using the same parameters as used in FRS.

Related work – FRS

The parameters used for FRS are sex, age, total cholesterol (TC), high density lipoprotein (HDL), systolic blood pressure, diastolic blood pressure, diabetes mellitus, body mass index (BMI), and smoking.⁸ The data of a group of patients were collected in the 1960s and 1970s and these data are used in the Framingham equations. Since then the cases of coronary heart disease (CHD) mortality have been decreasing.⁹ The lifestyle and other risk factors have been changed, which affected the rate of CHD. The FRS is outdated for the modern population. So the FRS is not accurate enough.

Common risk factors used in FRS and proposed system

The details of common risk factors that are used in FRS and in the proposed system are as follows.

Sex

According to research, males are at greater risk of CVD than pre-menopausal females. Females after menopause are in the same risk category of CVD as males. On the other hand, the risk of stroke is equal for males and females.^{10–12}

Age

The risk of CVD increases with the age. According to studies, after the age of 65 years, 80% of deaths occur due to heart disease. With the increase in age, our heart does not work properly.¹⁰

TC

Elevated level of total blood cholesterol is one of the major causes of CVD.¹³ As blood cholesterol increases, the risk of CVD increases. Table 1 shows the various risk levels for TC.

HDL

Abnormality of HDL may increase the risk of atherosclerosis, risk levels are shown in Table 2. According to research, the HDL level is a strong, independent inverse predictor of CVD.^{14–16}

Blood pressure

It is the most important risk factor of heart diseases. Globally, various epidemiological studies have identified the direct relation between blood pressure elevation and rate of coronary artery disease and stroke.^{17–19} It has also been identified that, the combined effect of multiple risk factors causes an increased risk.²⁰ The normal blood pressure range according to age is shown in Table 3.

Diabetes mellitus

Diabetes is also one of the most important risk factors of heart diseases. It increases the risk of CVD. According to the Rancho Bernardo Study, diabetes increases the risk of

Table 1 Risk levels for total cholesterol

Total cholesterol	Risk
Less than 200 mg/dL (5.17 mmol/L)	Desirable
200–239 mg/dL (5.17–6.18 mmol/L)	Borderline high risk
240 and over (6.2 mmol/L)	High risk

Table 2 Risk levels for high density lipoprotein

High density lipoprotein	Risk
Less than 40 mg/dL (men) (1.03 mmol/L)	Increased risk of heart disease
Less than 50 mg/dL (women) (1.29 mmol/L)	Increased risk of heart disease
Greater than 60 mg/dL (1.55 mmol/L)	Some protection against heart disease

CVD up to 1.9 times in males and 3.3 times in females.²¹ This study had been performed for 14 years on approximately 2,400 persons. The relation between diabetes and CVD is not uniform in all populations.^{22,23}

BMI

The BMI index is a statistical measurement. It shows the ratio between weight and height,²⁴ as shown in Table 4.

Smoking

It is one of the most important factors which increase the risk of myocardial infarction (MI). Different research studies show that smoking/tobacco use increases the risk of MI.²⁵ Proving the harmful effects of light smoking with adequate accuracy requires large cohort studies and only few studies have attempted this.^{26,27} According to studies, smoking is the main cause of 50% of deaths unrelated to CVD, which are avoidable and only one half of deaths happen due to CVD.^{28–31}

Proposed CVD prediction system

The proposed system provides a comprehensive predictive tool for predicting the CVD risk. This system predicts CVD at an early stage. The clinicians, doctors, pathologists as well as general public may use this tool very easily. It predicts the chances of CVD using the physical tests for CVD. Using

Table 3 Blood pressure range according to age

Age group	Lower limit ^a	Average limit ^a	Upper limit ^a
15–19	105/73	117/77	120/81
20–24	108/75	120/79	132/83
25–29	109/76	121/80	133/84
30–34	110/77	122/81	134/85
35–39	111/78	123/82	135/86
40–44	112/79	125/83	137/87
45–49	115/80	127/84	139/88
50–54	116/81	129/85	142/89
55–59	118/82	131/86	144/90
60–64	121/83	134/87	147/91
65–69	122/84	136/88	149/92
≥70	123/85	138/89	151/93

Note: ^aSystolic/diastolic blood pressure range (mmHg).

Table 4 BMI range

Category	BMI range, kg/m ²
Severely underweight	<16.5
Underweight	16.5 to 18.5
Normal	18.5 to 25
Overweight	25 to 30
Obese	>30

Abbreviation: BMI, body mass index.

this tool, the clinicians may fulfill their promises and commitments for ensuring the sustained fight against this deadly disease. The activity diagram showing the structure of the proposed model is shown in Figure 1. The CHD forecast based on a study by Indrayan³² showed the number of cases, both male and female, of CHD in India (Table 5). Here, in the proposed system for prediction, only the factors considered in the FRS have been considered here.

Materials and methods

The current study has been conducted during 2009 to 2015 on a cohort of CVD patients aged 20 years and above. During data collection, the personal details of patients were not collected. To build a knowledge based CVD prediction system, the QNN has been used in this approach for better adaptability and classification. This study has been conducted after the approval of Thapar University Research Board (TURB), who deemed that the patients' consent was not required for using retrospective/existing cohort without any identifiers linking to individuals.

Data description – patients' data

This study has been conducted with the data of 689 patients, showing the symptoms of CVD. The parameters of every individual patient have been diagnosed and analyzed by the three individual physicians. In Table 6, the various input parameters are shown, which have been used for prediction of heart disease and used in the FRS. The FRS is designed for adults, ie, aged 20 years and above. The training and evaluation was performed with the help of a cardiologist/physician. The data consist of patients' records with doctors' predictions/diagnosis. The whole dataset is divided into training, validation, and testing. Furthermore, for validation purpose, the dataset of the famous Framingham study conducted on 5,209 CVD patients, was taken from the University of Washington, Seattle, WA, USA.³³

QNN used for CVD prediction system

QNN based on multi-level transfer function was introduced by Karayiannis et al.^{34–36} In comparison with artificial neural

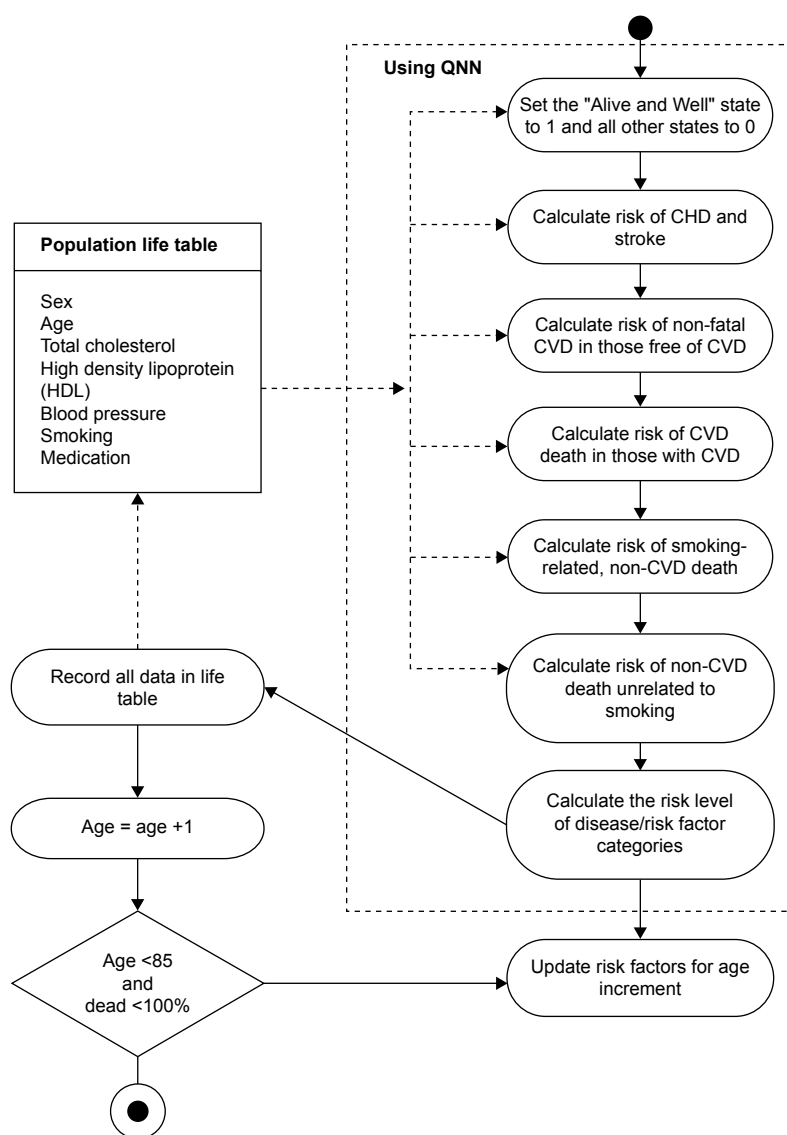


Figure 1 Activity diagram of proposed system.

Abbreviations: QNN, quantum neural network; CHD, coronary heart disease; CVD, cardiovascular disease.

network, the hidden units of QNN use the nonlinear activation functions instead of linear, as used in artificial neural network. Here, the nonlinear activation function consists of linear superposition of multi-sigmoid function. In the recent past, various models of QNN have been used in other research areas such as, prediction of weather, disease diagnosis, voice recognition, and machine translation.^{34–38} In such manner, more states can be expressed in a hidden layer neural cell

in comparison to traditional sigmoid function in which only two states may be expressed.

Algorithm for training the QNN

The QNN with input, output, and hidden units are shown in Figure 2. Here, only one hidden layer has been used. Every node of hidden layer represents three sub-states in itself with the difference of quantum interval θ^r with

Table 5 Forecast of number of cases (both male and female) of CVD in India

Year/age group	20–29	30–39	40–49	50–59	60–69	Total
2000	4,511,192	5,619,470	5,909,011	7,334,040	5,349,313	28,723,024
2005	6,150,408	7,566,792	7,659,859	10,308,145	6,894,266	38,579,471
2010	8,317,184	9,608,484	9,682,192	13,981,687	9,137,258	50,726,804
2015	10,492,011	13,134,021	12,628,644	18,230,363	11,884,240	66,369,279

Abbreviation: CVD, cardiovascular disease.

Table 6 Input parameters

Serial number	Parameters
1	Age
2	Sex
3	Tobacco smoke
4	High density lipoprotein
5	Systolic blood pressure
6	Total cholesterol
7	Medication

quantum level r . Let us assume that n_s denotes the number of grades or excitation levels, η is learning rate which is a small random value, δ_k is error rate of output layer, and δ_j error rate of hidden layer; where n_i denotes the input to the input layer; O_j and O_k denote the output of hidden and output layer, respectively. The weights between input and hidden layers are denoted by W_{ij} and the weights between hidden and output layers are denoted by W_{kj} .³⁷⁻⁴¹

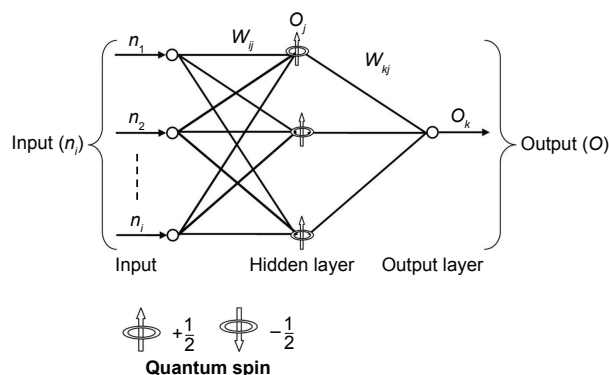
The initial weights are small random numbers and t denotes target value. Given R training pairs $\{n_1, t_1; n_2, t_2; n_R, t_R\}$; where n_i ($J \times 1$) is input and t_i ($K \times 1$) is target values for given inputs, as shown in Figure 2.

The error signal term of the output (δ_k) and hidden layers (δ_j) are written as,

$$\delta_k = (t_k - O_k) O_k (1 - O_k) \quad \{k = 1, 2, 3, \dots, K\} \quad (1)$$

$$\delta_j = O_j (1 - O_j) \sum_k W_{kj}^{New} \delta_k \quad \{j = 1, 2, 3, \dots, J\} \text{ and } \{k = 1, 2, 3, \dots, K\} \quad (2)$$

Consequently, output layer weights (W_{kj}^{New}) and hidden layer weights (W_{ji}^{New}) are adjusted as,

**Figure 2** Architecture of quantum neural network.

Notes: n_i denotes the input to the input layer; O_j and O_k denote the output of hidden and output layer, respectively. The weights between input and hidden layers are denoted by W_{ij} and the weights between hidden and output layers are denoted by W_{kj} .

$$W_{kj}^{New} = W_{kj}^{old} + \eta \delta_k O_j \quad \{j = 1, 2, 3, \dots, J\} \text{ and } \{k = 1, 2, 3, \dots, K\} \quad (3)$$

$$W_{ji}^{New} = W_{ji}^{old} + \eta \delta_j O_i \quad \{i = 1, 2, 3, \dots, I\} \text{ and } \{j = 1, 2, 3, \dots, J\} \quad (4)$$

The simple sigmoid function has been used as the activation function from input to hidden layer and is expressed as:

$$sgm(x) = (1 / (1 + \exp(-x))) \quad (5)$$

The sigmoid function with various graded levels has been used as the activation function for each hidden neuron, ie, from hidden to output layer is expressed as follows:

$$sgm(x) = \sum_{r=1}^{n_r} (1 / (1 + \exp(-x + \theta^r))) \quad (6)$$

After i iterations, when minimum possible error is obtained, then increase quantum interval by very small quantum interval $\Delta\theta$

$$\theta^r = \theta^r + \Delta\theta \quad (7)$$

The step by step details of the above discussed algorithm is expressed by in the flowchart shown in Figure 3.

Experiments and result

During the training process of QNN, the best possible weights have been identified for each node of every layer by conducting the different experiments as shown in Figure 3. The whole dataset has been randomly divided and selected into training set, validation set, and test set as shown in Table 7.

The architecture of the QNN consists of seven input nodes, 85 hidden nodes, and one output node as shown in Figure 2. The numbers of hidden nodes have been identified after several experiments. During our study, we have consulted several physicians and identified four groups of risk categories according to percentage level as “normal”, “low risk range”, “intermediate risk range”, and “highest risk range”. The percentage for each level is shown in Table 8.

During training of the proposed system, we presumed the baseline values of all the risk parameters, these values are identified on the basis of our study and consulting physicians, as shown in Table 9. The baseline value varies on the basis of “age and gender category”.⁴² In our study, we presumed

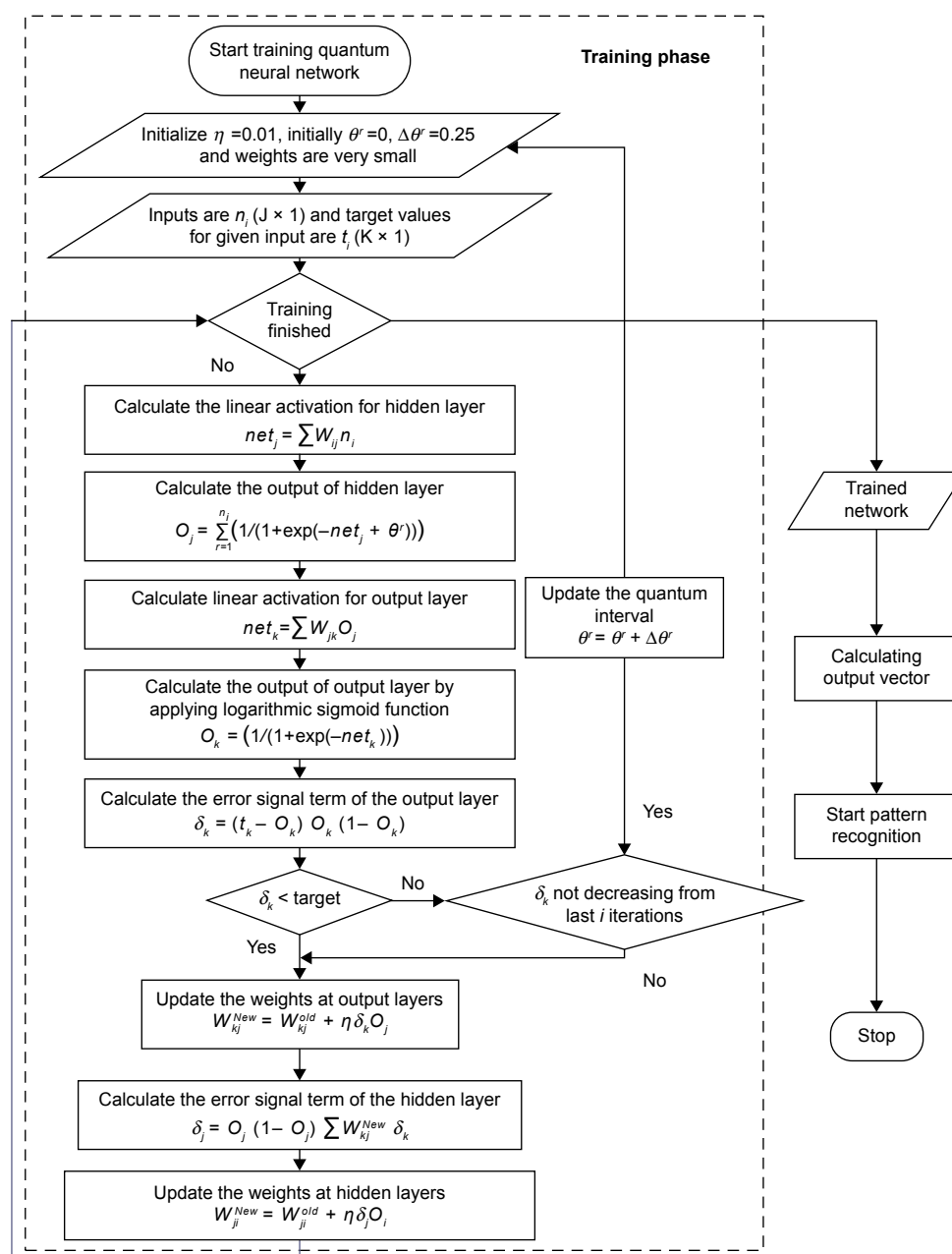


Figure 3 Flowchart of quantum neural network for heart disease prediction system.

Notes: θ' denotes quantum interval between sub-states, with the difference of quantum level i . n_i denotes the number of grades or excitation levels, η is learning rate, δ_k is error rate of output layer, and δ_j error rate of hidden layer. n_i denotes the input to the input layer. O_j and O_k denote the output of hidden and output layer, respectively. The weights between input and hidden layers are denoted by W_{ij} and the weights between hidden and output layers are denoted by W_{kj} . t denotes target value.

if the values lie up to baseline values, the patient should be treated as a healthy person without CVD.

To attain maximum performance, fine-tuning of the network architecture was performed. Particularly, number of

nodes in the hidden layer was adjusted for QNN. In this study, we have only considered the risk factors used in FRS, significant for CVD. After analysis of patients' data, the patients may be categorized on the basis of reference range.

Table 7 Data partition set

Data partition set	Records	Percentage
Training set	189	27.44
Validation set	250	36.28
Test set	250	36.28
Total	689	100.00

Table 8 Percentage level of risk categories

Risk category	Percentage
Normal (under baseline risk)	<26
Low risk range	26–51
Intermediate risk range	51–76
Highest risk range	≥76

Table 9 Baseline value of parameters

Serial no	Age	Sex	TC (mg/dL)	SBP (mmHg)	HDL (mg/dL)
1	20	Male	172	120	60
2	25	Male	174	121	60
3	30	Male	174	122	60
4	35	Male	174	123	60
5	40	Male	174	125	60
6	45	Male	175	127	60
7	50	Male	178	129	60
8	55	Male	182	131	60
9	60	Male	182	134	60
10	65	Male	182	136	60
11	70	Male	182	138	60
12	20	Female	165	120	70
13	25	Female	167	121	70
14	30	Female	167	122	70
15	40	Female	167	123	70
16	45	Female	169	125	70
17	35	Female	172	127	70
18	50	Female	172	129	70
19	55	Female	182	131	70
20	60	Female	182	134	70
21	65	Female	182	136	70
22	70	Female	182	138	70

Abbreviations: TC, total cholesterol; HDL, high density lipoprotein; SBP, systolic blood pressure.

The collected patient information indicated that 63% of them were suffering from high blood pressure followed by 23% with normal blood pressure, and 8% with lower blood pressure. Twenty-seven percent of these patients exhibited increased HDL while 73% had low HDL. Similarly, 82% of the patients exhibited TC in higher range as exhibited in Table 10.

Besides the earlier discussed parameters, 75.65% of patients were nonsmokers and 24.35% were smokers. An amount of 85.50% of patients were on medication and 14.50% were not taking any medication, as shown in Table 11.

Out of 689 patients' data, the data of 250 CVD patients with the doctors' diagnoses (predictions) were collected for evaluation and validation purposes. The proposed system achieved 98.57% accuracy in predicting the level of CVD risk in patients. The result of our proposed system is significantly higher than the other existing approaches as shown in Table 12.

Table 10 Percentage-wise patient distribution for various risk factors

Risk factors	Percentage-wise patient distribution		
	Lower	Average	High
Blood pressure	8%	29%	63%
Total cholesterol	6%	12%	82%
High density lipoprotein	73%	N/A	27%

Abbreviation: N/A, not applicable.

Table 11 Percentage-wise patient distribution for smoking and being on medication

Risk factors	Percentage-wise patient distribution	
	Yes	No
Smoker	24.35%	75.65%
Medication	85.50%	14.50%

It is evident that FRS is a commonly used method by doctors/practitioners worldwide for the prediction of CVDs. We carried out a comparison of our system with the FRS method. The same dataset was used for comparison of both the systems. The tool for FRS is available at <http://cvrisk.mvm.ed.ac.uk/calculator/calc.asp>.⁴³ As discussed earlier, the FRS is outdated in the present scenario, and has many drawbacks as it underestimates the risk. This tool is not designed for a person below 20 years of age.

The proposed system based on QNN shows an average accuracy of 98.57% using the same data used for FRS accuracy prediction level, as shown in Table 13. Figure 4 shows the accuracy graph of the proposed system.

The proposed system has been tested and validated using the dataset of the famous Framingham study. The dataset of 5,209 CVD patients of the Framingham study has been taken from the University of Washington, Seattle, WA, USA (<http://courses.washington.edu/b513/datasets/datasets.php?class=513>).³³ The test results also confirm the effectiveness of the proposed system, as shown in Table 14.

In the Framingham study dataset, out of 5,209 patients, 378 deaths were recorded due to MI and 605 due to CHD. Out of the 378 deaths due to MI, 32.93% of patients had a very high risk prediction and 41.46% had an intermediate risk prediction, only 25.62% of patients had a low risk prediction. In case of deaths recorded due to CHD, out of 605 patients, 37.17% had a high risk prediction and 43.77% an intermediate risk prediction, while 19.06% were placed in the low risk category by the proposed system. In general, all the recorded cases of death were predicted to be at risk by the proposed system, out of which approximately 77% of deaths were predicted as very high risk category and only

Table 12 Comparison of the results of the proposed algorithm with similar algorithms

Method	Accuracy
Proposed algorithm with quantum neural network	98.57%
Fuzzy-evidential based theories ³	91.58%
Neural network ⁴	84%
Bayesian classification and neural network ⁵	80%
Probability theory (logistic regression) ⁶	77%
Framingham risk score ⁴¹	19.22%

Table 13 Data showing comparison of the proposed system and FRS of random testing on different experimental values

Parameters used for comparison of proposed system and FRS (patient record)										Result and accuracy of proposed QNN based system				Accuracy group-wise		
Age (Years)	Sex	TC (mmol/L)	HDL (mmol/L)	Smoker	SBP (mmHg)	Medication	Output Risk category	Manually equated/calculated desired result	Desired result based on doctor's diagnosis	Output Risk category	% accuracy QNN	group-wise	FRS	% accuracy group-wise	FRS	Accuracy group-wise
40	Male	4.20	2.20	No	117	Yes	Normal	18.75% Normal	Normal	18.71% Normal	99.79	99.94	0.99	5.28	0.99	19.60
35	Female	5.00	1.81	No	121	Yes	Low risk	37.50% Low risk	Low risk	37.50% Low risk	100.00		0.77	2.04	0.77	
55	Female	5.28	1.68	No	132	Yes	Low risk	43.75% Low risk	Low risk	43.75% Low risk	100.00		6.32	14.44	6.32	
45	Male	5.28	1.68	Yes	126	Yes	Low risk	43.75% Low risk	Low risk	43.70% Low risk	99.89		9.45	21.60	9.45	
70	Male	5.83	1.42	No	140	Yes	Intermediate risk	56.25% Intermediate risk	Intermediate risk	56.22% Intermediate risk	99.95		27.71	49.26	27.71	
65	Female	4.74	1.94	No	135	Yes	Low risk	31.25% Low risk	Low risk	31.24% Low risk	99.97		8.34	26.70	8.34	
45	Male	7.17	0.78	No	190	Yes	Highest risk	87.50% Highest risk	Highest risk	87.50% Highest risk	100.00		15.64	17.87	15.64	
60	Male	5.28	1.68	No	132	Yes	Low risk	43.75% Low risk	Low risk	43.70% Low risk	99.89	99.96	13.49	30.83	13.49	20.49
50	Male	6.37	1.17	Yes	134	Yes	Intermediate risk	68.75% Intermediate risk	Intermediate risk	68.76% Intermediate risk	100.01		23.37	33.99	23.37	
60	Female	4.20	2.20	No	130	Yes	Normal	18.75% Normal	Normal	18.76% Normal	100.05		4.50	24.00	4.50	
25	Male	6.37	1.17	Yes	126	Yes	Intermediate risk	68.75% Intermediate risk	Intermediate risk	68.76% Intermediate risk	100.01		1.68	2.44	1.68	
60	Male	4.20	2.20	No	126	Yes	Normal	18.75% Normal	Normal	18.71% Normal	99.79		7.43	39.60	7.43	
20	Female	7.74	0.65	No	160	No	Highest risk	93.75% Highest risk	Highest risk	93.75% Highest risk	100.00		0.92	0.99	0.92	
35	Male	6.63	1.04	Yes	129	Yes	Intermediate risk	75.00% Intermediate risk	Intermediate risk	75.00% Intermediate risk	100.00		8.67	11.57	8.67	
60	Male	8.29	0.52	No	190	Yes	Highest risk	100.00% Highest risk	Highest risk	100.00% Highest risk	100.00	99.97	46.52	46.52	46.52	30.57
60	Female	7.74	0.65	No	180	Yes	Highest risk	93.75% Highest risk	Highest risk	93.75% Highest risk	100.00		31.23	33.32	31.23	
25	Female	8.29	0.52	No	165	Yes	Highest risk	100.00% Highest risk	Highest risk	100.00% Highest risk	100.00		3.65	3.65	3.65	
25	Male	5.83	1.42	Yes	123	Yes	Intermediate risk	56.25% Intermediate risk	Intermediate risk	56.22% Intermediate risk	99.95		1.00	1.79	1.00	
45	Male	5.83	1.42	No	129	No	Intermediate risk	56.25% Intermediate risk	Intermediate risk	56.22% Intermediate risk	99.95		6.26	11.13	6.26	
65	Female	7.74	0.65	No	170	Yes	Highest risk	93.75% Highest risk	Highest risk	93.75% Highest risk	100.00		35.44	37.81	35.44	
70	Male	5.28	1.68	Yes	136	Yes	Low risk	43.75% Low risk	Low risk	43.70% Low risk	99.89		34.89	79.74	34.89	
40	Female	4.74	1.94	No	125	Yes	Low risk	31.25% Low risk	Low risk	31.25% Low risk	100.00	99.97	1.26	4.04	1.26	15.77
65	Female	6.37	1.17	No	145	Yes	Intermediate risk	68.75% Intermediate risk	Intermediate risk	68.76% Intermediate risk	100.01		20.08	29.20	20.08	
45	Female	5.00	1.81	No	122	Yes	Low risk	37.50% Low risk	Low risk	37.50% Low risk	100.00		2.16	5.75	2.16	
25	Male	5.28	1.68	No	120	Yes	Low risk	43.75% Low risk	Low risk	43.70% Low risk	99.89		0.17	0.39	0.17	
40	Female	6.09	1.30	No	125	Yes	Intermediate risk	62.50% Intermediate risk	Intermediate risk	62.50% Intermediate risk	100.00		3.08	4.93	3.08	
40	Male	7.74	0.65	Yes	160	No	Highest risk	93.75% Highest risk	Highest risk	93.73% Highest risk	99.98		24.33	25.96	24.33	
65	Male	5.28	1.68	No	134	Yes	Low risk	43.75% Low risk	Low risk	43.70% Low risk	99.89		17.55	40.10	17.55	
65	Male	3.65	2.46	Yes	124	Yes	Normal	6.25% Normal	Normal	6.25% Normal	100.00	99.90	14.53	23.25	14.53	31.00
55	Male	6.09	1.30	Yes	134	Yes	Intermediate risk	62.50% Intermediate risk	Intermediate risk	62.50% Intermediate risk	100.00		26.34	42.15	26.34	
65	Female	5.54	1.55	No	140	Yes	Low risk	50.00% Low risk	Low risk	49.80% Low risk	99.60		13.11	26.21	13.11	
55	Male	4.20	2.20	No	123	No	Normal	18.75% Normal	Normal	18.71% Normal	99.79		5.00	26.67	5.00	
55	Male	6.92	0.91	No	139	Yes	Highest risk	81.25% Highest risk	Highest risk	81.26% Highest risk	100.01		23.89	29.40	23.89	
45	Male	8.29	0.52	Yes	170	No	Highest risk	100.00% Highest risk	Highest risk	100.00% Highest risk	100.00		38.31	38.31	38.31	
20	Male	8.29	0.52	No	160	Yes	Highest risk	100.00% Highest risk	Highest risk	100.00% Highest risk	100.00	98.36	1.10	1.10	1.10	13.96
35	Male	8.29	0.52	Yes	175	Yes	Highest risk	100.00% Highest risk	Highest risk	99.96% Highest risk	99.96		21.32	21.32	21.32	
55	Female	5.83	1.42	No	135	Yes	Intermediate risk	56.25% Intermediate risk	Intermediate risk	50.75% Low risk	90.22		8.85	15.74	8.85	

30	Female	4.74	1.94	No	115	No	31.25%	Low risk	Low risk	31.25%	Low risk	100.00	0.24	0.78
50	Male	7.74	0.65	Yes	175	No	93.75%	Highest risk	Highest risk	93.75%	Highest risk	99.98	40.08	42.75
35	Female	5.28	1.68	No	120	Yes	43.75%	Low risk	Low risk	43.75%	Low risk	100.00	0.91	2.07
60	Female	4.45	2.07	No	130	Yes	25.00%	Normal	Normal	25.01%	Normal	100.04	99.99	5.17
55	Female	3.91	2.33	No	125	Yes	12.50%	Normal	Normal	12.50%	Normal	100.00	2.52	20.69
60	Male	6.92	0.91	Yes	160	Yes	81.25%	Highest risk	Highest risk	81.26%	Highest risk	100.01	44.34	54.57
40	Male	4.74	1.94	No	120	Yes	31.25%	Low risk	Low risk	31.21%	Low risk	99.87	1.57	5.02
50	Male	4.45	2.07	No	123	Yes	25.00%	Normal	Normal	25.00%	Normal	100.00	3.93	15.71
35	Female	3.65	2.46	No	110	Yes	6.25%	Normal	Normal	6.25%	Normal	100.00	0.17	2.75
25	Female	4.20	2.20	No	110	Yes	18.75%	Normal	Normal	18.75%	Normal	100.00	0.04	0.24
20	Female	4.20	2.20	No	110	Yes	18.75%	Normal	Normal	18.75%	Normal	100.00	99.28	0.01
50	Female	4.45	2.07	No	125	Yes	25.00%	Normal	Normal	24.01%	Normal	96.04	2.47	9.90
40	Female	6.37	1.17	No	129	Yes	68.75%	Intermediate risk	Intermediate risk	68.75%	Intermediate risk	100.00	4.10	5.97
65	Male	6.92	0.91	No	175	No	81.25%	Highest risk	Highest risk	81.26%	Highest risk	100.01	36.06	44.38
55	Male	3.91	2.33	No	121	Yes	12.50%	Normal	Normal	12.50%	Normal	100.00	4.10	32.78
25	Female	6.37	1.17	No	125	Yes	68.75%	Intermediate risk	Intermediate risk	68.00%	Intermediate risk	98.91	0.60	0.87
40	Female	4.45	2.07	No	115	Yes	25.00%	Normal	Normal	25.00%	Normal	100.00	0.75	2.99
30	Female	5.83	1.42	No	120	Yes	56.25%	Intermediate risk	Intermediate risk	56.25%	Intermediate risk	100.00	99.97	0.71
60	Male	5.54	1.55	No	134	Yes	50.00%	Low risk	Low risk	50.00%	Low risk	100.00	15.57	31.14
50	Male	6.92	0.91	Yes	190	Yes	81.25%	Highest risk	Highest risk	81.26%	Highest risk	100.01	30.24	37.22
50	Female	6.63	1.04	No	180	Yes	75.00%	Intermediate risk	Intermediate risk	75.00%	Intermediate risk	100.00	11.20	14.94
35	Male	4.20	2.20	No	116	Yes	18.75%	Normal	Normal	18.71%	Normal	99.79	0.46	2.46
55	Female	4.45	2.07	No	124	Yes	25.00%	Normal	Normal	25.00%	Normal	100.00	3.36	13.46
35	Female	6.63	1.04	No	160	Yes	75.00%	Intermediate risk	Intermediate risk	75.00%	Intermediate risk	100.00	2.87	3.82
20	Male	5.83	1.42	Yes	122	Yes	56.25%	Intermediate risk	Intermediate risk	56.22%	Intermediate risk	99.95	99.98	0.28
55	Female	6.63	1.04	No	140	No	75.00%	Intermediate risk	Intermediate risk	75.00%	Intermediate risk	100.00	14.64	19.52
70	Male	6.92	0.91	Yes	180	Yes	81.25%	Highest risk	Highest risk	81.26%	Highest risk	100.01	56.33	69.33
40	Female	5.54	1.55	No	126	Yes	50.00%	Low risk	Low risk	50.00%	Low risk	100.00	2.21	4.42
45	Female	4.74	1.94	No	120	Yes	31.25%	Low risk	Low risk	31.23%	Low risk	99.94	1.72	5.50
25	Female	4.74	1.94	No	115	Yes	31.25%	Low risk	Low risk	31.25%	Low risk	100.00	0.09	0.30
30	Male	3.65	2.46	Yes	112	No	6.25%	Normal	Normal	6.25%	Normal	100.00	0.36	5.75

Abbreviations: TC, total cholesterol; HDL, high density lipoprotein; SBP, systolic blood pressure; QNN, quantum neural network; FRS, Framingham risk score.

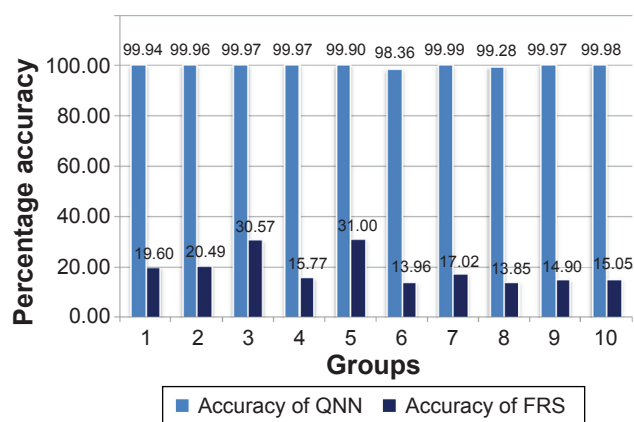


Figure 4 Graph showing accuracy of proposed system and Framingham risk score (FRS).

Abbreviation: QNN, quantum neural network.

23% were predicted as low risk. Here, in one of the most important causes for mortality, it must be taken into consideration that medication can affect mortality even though the patient is at low risk. The CVD risk predictions by the proposed system, using the dataset of the Framingham study, confirmed the potential risk of death, deaths which actually occurred and had been recorded as due to MI in the dataset of the Framingham study.

Discussion

The FRS and European heart score are commonly used methods which are considered as most effective methods for CVD risk prediction. However, these methods have some major drawbacks such as being less accurate and based on a very old study conducted in the 1960s wherein the parameters taken into consideration were also limited. The validity of the Framingham equation is undetermined as the lifestyle of the population has changed completely. The newly available risk parameters are not included. On the other hand, these systems are based on conventional statistical methods and are not updated with time according to changed environmental and lifestyle conditions. Presently, the accuracy of the FRS is only 19.22%, as evaluated and compared with the proposed system using the same data. The major drawbacks of these studies are their ineffectiveness with regard to value ranges, ie, the FRS is only applicable

for patients aged 20 to 100 years and the maximum threshold limit of CVD risk is 30%.

On the other hand in European heart score, if blood pressure is over 180 mmHg or cholesterol over 8 mmol/L, patients are automatically classified as high risk without considering the medication factor.⁴⁴ In many situations, with proper and effective medication, a patient with high risk parameters such as blood pressure, cholesterol, may survive for a very long time.

Other than the FRS and European heart score, the fuzzy-evidential based CVD prediction system has been proposed by Khatibi and Montazer in 2010, but they have achieved only 91.58% accuracy.³ Akay proposed the neural network based coronary artery disease prediction system in 1992, achieved an accuracy of 84%, and used the records of 100 patients for testing, but they have only considered the clinical parameters.⁴ The Bayesian classification and neural network based ischemic heart disease prediction system was proposed by Kukar et al in 1999, but in their study, ECG data were used as input parameter for prediction.⁵ The probability theory (logistic regression) angiographic coronary disease based system was proposed by Detrano et al in 1989, they collected the data of 425 patients and achieved only 77% accuracy. In their study, only clinical and noninvasive test data were considered as input parameters for their system.⁶

In the current study, we have introduced a new machine learning based system which uses the QNN for learning and for its knowledge buildup. On the basis of knowledge gained from data patterns, the proposed system predicts the CVD risk with very high accuracy and is able to update itself with time. Over a period of time, as the data increase, it will update and correct itself on the basis of the current trends and data pattern.

The proposed CVD system has been tested and validated using Indian patients' data and American patients' data, and achieved the highest accuracy among existing systems.

Conclusion

As compared and evaluated during experiments, the proposed CVD prediction system shows the best results on the dataset used, when compared to the FRS. This system has been

Table 14 Validation based on the dataset from the Framingham study of 5,209 American CVD patients

Out of 5,209 CVD patients	Very high risk ^a	Intermediate risk ^a	Low risk ^a	Normal ^a
Deaths recorded due to myocardial infarction =378	32.93%	41.46%	25.61%	Nil
Deaths recorded due to coronary heart disease =605	37.17%	43.77%	19.06%	Nil

Note: ^aPredicted by the proposed system.

Abbreviation: CVD, cardiovascular disease.

evaluated and validated with the data of patients and the doctors' diagnoses (predictions). This system will be serving as an aid to medical practitioners in planning better medication strategies. An early diagnosis may be effectively performed by using this system. Data showing prediction accuracy of random testing on different experimental values are shown in Table 13. In predicting the risk level, an overall accuracy of 98.57% has been achieved by the proposed system, whereas with the FRS the highest level of accuracy was 19.22%. This system's accuracy is considerably higher compared to other existing approaches. Thus, this system must be used instead of well-known FRS. Hence, the proposed system will serve as a significant tool for doctors/practitioners.

Acknowledgment

The authors would like to thank Dr Praveen Garg, cardiologist of Rameshwar Das Memorial Hospital, Jagadhri, India for his valuable guidance and continued support throughout the work.

Disclosure

The authors report no conflicts of interest in this work.

References

- who.int/en/ [homepage on the Internet]. Cardiovascular diseases (CVDs) Fact sheet N°317; WHO; 2013 [updated January 2015]. Available from: <http://www.who.int/mediacentre/factsheets/fs317/en/index.html>. Accessed May 27, 2016.
- Wu R, Peters W, Morgan MW. The next generation clinical decision support: linking evidence to best practice. *J Healthc Inf Manag*. 2002; 16(4):50–55.
- Khatibi V, Montazer GA. A fuzzy-evidential hybrid inference engine for coronary heart disease risk assessment. *Expert Syst Appl*. 2010; 37(12):8536–8542.
- Akay M. Noninvasive diagnosis of coronary artery disease using a neural network algorithm. *Biol Cybern*. 1992;67(4):361–367.
- Kukar M, Kononenko I, Groselj C, Kralj K, Fettich J. Analysing and improving the diagnosis of ischaemic heart disease with machine learning. *Artif Intell Med*. 1999;16(1):25–50.
- Detrano R, Janosi A, Steinbrunn W, et al. International application of a new probability algorithm for the diagnosis of coronary artery disease. *Am J Cardiol*. 1989;64(5):304–310.
- Narain R, Saxena S, Goyal AK. Cardiovascular Disease Prediction based on Physical factors using Quantum Neural Network. *International Journal of Information Processing*. 2014;8(3):83–94.
- Mainous AG 3rd, Koopman RJ, Diaz VA, Everett CJ, Wilson PW, Tilley BC. A coronary heart disease risk score based on patient reported information. *Am J Cardiol*. 2007;99(9):1236–1241.
- Capewell S, Ford ES, Croft JB, Critchley JA, Greenlund KJ, Labarthe DR. Cardiovascular risk factor trends and potential for reducing coronary heart disease mortality in the United States of America. *Bull World Health Organ*. 2010;88(2):120–130.
- Lloyd-Jones DM, Larson MG, Beiser A, Levy D. Lifetime risk of developing coronary heart disease. *Lancet*. 1999;353(9147):89–92.
- Legato MJ, Padus E, Slaughter E. Women's perceptions of their general health with special reference to their risk of coronary artery disease: results of a national telephone survey. *J Womens Health*. 1997;6(2):189–198.
- Pilote L, Hlatky MA. Attitudes of women toward hormone therapy and prevention of heart disease. *Am Heart J*. 1995;129(6):1237–1238.
- Ezzati M, Hoon SV, Rodgers A, et al. Estimates of global and regional potential health gains from reducing multiple selected major risk factors. *Lancet*. 2003;362(9380):271–280.
- Assmann G, Schulte H, von Eckardstein A, Huang Y. High-density lipoprotein cholesterol as a predictor of coronary heart disease risk. The PROCAM experience and pathophysiological implications for reverse cholesterol transport. *Atherosclerosis*. 1996;124 Suppl: S11–S20.
- Turner RC, Millns H, Neil HA, et al. Risk factors for coronary artery disease in non-insulin dependent diabetes mellitus: United Kingdom Prospective Diabetes Study (UKPDS: 23). *BMJ*. 1998;316(7134):823–828.
- Gordon T, Castelli WP, Hjortland MC, Kannel WB, Dawber TR. High density lipoprotein as a protective factor against coronary heart disease. The Framingham Study. *Am J Med*. 1977;62(5):707–714.
- McMahon R, Peto R, Cuttler J, et al. Blood pressure, stroke and coronary heart disease. Part 1, Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet*. 1990;335(8692):765–774.
- Kannel WB. Blood pressure as a cardiovascular risk factor: prevention and treatment. *JAMA*. 1996;275(20):1571–1576.
- Walker WG, Neaton JD, Cutler JA, Neuwirth, Cohen JD. Renal function change in hypertensive members of the Multiple Risk Factor Intervention Trial. Racial and treatment effects. The MRFIT Research Group. *JAMA*. 1992;268(21):3085–3091.
- Anderson KM, Wilson PW, Odell PM, Kannel WB. An updated coronary risk profile. A statement for health professionals. *Circulation*. 1991; 83(1):356–362.
- Barrett-Connor EL, Cohn BA, Wingard DL, Edelstein SL. Why Is diabetes mellitus a stronger risk factor for fatal ischemic heart disease in women than in men? The Rancho Bernardo Study. *JAMA*. 1991;265(5): 627–631.
- American Diabetes Association. Detection and management of lipid disorders in diabetes. *Diabetes Care*. 1993;16(5):828–834.
- Head J, Fuller JH. International variations in mortality among diabetic patients: the WHO Multinational Study of Vascular Disease in Diabetics. *Diabetologia*. 1990;33(8):477–481.
- Dudina A, Cooney MT, Bacquer DD, et al. Relationships between body mass index, cardiovascular mortality, and risk factors: a report from the SCORE investigators. *Eur J Cardiovasc Prev Rehabil*. 2011;18(5): 731–742.
- Tverdal A, Thelle D, Stensvold I, Leren P, Bjartveit K. Mortality in relation to smoking history: 13 years' follow-up of 68,000 Norwegian men and women 35–49 years. *J Clin Epidemiol*. 1993;46(5): 475–487.
- Doll R, Peto R, Wheatley K, Gray R, Sutherland I. Mortality in relation to smoking: 40 years' observations on male British doctors. *BMJ*. 1994; 309(6959):901–911.
- Willett WC, Green A, Stampfer MJ, et al. Relative and absolute excess risks of coronary heart disease among women who smoke cigarettes. *N Engl J Med*. 1987;317(21):1303–1309.
- Rosengren A, Wilhelmsen L, Wedel H. Coronary heart disease, cancer and mortality in male middle-aged light smokers. *J Intern Med*. 1992; 231(4):357–362.
- Bartechi CE, MacKenzie TD, Schrier RW. The human costs of tobacco use. *N Engl J Med*. 1994;330:907–912.
- MacKenzie TD, Bartechi CE, Schrier RW. The human costs of tobacco use. *N Engl J Med*. 1994;330:975–980.
- Manson JE, Tosteson H, Ridker PM, et al. The primary prevention of myocardial infarction. *N Engl J Med*. 1992;326(21):1406–1416.
- Indrayan A. Forecasting vascular disease cases and associated mortality in India. *NCMH Background Papers-Burden of disease in India*. 2005;(9): 198–215.
- Biostat 513 Datasets [database on the Internet]. University of Washington. Available from: <http://courses.washington.edu/b513/datasets/datasets.php?class=513>. Accessed May 29, 2016.

34. Zhou J, Gan Q, Krzyzak A, Suen CY. Recognition of handwritten numerals by quantum neural network with fuzzy features. *Int J Doc Anal Recognit*. 1999;2:30–36.
35. Karayiannis NB, Mukherjee A, Glover JR, et al. An evaluation of quantum neural networks in the detection of epileptic seizures in the neonatal electroencephalogram. *Soft Comput*. 2006;10(4):382–396.
36. Karayiannis NB, Mukherjee A, Glover JR, et al. Detection of pseudosinusoidal epileptic segments in the neonatal EEG by cascading a rule-based algorithm with a neural network. *IEEE Trans Biomed Eng*. 2006;53(4):633–641.
37. Narayan R, Chakraverty S, Singh VP. Machine translation using quantum neural network for simple sentences. *International Journal of Information and Computation Technology*. 2013;3(7):683–690.
38. Narayan R, Singh VP, Chakraverty S. Quantum neural network based machine translator for Hindi to English. *ScientificWorldJournal*. 2014; 2014:485737.
39. Narayan R, Singh VP, Chakraverty S. Quantum Neural Network based Parts of Speech Tagger for Hindi. *International Journal of Advancements in Technology*. 2014;5(2):137–152.
40. Narayan R, Chakraverty S, Singh VP. *Neural Network based Parts of Speech Tagger for Hindi. Proceedings of 3rd International conference, Advances and control and Optimisation of dynamical systems, Kanpur, India, March 13–15, 2014*. Kanpur, India: proceedings of IFAC-Elsevier; 2014.
41. Narayan R, Chakraverty S, Singh VP. Quantum neural network based machine translator for English to Hindi. *Applied Soft Computing*. 2016; 38:1060–1075.
42. Schaefer JE, Lamon-Fava S, Cohn SD, et al. Effects of age, gender, and menopausal status on plasma low density lipoprotein cholesterol and apolipoprotein B levels in the Framingham Offspring Study. *J Lipid Res*. 1994;35(5):779–792.
43. Cardiovascular Risk Calculator. The University of Edinburgh [updated: May 19, 2010]. Available from: <http://cvrisk.mvm.ed.ac.uk/calculator/calc.asp>. Accessed May 29, 2016.
44. European Society of Cardiology [homepage on the Internet]. HeartScore. Available from: <https://escol.escardio.org/heartscore/download.aspx>. Accessed May 29, 2016.

Patient Preference and Adherence

Publish your work in this journal

Patient Preference and Adherence is an international, peer-reviewed, open access journal that focuses on the growing importance of patient preference and adherence throughout the therapeutic continuum. Patient satisfaction, acceptability, quality of life, compliance, persistence and their role in developing new therapeutic modalities and compounds to optimize

clinical outcomes for existing disease states are major areas of interest for the journal. This journal has been accepted for indexing on PubMed Central. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/patient-preference-and-adherence-journal>

Dovepress

Cardiovascular Disease Prediction based on Physical Factors using Quantum Neural Network

Renu Narain^a, Sanjai Saxena^b, Achal Kumar Goyal^c

^aDepartment of Biotechnology, Thapar University, Patiala, Punjab-147004, India.
Contact: renu_narain@hotmail.com

^bDepartment of Biotechnology, Thapar University, Patiala, Punjab-147004, India.

^cComputer Center, Gurukul Kangri University, Haridwar, Uttarakhand, India.

This Paper presents the intelligent system for prediction of Cardiovascular Disease. The diagnosis of Cardiovascular Disease is tedious task for a Medical Practitioner. The correct presumptions are most important for detection of heart disease which is caused due to various factors or symptoms. In order to make it effective and reliable this system automatically learns and extracts the significant patterns of the heart disease using the Novel approach of Quantum Neural Network. Our System can predict the Cardiovascular Disease at its primary stage using the Physical, physiological and clinical risk Factors, very efficiently by Quantum Neural Network (QNN). The prediction with QNN shows better results than other available algorithms using same parameters.

Keywords : Atherosclerosis, Myocardial Infarction, Physical Risk Factors, Quantum Neural Network.

1. INTRODUCTION

Cardiovascular disease is the foremost cause of mortality and disability in most of the countries. Deaths due to cardiovascular disease have declined over time in developed nations [1, 2]. Although these data are encouraging, there is a lack of knowledge about national trends in the risk factors that lead to cardiovascular disease. The future incidence of cardiovascular disease can be determined through knowledge of risk factors and their trends in younger populations in whom clinically evident heart disease has not yet developed. By the time that heart problems are detected, the underlying cause (atherosclerosis) is usually quite advanced, having progressed for decades. There is therefore increased emphasis on preventing atherosclerosis by modifying risk factors, such as healthy eating, exercise and avoidance of smoking. The importance of risk factors is further underscored by the substantial reductions in atherosclerosis that can be achieved by primary prevention [3].

Previous studies shows that people of South Asian origin have higher rates of this disease than people of European origin, a finding that cannot be explained by differences in conventional cardiovascular risk factors, such as smoking, raised blood pressure, diabetes, or high cholesterol [4,5].

2. RELATED WORK

Many risk assessments are based on relatively simplistic strategies that have a clinician identify whether a "risk factor" is present. However, national guidelines advocate the use of scoring systems for CHD risk. Sophisticated methods like the Framingham Risk Score (FRS), the European Systematic Coronary Risk Evaluation (SCORE), the Sheffield risk and treatment table allow calculation of an individual's risk as a function of the patient's values on selected established coronary heart disease risk factors [6]. Listed below are the variables used for risk score computation by the different models used:

2.1. Framingham Risk Score

The variables used for risk score computation are Gender, age, Total Cholesterol (TC), HDL, Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), Diabetes Mellitus (DM), Body Mass Index (BMI) and smoking [6]. The data of group of patients was collected in the 1960s and 1970s and this data is used in the Framingham equations. Since then the cases of CHD mortality has been decreasing and this change is due to the change in the prevalence of risk factors taken into the consideration in Framingham equation. The lifestyle and other risk factors have changed and affected the incidence of CHD. If so, in modern population, the validity of the Framingham equation may be undermined. For example, some other risk factor such as serum fibrinogen levels, homocysteine, or an as yet undetected factor may have altered which would result in the inaccurate Framingham predictions.

2.2. The Joint British Cardiac Society (BCS) / British Hypertension Society (BHS) / British Hyperlipidemia Association (BHA) Risk Score Risk Score

The variables used for risk score computation are Gender, age, Total Cholesterol, HDL, SBP, DBP, DM, Smoking, ECG-LVH if available

2.3. The European SCORE

According to the national guidelines, several European countries uses the European Heart SCORE model as the basis for primary prevention and treatment of Cardio Vascular Disease (CVD). The variables used in individuals' 10 year risk assessment are age, gender, smoking, SBP and Total cholesterol level [7].

2.4. FINRISK Model

It is based on its European origin. It included Diabetes and LDL [8]. The FINRISK model was derived from the occurrence of acute myocardial infarction or coronary heart disease death during a 10 year follow-up of populations in eastern and southwestern Finland, based on the 1982 and 1987 cohorts.

2.5. Berto Model

The model from Berto and co-workers starts with healthy individuals and model their life until different coronary heart diseases or death by other causes [9]. This model does not go any further after the first CHD event.

2.6. CDC Model

CDC (Centers for Disease Control and prevention) diabetes cost-effectiveness group made a Markov model which models whether an individual gets CHD or continues to stay "normal". In contrary to *e.g.*, Berto's model, this model doesn't stop at first event, but also models that one can stay in the health states "angina" and "history of CA/MI" for several years.

2.7. Huse Model

Huse and co-workers have, as CDC, modelled up to all individuals are dead [10]. This is however a model which evaluates the difference between different statins. In this model it is possible to start and stop giving medications during the process, which is rather uncommon for models like these.

2.8. Nyman *et al.*

has an easily understood model which models the whole process up to death [11]. In this model, all individuals who survive the first year after a cardiovascular event will end up in a "chronic state" and be there until their death.

2.9. Weinstein's Model

It is possibly the most extensive of those mentioned here [12]. This model has five health states; persons without CHD, persons with new CHD, persons with CHD, persons who survived up to the age of 85 years, and death. Hence this model has two different absorbing states; survive 85 and death. Individuals in the state "new CHD" will only be there the first year, and after that transfer to "CHD", "survive until 85" or die.

2.10. CANFIS Model

An approach based on Co-Active Neuro-Fuzzy Inference System (CANFIS) was presented for prediction of heart disease. The proposed CANFIS model combined the neural network

Table 1
List of Parameters Used for Prediction

S.No.	Parameters
1	Tobacco Smoke
2	High Blood Pressure
3	Body Mass Index
4	Stress
5	Alcohol
6	Age
7	Gender
8	Family History

adaptive capabilities and the fuzzy logic qualitative approach which is then integrated with genetic algorithm to diagnose the presence of the disease. Coactive Neuro-fuzzy modeling was proposed as a dependable and robust method developed to identify a nonlinear relationship and mapping between the different attributes. It has been shown that of Genetic Algorithm (GA) is a very useful technique for auto-tuning of the CANFIS parameters and selection of optimal feature set [13].

3. MATERIALS AND METHODS

3.1. Data Description - Patients Data

The data for this study have been collected from 572 patients who have symptoms of heart disease. In this work, parameters significant to cardiovascular Disease are used with their weightage and the priority levels are advised by the medical experts. Table 1 shows the various input parameters for the prediction of heart disease.

For training and evaluation purpose we have carefully predicted the level of heart disease by taking the help of Cardiologist/ Physician. The data consists of patients' record with doctor's predictions/ diagnosis. The Quantum Neural Network is trained with the selected significant patterns. During analysis, the column type is identified. During data analysis, the last column is considered as the target one and other columns will be considered as input columns. The whole dataset is divided into training, validation and testing as shown in Ta-

Table 2
Data Partition Set

S.No.	Data Partition Set	Partition Set	Records Percentage
1	Training Set	72	12.58
2	Validation Set	250	43.71
3	Test Set	250	43.71
4	Total	572	100

ble 2.

3.2. Parameters Used for Prediction

The major risk factors that can be modified, treated or controlled by changing our lifestyle or taking medicine: -

3.2.1. Tobacco Smoke

Tobacco consumption increases the risk of Myocardial Infarction (MI). Most of the previous studies which shows increase in Tobacco consumption increases the risk of Myocardial Infarction used a cutoff point of 10, 15, or 20 cigarettes per day [14, 15]. The harmful effects of light smoking with adequate accuracy requires large cohort studies and only few studies have attempted this [16, 17]. In the Framingham Heart study, cardiovascular mortality increased 18 percent in men and 31 percent in women for each 10 cigarettes smoked per day [18]. In addition, the use of tobacco products in individuals with other cardiac risk factors was found to have a synergistic effect on Coronary Artery Disease morbidity and mortality: smoking was found to increase the risk for Coronary Artery Disease, stroke, heart failure and peripheral vascular disease at every level of blood pressure.

Smokers' risk of developing coronary heart disease is 2-4 times that of non-smokers. Cigarette smoking is a powerful independent risk factor for sudden cardiac death in patients with coronary heart disease; smokers have about twice the risk of non-smokers. People who smoke or pipes seem to have a higher risk of death from coronary heart disease (and possibly stroke) but their risk isn't as great as cigarette smokers'. Exposure to other peo-

ple's smoke increase the risk of heart disease even for non- smokers. There is overwhelming evidence for an adverse effect of smoking on the risk of coronary heart disease and other atherosclerotic disease. Smoking is responsible for 50% of all avoidable deaths and one half of these are due to cardiovascular disease [19, 20]. This adverse effect of smoking is related to the amount of tobacco smoked daily and the duration of smoking [21].

3.2.2. High Blood Pressure

High blood pressure increases the heart's workload, causing the heart to thicken and become stiffer. It also increases the risk of stroke, heart attack, kidney failure and congestive heart failure. The risk of heart attack or stroke increases several times when high blood pressure exists with obesity, smoking, high blood cholesterol levels or diabetes. Numerous observational epidemiological studies in geographically and ethnically diverse populations have established a direct relation between blood pressure elevation and incidence of Coronary Artery Disease and stroke. The importance of elevated blood pressure as a risk factor for coronary heart disease, heart failure, cerebrovascular disease and renal failure in both men and women has been demonstrated in a large number of epidemiological studies [18]. The agewise reference range for high blood pressure is given in Table 3.

3.2.3. Body Mass Index (BMI)

The body Mass Index (BMI), or Quetelet index, is a statistical measurement which compares a person's weight and height. Though it does not actually measure the percentage of body fat, it is a useful tool to estimate a healthy body weight based on how tall a person is. Due to its ease of measurement and calculation, it is the most widely used diagnostic tool to identify weight problem within a population including: underweight, overweight and Obesity.

$$BMI = Weight(Kg)/Height^2(m^2) \quad (1)$$

Table 3

Age Wise Reference Range of Blood Pressure

Age Group	Lower Limit*	Average Limit*	Upper Limit*
15-19	105/73	117/77	120/81
20-24	108/75	120/79	132/83
25-29	109/76	121/80	133/84
30-34	110/77	122/81	134/85
35-39	111/78	123/82	135/86
40-44	112/79	125/83	137/87
45-49	115/80	127/84	139/88
50-54	116/81	129/85	142/89
55-59	118/82	131/86	144/90
60-64	121/83	134/87	147/91
65-69	122/84	136/88	149/92
70 and above	123/85	138/89	151/93

(*systolic/diastolic blood pressure range)

3.2.4. Stress

Individual response to stress may be a contributing factor. Some scientists have noted a relationship between coronary heart disease risk and stress in a person's life, their health behaviors and socioeconomic status. These factors may affect established risk factors. For example, people under stress may over eat; start smoking more than they otherwise would. Stress is considered a contributing risk factor for heart disease because its effects on the heart are not completely understood. Also, the effects of emotional stress, behavior habits, and socioeconomic status on the risk of heart disease and heart attack have not been proven. That is because we all deal with stress differently: how much and in what way stress affects us varies from person to person.

3.2.5. Alcohol

Drinking too much alcohol can raise pressure, cause heart failure and lead to stroke. It can contribute to high triglycerides, cancer and other diseases, and produce irregular heartbeats. It contributes to obesity, alcoholism, suicide and accidents. The risk of heart disease in people who drink moderate amounts of alcohol is lower than in non- drinkers. The role of alcohol in Coronary Artery Disease risk is

complicated by difficulties in obtaining accurate data on individual alcohol consumption. In a number of studies, moderate alcohol intake has been associated with decreased coronary risk [22], and this protective effect may be mediated by an increase in HDL cholesterol. A study shows that 60 ml alcohol is not too much harmful but more than 60 ml increases the risk of cardiovascular diseases. Reduced mortality of light or moderate drinkers is due to their lower mortality from coronary heart disease, independent of the type of drink [23, 24].

3.2.6. Age

The risk of heart disease increases with the age. In fact, about 4 of every 5 deaths due to heart disease occur in people older than 65. With the increase in age, our heart tends not to work as well. The walls of the heart may thicken and arteries become stiff and hard, making the heart less able to pump blood to the muscles of the body. Due to these changes, the risk of developing cardiovascular disease increases with the age. Women 65 and older have about the same risk of cardiovascular disease as men of the same age.

3.2.7. Gender

Studies shows that men have a higher risk of heart attack than women. But the difference narrows after women reach menopause. After the age of 65, the risk of heart disease is about the same between the sexes when other risk factors are similar.

3.2.8. Family History

A family history of early heart disease is also a risk factor, including heart disease in the father or a brother diagnosed before age 55 years and in the mother or a sister diagnosed before age 65 years.

4. PROPOSED CARDIOVASCULAR DISEASE PREDICTION SYSTEM

This System provides the comprehensive predictive tool for predicting cardiovascular disease; this system predicts the cardiovascular disease in early stage by clinicians, doctors,

pathologists as well as by general public also may predict the chances of having the cardiovascular diseases using the important physiological tests caused for cardiovascular disease, so that the clinicians may fulfill their promises and commitments for ensuring sustained fight against this deadly disease. In our model for predicting, the major physiological factors are considered, which have not been considered altogether in any system, developed so far.

4.1. Quantum Neural Network

Quantum Neural Network based on multi-level transfer function was introduced by Karayianis *et al.*. The major difference between conventional neural network and quantum neural network is the form of the nonlinear activation functions of their hidden units. The activation function of quantum neural network express as linear superposition of multi-sigmoid function. Various models of Quantum Neural Network based on Multi-level transfer function have been used in the prediction of weather, disease diagnosis and voice recognition [25,26,27]. In such manner, more states are expressed in a hidden layer neural cell in comparison to traditional sigmoid function in which only two states are expressed. The quantum interval for each sigmoid function is different. It is obtained by training of quantum neural network. The model of Quantum Neural Network is useful in the indetermination and crossed pattern problem. Multilevel transfer function of neural cell in hidden layer represents the feature of fuzziness. It's effective to classify indeterminate data.

4.2. Algorithm for Training the Quantum Neural Network

The QNN with input, output and hidden units are shown in Figure 1. Here, only one hidden layer has been used. Every node of hidden layer represents three sub states in itself with the difference of quantum interval θ^r with quantum level r , let us assume that n_s denotes the number of grades or excitation levels, is learning rate which is a small random value, k is error rate of output layer and j error rate of hidden layer. Where n_i denotes the input to the

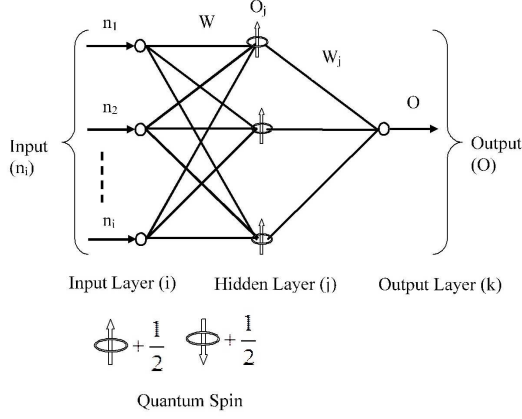


Figure 1. Architecture Diagram of Quantum Neural Network for Cardiovascular Disease Prediction System

input layer. O_j and O_k denotes the output of hidden and output layer, respectively. The weights between input and hidden layers are denoted by W_{ij} and the weights between hidden and output layers are denoted by W_{kj} . The initial weights are small random numbers and t denotes Target value. First apply the inputs to the network and calculate the output [28-36]. Given R training pairs $\{n_1, t_1; n_2, t_2; n_R, t_R\}$ Where $n_i (J \times 1)$ is input and $t_i (K \times 1)$ is target values for given inputs, for the present quantum neural network as shown in Figure 1. The error signal term of the output (δ_k) and hidden layers (δ_j) are written as,

$$\delta_k = (t_k - O_k)O_k(1 - O_k) \quad (2)$$

$$\{k = 1, 2, 3, \dots, K\}$$

$$\delta_j = O_j(1 - O_j) \sum_k W_{kj}^{New} \delta_k \quad (3)$$

$\{j = 1, 2, 3, \dots, J\} \{k = 1, 2, 3, \dots, K\}$ Consequently, output layer weights (W_{kj}^{New}) and hidden layer weights (W_{ji}^{New}) are adjusted as-

$$W_{kj}^{New} = W_{kj}^{Old} + \eta \delta_k O_j \quad (4)$$

$$\{j = 1, 2, 3, \dots, J\} \{k = 1, 2, 3, \dots, K\}$$

$$W_{ji}^{New} = W_{ji}^{Old} + \eta \delta_j O_i \quad (5)$$

$\{i = 1, 2, 3, \dots, I\} \{j = 1, 2, 3, \dots, J\}$ The simple sigmoid function has been used as the acti-

vation function from input to hidden layer and is expressed as:

$$sgm(x) = (1/(1 + \exp(-x))) \quad (6)$$

The sigmoid function with various graded levels has been used as the activation function for each hidden neuron *i.e.*, from hidden to output layer is expressed as follows:

$$sgm(x) = (1/(1 + \exp(-x + \theta^r))) \quad (7)$$

After i iterations when minimum possible error is obtained, then increase quantum interval by very small quantum interval $\Delta\theta$

$$\Delta^r = \Delta^r + \Delta\theta \quad (8)$$

4.3. Results and Discussions

The proposed system predicts the rate of heart disease or the percentage rate of disease on the basis of physical risk factors; it also predicts the future condition in advance. The Proposed system is based on Quantum Neural Network, means it is predictive as well as adopts the predictive knowledge based on the past experience *i.e.*, on the basis of past data it learns, and gradually maximizes the accuracy level of prediction. We have categorized the risk level into four groups according to percentage level as Normal, Low Risk Range, Intermediate risk range and highest risk range. The percentage for each level is shown in Table 4.

The dataset is divided into training set, validation set and test set as shown in Table 2. The architecture of the quantum neural network used in this study is having 9 input nodes, 115 hidden nodes and 1 output nodes as shown in Figure 1. The Quantum Neural Network is trained with the above mentioned number of nodes and best training performance graph is shown in the Figure 2. The number of input nodes is determined by the physical parameters, the number of hidden nodes is determined through trial and error; and the number of output node is represented as percentage level of cardiovascular disease. Training a quantum neural network is the process of setting the best weights on the inputs of each of the units. The purpose is to use the training set to produce

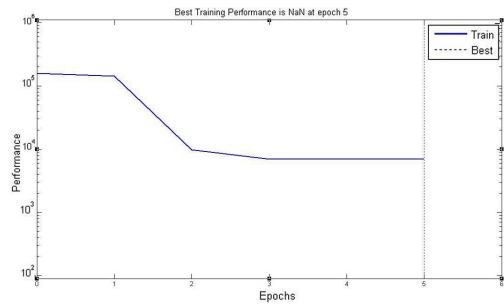


Figure 2. Graph Showing Trained Quantum Neural Network

Table 4
Percentage Level of Risk Categories

Risk Category	Percentage
Normal	Below 26
Low Risk Range	26-51
Intermediate risk range	51-76
Highest risk range	76 and above

weights where the output of the network is as close to the desired output as possible for as many of the examples in the training set as possible. The training set is a part of the input dataset used for neural network training, *i.e.*, for adjustment of network weights. To attain maximum performance fine tuning of the parameters were executed. Particularly, number of nodes in the hidden layer was adjusted when using Quantum Neural Network while the spread parameter was optimized. The optimum number of nodes in the hidden layer was determined by altering the number of nodes from 1 to 115 while fixing the settings for the remaining parameters. The results are plotted as shown in Figure 3. The data have been analyzed and shows 98% accuracy in prediction as matched with the available data and results discussed with the physicians. The data for this study have been collected from 250 patients who have symptoms of cardiovascular diseases. The accuracy of the proposed algorithm is compared with other available algorithms and showed better results for accuracy as shown in Table 5. This work described here shows that the prediction of risk from car-

Table 5
Data Partition Set

S. No.	Method	Accuracy
1	Proposed Algorithm with QNN	98%
2	Framingham Risk Score	85%
3	Bayesian classification and neural network [31]	80%
4	Neural network [32]	84%
5	Neural network [33]	48%
6	Probability theory[34]	77%
7	Clustering (CLASSIT conceptual system) [35]	78.90%
8	Fuzzy sets and evidence theories [36]	91.58%

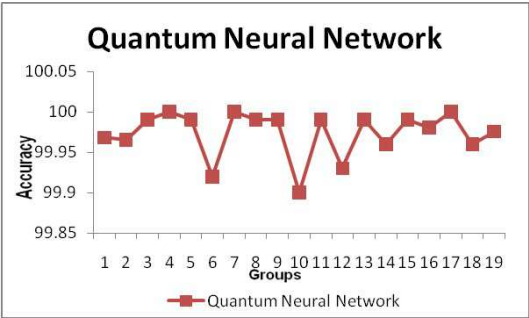


Figure 3. Plot of the Number of Nodes in Hidden Layer and Percentage Error

diovascular diseases gives best results on the dataset used. The results generated by this system have been verified with the physicians and are found correct. This QNN model will help the doctors to plan for a better medication and provide the patient with early diagnosis as it performs reasonably well even without retraining. In conclusion, when the QNN was trained and tested after optimizing the input parameters, the overall predictive accuracy obtained was 98%Data Showing Prediction accuracy of Random Testing on different Experimental values has been given in Table 6.

Table 6
Graph Showing Trained Quantum Neural Network

A g e	Sm ok ing	S B P	D B P	B M I	St re ss	Al co hol	Ge nd er	Family History	Desired Output	Result QNN	Accu -racy	Group wise Accuracy	Final Accu -racy
30	0	113	78	26	0	0	0	0	12.5	12.5	100	99.97	99.97
35	0	117	80	27	0	0	0	0	25	25	100		
40	0	112	79	25	0	0	0	0	0	0	100		
40	1	136	86	39	1	1	1	1	93.75	93.75	100		
45	0	127	84	30	0	0	0	0	50	50	100		
50	0	118	81	26	0	0	0	0	6.25	6.23	99.68		
55	0	125	84	27	0	0	0	0	25	25	100		
60	0	121	83	25	0	0	0	0	0	0	100		
65	1	138	88	32	1	1	1	1	56.25	56.25	100		
70	1	140	89	32	1	1	1	1	56.25	56.25	100		
20	0	109	75	26	0	0	0	0	6.25	6.23	99.68	99.97	
25	0	121	80	30	0	0	0	0	50	50	100		
40	0	115	80	26	0	0	0	0	12.5	12.5	100		
40	1	128	84	34	1	1	1	1	62.5	62.5	100		
45	0	123	82	28	0	0	0	0	31.25	31.25	100		
50	0	121	82	27	0	0	0	0	18.75	18.75	100		
55	0	128	85	28	0	0	0	0	37.5	37.5	100		
60	1	136	87	32	1	1	1	1	56.25	56.25	100		
65	1	141	89	35	1	1	1	1	68.75	68.73	99.97		
70	0	134	88	28	0	0	1	0	37.5	37.5	100		
25	0	115	78	27	0	0	0	0	25	25	100	99.99	
25	1	129	82	37	1	1	1	1	81.25	81.25	100		
35	0	114	79	26	0	0	0	0	12.5	12.5	100		
35	1	128	83	35	1	1	1	1	68.75	68.73	99.97		
40	1	133	85	37	1	1	1	1	81.25	81.25	100		
45	1	139	88	40	1	1	1	1	100	100	100		
50	0	129	85	30	0	0	0	0	50	50	100		
65	0	122	84	25	0	0	1	0	0	0	100		
65	0	136	88	30	0	0	1	0	50	50	100		
70	0	136	88	29	0	0	1	0	43.75	43.72	99.93		
20	0	113	76	27	0	0	0	0	18.75	18.75	100	100	
25	1	133	84	40	1	1	1	1	100	100	100		
35	1	132	85	38	1	1	1	1	87.5	87.5	100		
40	1	131	85	36	1	1	1	1	75	75	100		
45	1	130	85	34	1	1	1	1	62.5	62.5	100		
45	1	138	87	39	1	1	1	1	93.75	93.75	100		
55	1	139	88	37	1	1	1	1	81.25	81.25	100		
60	0	124	84	26	0	0	0	0	12.5	12.5	100		
65	0	127	85	27	0	0	1	0	18.75	18.75	100		
70	0	131	87	27	0	0	1	0	25	25	100		

A g e	Sm ok ing	S B P	D B P	B M I	St re ss	Al co hol	Ge nd er	Family History	Desired Output	Result QNN	Accu -racy	Group wise Accuracy	Final Accu -racy
20	0	116	77	28	0	0	0	0	31.25	31.25	100	99.99	
25	0	117	78	28	0	0	0	0	31.25	31.25	100		
30	0	118	79	28	0	0	0	0	31.25	31.25	100		
35	1	135	86	40	1	1	1	1	100	100	100		
45	1	132	85	35	1	1	1	1	68.75	68.73	99.97		
50	1	134	86	35	1	1	1	1	68.75	68.73	99.97		
55	1	141	89	38	1	1	1	1	87.5	87.5	100		
60	1	137	88	34	1	1	1	1	62.5	62.5	100		
65	0	129	86	27	0	0	1	0	25	25	100		
70	0	132	87	28	0	0	1	0	31.25	31.25	100		
20	1	122	79	32	1	1	1	1	56.25	56.25	100	99.92	
30	0	112	77	26	0	0	0	0	6.25	6.23	99.68		
35	0	122	81	29	0	0	0	0	43.75	43.72	99.93		
40	0	117	80	27	0	0	0	0	18.75	18.75	100		
45	0	117	80	26	0	0	0	0	6.25	6.23	99.68		
50	0	123	83	27	0	0	0	0	25	25	100		
55	0	129	85	29	0	0	0	0	43.75	43.72	99.93		
60	1	142	89	37	1	1	1	1	81.25	81.25	100		
70	0	127	86	26	0	0	1	0	12.5	12.5	100		
70	1	145	91	36	1	1	1	1	75	75	100		
25	0	109	76	25	0	0	0	0	0	0	100	100	
25	1	124	81	34	1	1	1	1	62.5	62.5	100		
30	1	128	83	36	1	1	1	1	75	75	100		
40	0	119	81	27	0	0	0	0	25	25	100		
45	1	133	86	36	1	1	1	1	75	75	100		
50	1	136	87	36	1	1	1	1	75	75	100		
55	1	142	89	39	1	1	1	1	93.75	93.75	100		
60	1	139	88	35	1	1	1	1	68.75	68.73	99.97		
65	0	131	86	28	0	0	1	0	31.25	31.25	100		
70	0	129	86	27	0	0	1	0	18.75	18.75	100		
25	0	114	77	27	0	0	0	0	18.75	18.75	100	99.99	
25	1	126	81	35	1	1	1	1	68.75	68.73	99.97		
30	1	130	83	37	1	1	1	1	81.25	81.25	100		
40	0	120	81	28	0	0	0	0	31.25	31.25	100		
45	1	135	86	37	1	1	1	1	81.25	81.25	100		
50	1	137	87	37	1	1	1	1	81.25	81.25	100		
55	1	144	90	40	1	1	1	1	100	100	100		
60	0	129	85	28	0	0	0	0	31.25	31.25	100		
65	0	133	87	28	0	0	1	0	37.5	37.5	100		
70	1	147	92	38	1	1	1	1	87.5	87.5	100		
20	0	120	79	30	0	0	0	0	50	50	100	99.99	
20	1	131	82	39	1	1	1	1	93.75	93.75	100		
25	1	132	83	39	1	1	1	1	93.75	93.75	100		
30	1	127	82	35	1	1	1	1	68.75	68.73	99.97		
35	0	111	78	25	0	0	0	0	0	0	100		
35	1	131	84	37	1	1	1	1	81.25	81.25	100		
40	1	127	83	32	1	1	1	1	56.25	56.25	100		
50	1	132	86	34	1	1	1	1	62.5	62.5	100		
55	0	118	82	25	0	0	0	0	0	0	100		
65	0	126	85	26	0	0	1	0	12.5	12.5	100		

5. CONCLUSIONS

The proposed work shows that the prediction of risk from cardiovascular diseases gives best results based on physical factors. The result generated by this system has been evaluated and validated on data of patients with the Doctor's diagnosis (predictions). This proposed system will help the doctors to plan for a better medication and provide the patient with early diagnosis as it performs reasonably well even without retraining. The QNN has been trained and tested after optimizing the input parameters. The overall predictive accuracy obtained was 98% . Data Showing Prediction accuracy of Random Testing on different Experimental values has been given in (Table 6). Hence the proposed system will work as a significant tool for doctors/ practitioners.

REFERENCES

1. Tu J V, Nardi L and Fang J. For the Canadian Cardiovascular Outcomes Research Team. *National Trends in Rates of Death and Hospital Admissions Related to Acute Myocardial Infarction, Heart Failure and Stroke*, in *CMAJ*, 180(3):E118–E125, 2009.
2. Rosamond W, Flegal K and Furie K. Heart Disease and Stroke Statistics 2008 Update, *A Report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee*, in *Circulation*, 117:125–146, 2008.
3. Ford ES, Ajani UA and Croft J B. Explaining the Decrease in U.S. Deaths from Coronary Disease, 1980–2000, in *N Engl Journal of Medicine*, 356:2388–98, 2007.
4. McKeigue P M, Shah B and Marmot M G. Relation of Central Obesity and Insulin Resistance with High Diabetes Prevalence and Cardiovascular Risk in South Asians, in *Lancet*, 337:382–86, 1991.
5. McKeigue P M, Ferrie J E, Pierpoint T and Marmot M G. Association of Early-onset Coronary Heart Disease, in *South Asian Men with Glucose Intolerance and Hyperinsulinemia*, in *Circulation*, pages 152–161, 1993.
6. Arch G Mainous, Richelle J Koopman, Vanessa A Diaz, Charles J Everett, Peter W F Wilson and Barbara C Tilley. A Coronary Heart Disease Risk Score Based on Patient Reported Information, in *American Journal of Cardiology*, 99(9):1236–1241, 2007.
7. Stovring H, Harmsen C G, Wisloff T, Jarbol D E, Nexoe J, Nielsen J B and Kristiansen I S. Competing Risk Approach for the European Heart SCORE Model based on Cause-specific and All-cause Mortality, in *European Journal of Prev Cardiol.*, 20(5):827–836, 2013.
8. Raj Bhopal, Colin Fischbacher, Erkki Vartiainen, Nigel Unwin, Martin White and George Alberti. Predicted and Observed Cardiovascular Disease in South Asians: Application of FINRISK, Framingham and SCORE Models to Newcastle Heart Project data, in *Journal of Public Health*, 27(1):93–100, 2005.
9. Berto P, Munro V, Gaddi A, Negrini C, Hutton J and Mast O. Cost-Effectiveness Analysis for Statin Therapies in the Primary Prevention of Coronary Heart Disease in Italy, in *Clinical Drug Investigation*, 20:107–721, 2000.
10. Huse D M, Russell M W, Miller J D, Kraemer D F, D'Agostino R B and Ellison R C. Cost-Effectiveness of Statins, in *American Journal of Cardiology*, 82:1357–1363, 1998.
11. Nyman J A, Martinson M S, Nelson D, Nugent S, Collins D and Wittes J. Cost-Effectiveness of Gemfibrozil for Coronary Heart Disease Patients with Low Levels of High-Density Lipoprotein Cholesterol: The Department of Veterans Affairs High-Density Lipoprotein Cholesterol Intervention Trial, in *Arch. International Medicine*, 162:177–182, 2002.
12. Weinstein M C, Coxson P G, Williams L W, Pass T M, Stason W B and Goldman L. Forecasting Coronary Heart Disease Incidence, Mortality and Cost: the Coronary Heart Disease Policy Model, in *American Journal of Public Health*, 77:1417–1426, 1987.
13. Parthiban Latha and Subramanian R. Intelligent Heart Disease Prediction System Using CANFIS and Genetic Algorithm, in *International Journal of Biological, Biomedical and Medical Sciences*, 3(3):157–160, 2008.
14. Tverdal A, Thelle D and Stensvold I. Mortality in Relation to Smoking History: 13 years Follow-up of 68,000 Norwegian Men and Women 35–49 years, in *Journal of Clinical Epidemiology*, 46:475–487, 1993.
15. Doll R, Peto R and Wheatley K. Mortality in Relation to Smoking: 40 years Observations on Male British Doctors, in *BMJ*, 309:901–911, 1994.
16. Willett W C, Green A and Stampfer M J. Relative and Absolute Excess Risks of Coronary

- Heart Disease Among Women Who Smoke Cigarettes, in *N Engl Journal of Medicine*, 317:1303–1309, 1987.
17. Rosengren A, Wilhelmsen L and Wedel H. Coronary heart disease, cancer and mortality in male middle-aged light smokers, in *Journal of Intern Medicine*, 231:357–62, 1992.
 18. Kannel W B. Blood Pressure as a Cardiovascular Risk Factor: Prevention and Treatment, in *JAMA*, 275:1571–1576, 1996.
 19. Bartechi C E, MacKenzie T D and Schrier R W. The Human Costs of Tobacco Use (first of two parts), in *N Engl Journal of Medicine*, 330:907–12, 1994.
 20. MacKenzie T D, Bartechi C E and Schrier R W. The Human Costs of Tobacco Use (second of two parts), in *N Engl Journal of Medicine*, 330:975–980, 1994.
 21. Manson J E, Tosteson H and Ridker P M. The Primary Prevention of Myocardial Infarction, in *N Engl Journal of Medicine*, 326:1406–1416, 1992.
 22. Pohorecky L A. Interaction of Alcohol and Stress at the Cardiovascular Level, in *Alcohol*, 7:537–542, 2000.
 23. Shaper A G, Wannamethee G and Walker M. Alcohol and Coronary Heart Disease: A Perspective from the British Regional Heart Study, in *International Journal of Epidemiology*, 23:482–94, 1994.
 24. Rehm J T, Bondy S J, Sempos C T and Vuong C V. Alcohol Consumption and Coronary Heart Disease Morbidity and Mortality, in *American Journal of Epidemiology*, 146:495–501, 1997.
 25. Zhou J, Gan Q and Krzyzak A. Recognition of Handwritten Numerals by Quantum Neural Network with Fuzzy Features, in *International Journal on Document Analysis and Recognition*, 30–36, 1999.
 26. Karayiannis N B, Mukherjee A and Glover J R. An Evaluation of Quantum Neural Networks in the Detection of Epileptic Seizures in the Neonatal Electroencephalogram, in *Soft Computing*, pages 382–396, 2006.
 27. Karayiannis N B, Mukherjee A and Glover J R. Detection of Pseudosinusoidal Epileptic Segments in the Neonatal EEG by Cascading a Rule Based Algorithm with a Neural Network, in *IEEE Transactions on Biomedical Engineering*, pages 633–641, 2006.
 28. Narayan R, Chakraverty S and Singh V P. Quantum Neural Network based Machine Translator for Hindi to English, in *The Scientific World Journal*, Article ID 485737, 2014.
 29. Chakraverty S, Gupta P and Sharma S. Neural Network-based Simulation for Response Identification of Two-Storey Shear Building Subject to Earthquake Motion, in *Journal of Neural Computing and Applications*, 3(19):367–375, 2010.
 30. Narayan R, Chakraverty S and Singh V P. Machine Translation using Quantum Neural Network for Simple Sentences, in *International Journal of Information and Computation Technology*, 3(7):683–690, 2013.
 31. Kukar M, Kononenko I, Groselj C, Kralj K and Fettich. Analyzing and Improving the Diagnosis of Ischaemic Heart Disease with Machine Learning, in *Artificial Intelligence in Medicine*, 16:25–50, 1999.
 32. Akay M. Noninvasive Diagnosis of Coronary Artery Disease using a Neural Network Algorithm, in *Biological Cybernetics*, 67:361–367, 1992.
 33. Haddad M, Adlassnig K P and Porenta G. Feasibility Analysis of a case based Reasoning System for Automated Detection of Coronary Heart Disease from Myocardial Scintigrams, in *Artificial Intelligence in Medicine*, 9:61–78, 1999.
 34. Detrano R, Janosi A, Steinbrunn W, Pfisterer M, Schmid J and Sandhu S. International Application of a New Probability Algorithm for the Diagnosis of Coronary Artery Disease, in *American Journal of Cardiology*, 64:304–310, 1989.
 35. Gennari J H, Langley P and Fisher D. Models of Incremental Concept Formation, in *Artificial Intelligence*, 40:11–61, 1989.
 36. Vahid Khatibi and Gholam Ali Montazer. A Fuzzy-Evidential Hybrid Inference Engine for Coronary Heart Disease Risk Assessment, in *Expert Systems with Applications*, 37:8536–8542, 2010.



Renu Narain received her M.Sc. degree from the MJP Rohilkhand University, India, in 2005. She is certified on IBM DB2 and RAD. She is presently a Ph.D candidate in the Department of Biotechnology at Thapar University, Patiala (Punjab), India.

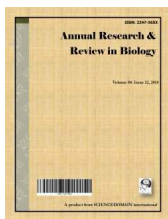


Sanjai Saxena did his Ph.D in Biological Science Biotechnology). Currently he is Associate Professor, Department of Biotechnology, Thapar University, Patiala, Punjab. He authored five book Chapters and published various research papers in national and international journals. He is reviewer and member of editorial board of many international journals. He has many patents on his name. He received many prestigious awards *viz.* Bharat Jyoti Award, Dr. M J Thirumalachar Merit Award, D Dharmadhikari Gold medal *etc.*, His research areas are Drug Discovery, Bioinformatics Ethnopharmacology, Plant-Microbe Interaction, Microbial

Secondary Metabolites, Clinical Microbiology and Role of Microbes in Development of Composite Structures for Industrial Applications



Achal Kumar Goyal did his Ph.D in computer science. Presently he is working as a Head, University Computer Center, Gurukul Kangri University, Haridwar (Uttarakhand). His research interest is in Interconnection Networks, Mobile Networks and Bioinformatics. He published various research papers in journals and conferences. He is member of Computer Society of India.



A Novel Heart Disease Prediction System Based on Quantum Neural Network Using Clinical Parameters

Renu Narain^{1*}, Sanjai Saxena¹ and Achal Kumar Goyal²

¹Department of Biotechnology, Thapar University, Patiala, Punjab -147004, India.

²University Computer Center, Gurukul Kangri University, Haridwar, Uttarakhand, India.

Authors' contributions

This work was carried out in collaboration between all authors. Author RN had designed the study with the help of Author SS. Author RN had written the draft of the manuscript. Authors SS and AKG had managed the analyses of the study. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/ARRB/2017/10456

Editor(s):

(1) George Perry, University of Texas at San Antonio, USA.

Reviewers:

(1) Preecha Yupapin, King Mongkut's Institute of Technology Ladkrabang, Thailand.

(2) Ladislav Zjavka, VSB Technical University of Ostrava, Czech Republic.

Complete Peer review History: <http://www.sciencedomain.org/review-history/19878>

Original Research Article

Received 28th March 2014

Accepted 29th April 2014

Published 5th July 2017

ABSTRACT

Aims: The diagnosis of Heart disease at earliest possible stage is very crucial to increase the chance of successful treatment and to reduce the mortality rate. The interpretation of cardiovascular disease is time-consuming and requires analysis by an expert physician. Thus there is a need of expert system which may provide quick and accurate prediction of Heart disease at early possible stage, without the help of physician.

Place and Duration of Study: The study was carried out during 2010 to 2013 in the vicinity of Yamuna Nagar, Haryana, India.

Methodology: The data used for this study consists of clinical values (Diabetes Mellitus, Low Density Lipoprotein, Triglycerides and High Density Lipoprotein) and has been collected from various Hospitals of 689 patients, who have symptoms of heart disease. All these cases are analyzed after careful scrutiny with the help of the Physicians. For training and evaluation purpose we have carefully predicted the level of heart disease by taking the help of Cardiologist/ Physician. The data consists of patients' record with doctor's predictions/ diagnosis.

*Corresponding author: E-mail: renu_narain@hotmail.com;

Results: The obtained result of Heart disease prediction match with the expert physician's opinion with 96.97% accuracy and shows high degrees of sensitivity and specificity.

Conclusion: The proposed Heart Disease Prediction System based on Quantum Neural Network gives the high degrees of accuracy in predicting the risk of cardiovascular diseases, are also the best results based on clinical factors. The result generated by this system has been evaluated and validated on data of patients with the Doctor's diagnosis. This system will help the doctors to plan for a better medication and provide the patient with early diagnosis as it performs reasonably well even without retraining. Such an expert system may also prove useful in combination with other systems to providing diagnostic and predictive medical opinions in a timely manner.

Keywords: Myocardial infarction; quantum neural network; atherosclerosis; clinical risk factors.

1. INTRODUCTION

Heart disease is one of the major causes of death worldwide. According to World Health Organization's report, cardiovascular diseases (CVD) are the number one cause of death globally. More people die annually from CVDs than from any other cause. An estimated 17.3 million people died from CVDs in 2008, representing 30% of all global deaths. Low and middle-income countries are disproportionately affected: over 80% of CVD deaths take place in low and middle-income countries and occur almost equally in men and women [1].

Nowadays, many hospitals use some sort of information systems to manage their healthcare or patient data [2]. Unfortunately, these data are infrequently used to support clinical decision making. The main objective is to develop an Intelligent Heart Disease Prediction System with Quantum Neural Network using heart disease databases to make intelligent clinical decisions which cannot be done in traditional decision support systems.

In this study, we used Quantum Neural Network for the identification of CVD patients and it also predicts the rate of risk. The conventional Neural Network differs from the Quantum Neural Network (QNN) in the form of nonlinear activation functions of their hidden units. A multilevel activation function is used in QNN instead of the ordinary sigmoid function. Each multilevel function consists of the sum of sigmoid functions shifted by the quantum intervals. All models were trained and tested on the test dataset using clinical risk factors.

The major clinical risk factors that are used in our model can be modified, treated or controlled by

changing our lifestyle or taking medicine and their normal reference ranges are shown in Table 1.

1.1 Diabetes Mellitus (DM)

Diabetes seriously increases the risk of developing cardiovascular disease. Both major types of diabetes mellitus, Type 1 (insulin dependent) diabetes and Type 2 (non-insulin-dependent) diabetes are associated with a significantly increased risk of coronary heart disease, cerebrovascular disease and peripheral vascular disease [3-8]. Even when glucose (blood sugar) levels are under control, diabetes increases the risk of heart disease and stroke, but the risks are even greater if blood sugar is not well controlled. About three –quarters of people of world population are diabetic. In this situation it is extremely important to work with healthcare provider to manage it. Coronary Artery Disease is a major complication of both insulin – dependent diabetes mellitus (IDDM) and non- insulin – dependent diabetes mellitus (NIDDM). In 14 years of follow- up, the Rancho Bernardo Study, in which 334 men and women with NIDDM were compared with 2137 men and women without diabetes, the relative risk for Coronary Artery Disease deaths was 1.9 in diabetic men and 3.3 in diabetic women compared with non- diabetic men and women after adjustment for other Coronary Artery Disease risk factors [9]. The relation between diabetes and cardiovascular diseases are not uniform in all populations. In the WHO Multinational Study of vascular Disease in Diabetics, the incidence of death in diabetic patients that was attributable to circulatory disease ranged from 32 percent in men and 0 percent in women in Tokyo to 67 percent in men and 47 percent in women in London [10].

Table 1. Reference range of clinical risk factors/ input parameters

#	Clinical risk factor	Lower limit	Average limit	Upper limit
1.	DM	80 mg/dl	140 mg/dl	> 200 mg/dl
2.	LDL	70 mg/dl	130 mg/dl (low Risk)	> 160(High Risk)
3.	HDL	35 mg/dl		80 mg/dl
4.	TG	149 mg/dl	199 mg/dl	499 mg/dl

1.2 Low Density Lipoprotein (LDL)

Low density lipoprotein (LDL) is one of the most fascinating macromolecular assemblies ever evolved by nature and has consequently attracted the attention of scientists for decades. As the principal transporter for cholesterol in the blood, circulating LDL guarantees a constant supply of cholesterol for tissues and cells and is an essential constituent of human life. The cellular uptake of LDL is mediated by specific receptors, either classical B/E receptor or the scavenger receptor pathway [11,12].

1.3 High Density Lipoprotein (HDL)

Abnormality of high density lipoprotein may increase the risk of atherosclerosis. Studies have shown that high-density lipoprotein (HDL) cholesterol levels are a strong, independent inverse predictor of cardiovascular disease [13-17]. According to the Framingham Heart Study, HDL level was considered as more persuasive risk factor for coronary heart disease than the level of low-density lipoprotein (LDL) [16].

1.4 Triglycerides (TG)

High triglycerides raise the risk of heart disease. The American Heart Association (AHA) recommends that a triglyceride level of 100 mg/dL(1.3 mmol/L) or lower is considered "optimal." The AHA says this optimal level would improve heart health. However, the AHA doesn't recommend drug treatment to reach this level. Instead, for those trying to lower their triglycerides to this level, lifestyle changes such as diet, weight loss and physical activity are encouraged. That's because triglycerides usually respond well to dietary and lifestyle changes.

1.5 Related Work

Many risk assessments are based on relatively simplistic strategies, the clinician may identify whether a "risk factor" is present. However, national guidelines advocate the use of scoring systems for CHD risk. Sophisticated methods like the Framingham Risk Score (FRS), the European

Systematic Coronary Risk Evaluation (SCORE), the Sheffield risk and treatment table allow calculation of an individual's risk, as a function of the patient's values on selected established coronary heart disease risk factors [18]. The variables used for risk score computation by the different models used, are listed below:

1.6 Framingham Risk Score

The variables used for risk score computation are Gender, age, total cholesterol (TC), HDL, systolic blood pressure (SBP), diastolic blood pressure (DBP), diabetes mellitus (DM), Body Mass Index (BMI) and smoking [18].

The Framingham equations used were developed from cohort data collected in the 1960s and 1970s. We know that since then there has been a fall in the incidence of CHD mortality and a change in the prevalence of smoking. Whilst it may be that the change in incidence is due to the change in the prevalence of risk factors included in the Framingham equation, we cannot be sure that there are not other extraneous factors that have varied over that time and may have affected the incidence of CHD. If so, the validity of the Framingham equation in modern populations may be undermined. For example, some other risk factors such as serum fibrinogen levels, homocysteine, chlamydia infection or any undetected factors may have altered which would distort the Framingham predictions.

1.7 The Joint British Cardiac Society (Bcs)/ British Hypertension Society (Bhs)/ British Hyperlipidemia Association (Bha) Risk Score

The variables used for risk score computation are Gender, age, Total Cholesterol, HDL, Systolic Blood Pressure (SBP), DBP, DM, Smoking, ECG-LVH if available.

1.8 The European Score

The European Heart SCORE model constitutes the basis for national guidelines for primary

prevention and treatment of cardiovascular disease (CVD) in several European countries. The model estimates individuals' 10-year CVD mortality risks from age, sex, smoking status, systolic blood pressure, and total cholesterol level. The SCORE model, however, is not mathematically consistent and does not estimate all-cause mortality [19].

1.9 Finrisk Model

It is based on its European origin. It included Diabetes and LDL [20]. The FINRISK model was derived from the occurrence of acute myocardial infarction or coronary heart disease death during a 10 year follow-up of populations in eastern and southwestern Finland, based on the 1982 and 1987 cohorts.

1.10 Berto Model

The model from Berto and co-workers starts with healthy individuals and model their life until different coronary heartdiseases (CHD) or death by other causes [21]. This model does not go any further after the first CHD event.

1.11 CDC Model

CDC (Centers for disease control and prevention) diabetes cost-effectiveness group made a Markov model which models whether an individual gets CHD or continues to stay "normal". In contrary to e.g. Berto's model, this model doesn't stop at first event, but also models that one can stay in the health states "angina" and "history of CA/MI" for several years.

1.12 Huse Model

Huse and co-workers have, as CDC, modeled up to all individuals are dead [22]. This is however a model which evaluates the difference between different statins. In this model it is possible to start and stop giving medications during the process, which is rather uncommon for models like these.

1.13 Nyman et al. Model

It models the whole process up to death [23]. In this model, all individuals who survive the first year after a cardiovascular event will end up in a "chronic state" and be there until their death.

1.14 Weinstein's Model

This model is possibly the most extensive of those mentioned here [24]. This model has five

health states; persons without CHD, persons with new CHD, persons with CHD, persons who survived up to the age of 85 years, and death. Hence this model has two different absorbing states; survive 85 and death. Individuals in the state "new CHD" will only be there the first year, and after that transfer to "CHD", "survive until 85" or die.

1.15 CANFIS Model

An approach based on coactive neuro-fuzzy inference system (CANFIS) was presented for prediction of heart disease. The proposed CANFIS model combined the neural network adaptive capabilities and the fuzzy logic qualitative approach which is then integrated with genetic algorithm to diagnose the presence of the disease. Coactive Neuro-fuzzy modeling was proposed as a dependable and robust method developed to identify a nonlinear relationship and mapping between the different attributes. It has been shown that of Genetic Algorithm (GA) is a very useful technique for auto-tuning of the CANFIS parameters and selection of optimal feature set [25].

2. MATERIALS AND METHODS

2.1 Data Description - Patients Data

The data for this study has been collected from various Hospitals of 689 patients, who have symptoms of heart disease. All these cases are analyzed after careful scrutiny with the help of the Physicians. Table 1 shows the various input parameters for the prediction of heart disease. For training and evaluation purpose we have carefully predicted the level of heart disease by taking the help of Cardiologist/ Physician. The data consists of patients' record with doctor's predictions/ diagnosis. The whole dataset is divided into training, validation and testing as shown in Table 2.

2.2 Proposed Cardiovascular Disease Prediction Systems

This System provides the comprehensive predictive tool for predicting cardiovascular disease; this system predicts the cardiovascular disease in early stage by clinicians, doctors, pathologists as well as by general public also, may predict the chances of having the cardiovascular diseases using the important Clinical tests caused for cardiovascular disease, so that the clinicians may fulfill their promises and

Table 2. Data Partition Set

#	Data partition set	Records	Percentage
1	Training set	189	27.44
2	Validation set	250	36.28
3	Test set	250	36.28
4	Total	689	100.00

commitments for ensuring sustained fight against this deadly disease. The major Clinical factors are considered for prediction, which have not been considered altogether in any system, developed so far.

2.3 Quantum Neural Network Used for Cardiovascular Disease Prediction System

Quantum Neural Network based on multi-level transfer function was introduced by Karayiannis et al. The major difference between conventional neural network and quantum neural network is the form of the nonlinear activation functions of their hidden units. The activation function of quantum neural network express as linear superposition of multi-sigmoid function. Various models of Quantum Neural Network based on Multi-level transfer function have been used in the prediction of weather, disease diagnosis and voice recognition [26,27,28]. In such manner, more states are expressed in a hidden layer neural cell in comparison to traditional sigmoid function. The quantum interval for each sigmoid function is different. It is obtained by training of quantum neural network. The model of Quantum Neural Network is useful in the indetermination and crossed pattern problem. Multilevel transfer function of neural cell in hidden layer represents the feature of fuzziness. It's effective to classify indeterminate data.

2.4 Algorithm for Training the Quantum Neural Network

The QNN used for our proposed system consists of input, output and hidden units, as shown in Fig. 2. Here, only one hidden layer has been used. In our proposed system, the QNN's every node of hidden layer represents only three sub states in itself with the difference of quantum interval θ^r with quantum level r . Here more than three sub states may be possible in every node of hidden layer, but with every additional sub state the computational complexity increases, but for our proposed system, with three sub states the performance of QNN is optimum with significant accuracy. Let us assume that n_s

denotes the number of grades or excitation levels, η is learning rate which is a small random value, δ_k is error rate of output layer and δ_j error rate of hidden layer. Where n_i denotes the input to the input layer. O_j and O_k denotes the output of hidden and output layer, respectively. The weights between input and hidden layers are denoted by W_{ij} and the weights between hidden and output layers are denoted by W_{kj} . The initial weights are small random numbers and t denotes Target value. First apply the inputs to the network and calculate the output [29,30,31].

Given R training pairs $\{n_1, t_1; n_2, t_2; n_R, t_R\}$

Where n_i (J X 1) is input and t_i (K X 1) is target values for given inputs, for the present quantum neural network as shown in Fig. 1.

The error signal term of the output (δ_k) and hidden layers (δ_j) are written as,

$$\delta_k = (t_k - O_k) O_k (1 - O_k) \quad \{k=1, 2, 3, \dots, K\} \quad (1)$$

$$\delta_j = O_j (1 - O_j) \sum_k W_{kj}^{New} \delta_k \quad \{j=1, 2, 3, \dots, J\} \text{ and } \{k=1, 2, 3, \dots, K\} \quad (2)$$

Consequently, output layer weights (W_{kj}^{New}) and hidden layer weights (W_{ji}^{New}) are adjusted as.

$$W_{kj}^{New} = W_{kj}^{old} + \eta \delta_k O_j \quad \{j=1, 2, 3, \dots, J\} \text{ and } \{k=1, 2, 3, \dots, K\} \quad (3)$$

$$W_{ji}^{New} = W_{ji}^{old} + \eta \delta_j O_i \quad \{i=1, 2, 3, \dots, I\} \text{ and } \{j=1, 2, 3, \dots, J\} \quad (4)$$

The simple sigmoid function has been used as the activation function from input to hidden layer and is expressed as:

$$sgm(x) = (1 / (1 + \exp(-x))) \quad (5)$$

The sigmoid function with various graded levels has been used as the activation function for each hidden neuron i.e. from hidden to output layer is expressed as follows:

$$sgm(x) = \sum_{r=1}^{n_s} \left(\frac{1}{1 + \exp(-x + \theta^r)} \right) \quad (6)$$

Here the initial value of θ^r is initialized by zero, then apply Eq1 to 6 and iterate this process several times upto i iterations until it reaches the lowest possible error and then apply Eq 7.

After i iterations when minimum possible error is obtained, then increase in quantum interval by very small quantum interval $\Delta\theta$, for this proposed system we have used the value of $\Delta\theta$ is 0.25.

$$\theta^r = \theta^r + \Delta\theta \quad (7)$$

Repeat this step until δ_k is decreased to an acceptable accuracy.

3. RESULTS AND DISCUSSION

The proposed system predicts the rate of heart disease or the percentage rate of disease on the basis of Clinical risk factors; it also predicts the future condition in advance. The Proposed system is based on Quantum Neural Network, means it is predictive as well as adopts the predictive knowledge based on the past experience i.e. on the basis of past data it learns, and gradually maximizes the accuracy level of prediction. We have categorized the risk level into

four groups according to percentage level as Normal, Low Risk Range, Intermediate risk range and highest risk range. The percentage for each level is shown in Table 3.

The results of our experimental analysis for significant patterns for cardiovascular disease prediction using Clinical risk factors are presented in this section. In this work, we have only used Clinical risk factors significant to cardiovascular Disease. After analysis of patients' data, the patients may be categorized on the basis of reference range. As shown in the Fig. 1, the total numbers of patients are distributed percentage wise patient distribution lies in Lower, Average and High Level Reference range of Diabetes Mellitus (DM), Low Density Lipoprotein (LDL), Triglycerides (TG) and High Density Lipoprotein (HDL).With the help of the dataset, the patterns significant to the cardiovascular disease prediction are extracted. The dataset is divided into training set, validation set and test set as shown in Table 2.

Table 3. Percentage level of risk categories

Risk category	Percentage
Normal	Below 26
Low risk range	26-51
Intermediate risk range	51-76
Highest risk range	76 and above

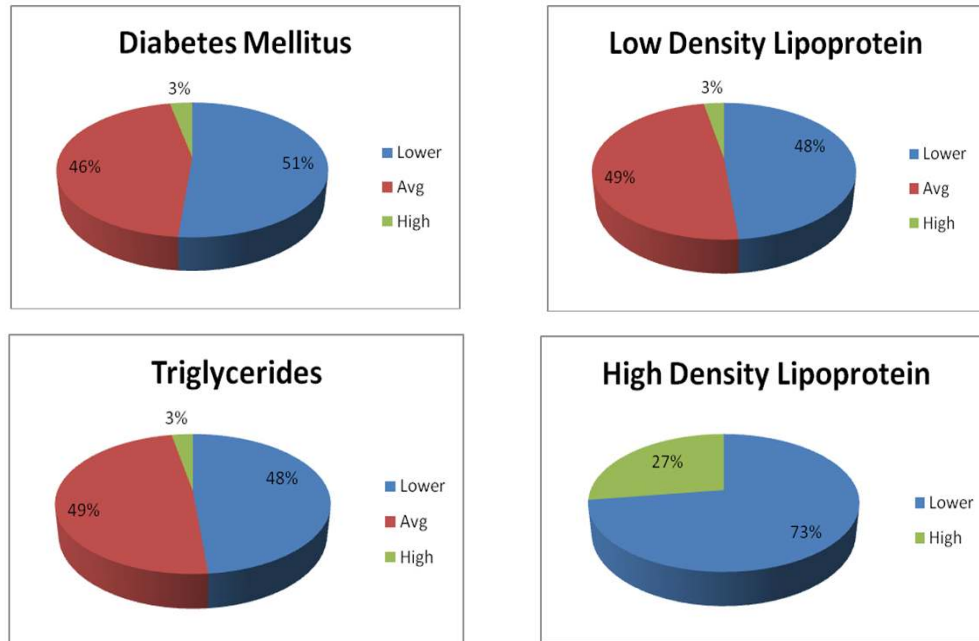


Fig. 1. Percentage wise patient distribution under lower, average and high level reference range

The Quantum Neural Network is trained with the selected significant patterns. The architecture of the quantum neural network used in this study is having 4 input nodes, 15 hidden nodes and 1 output nodes as shown in Fig. 2.

The Quantum Neural Network is trained with the above mentioned number of nodes and accuracy graph is shown in the Fig. 3 and its performance graph is shown in Fig. 4. The number of input nodes is determined by the clinical parameters, the number of hidden nodes is determined through tuning; and the number of output node is

represented as percentage level of cardiovascular disease.

The proposed work shows that the prediction of risk from cardiovascular diseases gives best results on the dataset used. The result generated by this system has been evaluated and validated on data of patients with the Doctor's diagnosis (predictions). When the QNN is trained and tested after optimizing the input parameters, the overall predictive accuracy obtained is 96.97%. Data Showing Prediction accuracy of Random Testing on different Experimental values has been given in Table 4.

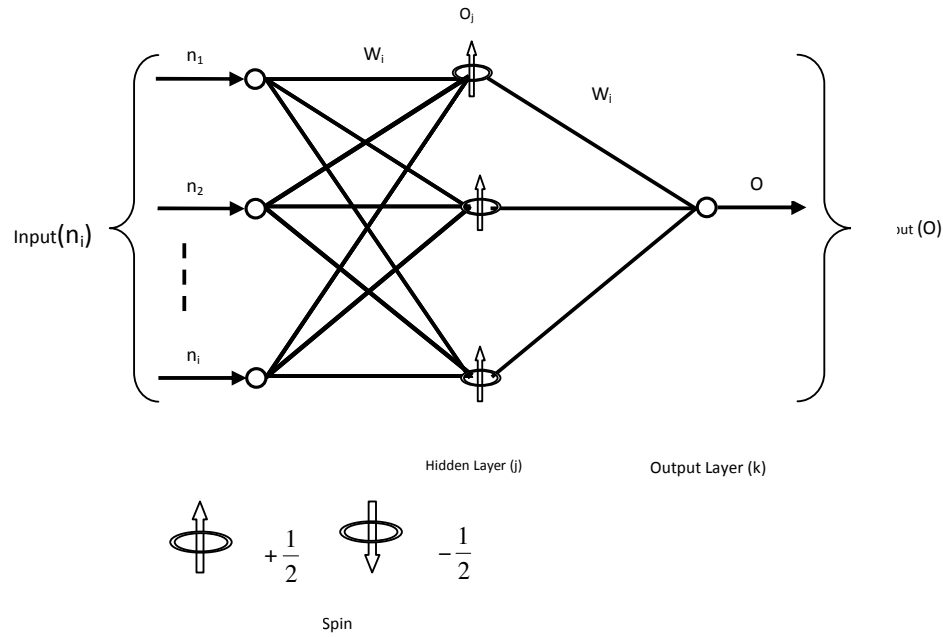


Fig. 2. Architecture diagram of quantum neural network for cardiovascular disease prediction system

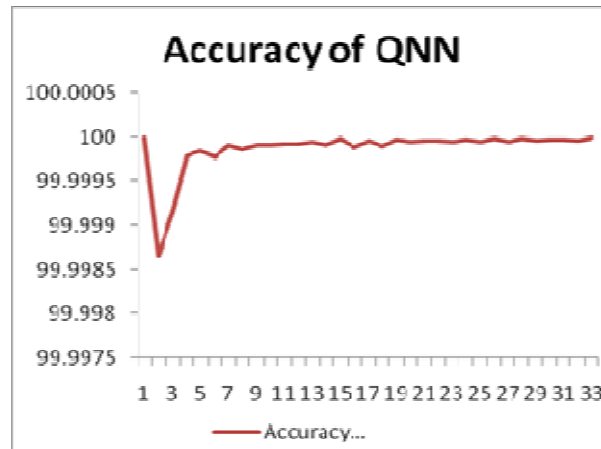


Fig. 3. Graph showing accuracy of CVD prediction system based on QNN

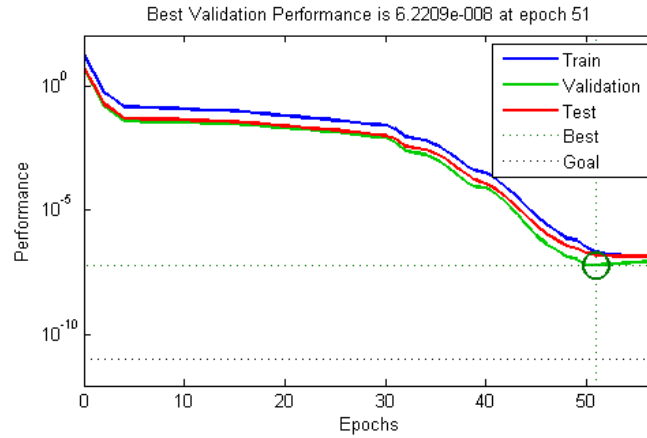


Fig. 4. Performance graph of quantum neural network

Table 4. Data showing prediction accuracy proposed on different experimental values

Diabetes	LDL	HDL	Triglycerides	Equated/ Calculated desired result		Desired result based on Doctor's diagnosis	Result and accuracy of proposed QNN based system	
				Output	Risk Cat.		Output	Risk Cat.
151.25	135.63	52.5	255.25	59.38	Int	Int	59.38	Int
121.25	111.25	72.5	183.38	34.38	Low	Low	34.38	Low
188.75	154.38	27.5	442.75	90.63	High	High	90.63	High
95	85	90	161.5	12.5	Nml	Nml	12.5	Nml
80	70	100	149	0	Nml	Nml	0.02	Nml
201	160	20	499	100	High	High	100	High
117.5	107.5	75	180.25	31.25	Low	Low	31.25	Low
143.75	131.88	57.5	217.75	53.13	Int	Int	53.13	Int
113.75	103.75	77.5	177.13	28.13	Low	Low	28.13	Low
91.25	81.25	92.5	158.38	9.38	Nml	Nml	9.38	Nml
196.25	158.13	22.5	480.25	96.88	High	High	96.88	High
139	130	60	199	50.45	Low	Low	51	Int
162.5	141.25	45	311.5	68.75	Int	Int	68.75	Int
83.75	73.75	97.5	152.13	3.13	Nml	Nml	3.13	Nml
192.5	156.25	25	461.5	93.75	High	High	93.75	High
106.25	96.25	82.5	170.88	21.88	Nml	Nml	21.88	Nml
128.75	118.75	67.5	189.63	40.63	Low	Low	40.63	Low
166.25	143.13	42.5	330.25	71.88	Int	Int	71.88	Int
181.25	150.63	32.5	405.25	84.38	High	High	84.38	High
87.5	77.5	95	155.25	6.25	Nml	Nml	6.26	Nml
185	152.5	30	424	87.5	High	High	87.5	High
155	137.5	50	274	62.5	Int	Int	62.5	Int
132.5	122.5	65	192.75	43.75	Low	Low	43.75	Low
125	115	70	186.5	37.5	Low	Low	37.5	Low
177.5	148.75	35	386.5	81.25	High	High	81.25	High
170	145	40	349	75	Int	Int	75	Int
98.75	88.75	87.5	164.63	15.63	Nml	Nml	15.63	Nml
136.25	126.25	62.5	195.88	46.88	Low	Low	46.87	Low
110	100	80	174	25	Nml	Nml	25	Nml
158.75	139.38	47.5	292.75	65.63	Int	Int	65.63	Int
173.75	146.88	37.5	367.75	78.13	High	High	78.13	High
102.5	92.5	85	167.75	18.75	Nml	Nml	18.75	Nml
147.5	133.75	55	236.5	56.25	Int	Int	56.25	Int

*Risk Cat.: Risk Category, Nml: Normal, Int: Intermediate, High: Highest

4. CONCLUSION

The proposed work shows that the prediction of risk from cardiovascular diseases gives best results based on clinical factors. The result generated by this system has been evaluated and validated on data of patients with the Doctor's diagnosis (predictions). This proposed system will help the doctors to plan for a better medication and provide the patient with early diagnosis as it performs reasonably well even without retraining. The QNN has been trained and tested after optimizing the input parameters. The optimization has been done by selecting minimum distorted data set of input parameters. The overall predictive accuracy obtained was 96.97%. Data Showing Prediction accuracy of Random Testing on different Experimental values has been given in (Table 4). Hence the proposed system will work as a significant tool for doctors/ practitioners.

ACKNOWLEDGEMENT

The authors would like to thank Dr. Praveen Garg, Cardiologist of Rameshwar Das Memorial Hospital, Jagadhri, Haryana, India for his valuable guidance and continued support throughout the work.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. WHO; 2013.
Available:<http://www.who.int/mediacentre/factsheets/fs317/en/index.html>
2. Mary Obenshain K. Application of data mining techniques to healthcare data. *Infection Control and Hospital Epidemiology*. 2004;25(8):690–695.
3. Pyorala K, Laakso M, Uusitupa M. Diabetes and atherosclerosis: An epidemiologic view. *Diabetes/Metab Rev*. 1987;3:463–524.
4. Krolewski AS, Kosinski EJ, Warram JH, et al. Magnitude and determinants of coronary artery disease in juvenileonset, insulin-dependent diabetes mellitus. *Am J Cardiol*. 1987;59:750–5.
5. Yudkin JS, Blauth C, Drury P, et al. Prevention and management of cardiovascular disease in patients with diabetes mellitus: An evidence base. *Diabet Med*. 1996;13:101–21.
6. Laakso M, Lehto S. Epidemiology of macrovascular disease in diabetes. *Diabetes Rev*. 1997;5:294–315.
7. Haffner SM, Miettinen H. Insulin resistance implications for type II diabetes mellitus and coronary heart disease. *Am J Med*. 1997;103:152–162.
8. Haffner SM. Management of dyslipidemia in adults with diabetes. *Diabetes Care*. 1998;21:160–78.
9. American Diabetes Association, Detection and management of lipid disorders in diabetes. *Diabetes Care*. 1993;16(2):106.
10. Head J, Fuller JH. International variations in mortality among diabetic patients: The Multinational Study of vascular Disease in Diabetics. *Diabetologia*. 1990;33:477– 481.
11. Brown MS, Goldstein JL. Receptor-mediated control of cholesterol metabolism. *Science*. 1976;191:150–154. DOI: 10.1126/science.174194
12. Steinberg D, Parthasarathy S, Carew S, Khoo JC, Witztum JL. Beyond cholesterol. Modifications of low-density lipoprotein that increase its atherogenicity. *N Engl J Med*. 1989;320:915–924.
13. Assmann G, Schulte H, von Eckardstein A, Huang Y. High-density lipoprotein cholesterol as a predictor of coronary heart disease risk: The PROCAM experience and pathophysiological implications for reverse cholesterol transport. *Atherosclerosis*. 1996;124:Suppl:S11–S20.
14. Curb JD, Abbott RD, Rodriguez BL, et al. A prospective study of HDL-C and cholesteryl ester transfer protein gene mutations and the risk of coronary heart disease in the elderly. *J Lipid Res*. 2004; 45:948–953.
15. Sharrett AR, Ballantyne CM, Coady SA, et al. Coronary heart disease prediction from lipoprotein cholesterol levels, triglycerides, lipoprotein(a), apolipoproteins A-I and B, and HDL density subfractions: The Atherosclerosis Risk in Communities (ARIC) Study. *Circulation*. 2001;104:1108–1113.
16. Gordon T, Castelli WP, Hjortland MC, Kannel WB, Dawber TR. High density lipoprotein as a protective factor against coronary heart disease: The Framingham Study. *Am J Med*. 1977;62:707–714
17. Turner RC, Millns H, Neil HA, et al. Risk factors for coronary artery disease in non-insulin dependent diabetes mellitus: United

- Kingdom Prospective Diabetes Study (UKPDS: 23). BMJ. 1998;316:823-828.
18. Arch Mainous G, Richelle Koopman J, Vanessa Diaz A, Charles Everett J, Peter Wilson WF, Barbara Tilley C. A coronary heart disease risk score based on patient reported information. Am J Cardiology. 2007;99(9):1236–1241.
19. Støvring H, Harmsen CG, Wisløff T, Jarbøl DE, Nexøe J, Nielsen JB, Kristiansen IS. Competing risk approach for the European Heart SCORE model based on cause-specific and all-cause mortality. Eur J PrevCardiol. 2013;20(5):827-36.
20. Bhopal Raj, Colin Fischbacher, Erkki Vartiainen, Nigel Unwin, Martin White, George Alberti. Predicted and observed cardiovascular disease in South Asians: Application of FINRISK, Framingham and SCORE models to Newcastle Heart Project data. Journal of Public Health. 2005;27(1):93–100.
21. Berto P, Munro V, Gaddi A, Negrini C, Hutton J, Mast O. Cost-effectiveness analysis for statin therapies in the primary prevention of coronary heart disease in Italy. Clin. Drug Invest. 2000;20:107-21.
22. Huse DM, Russell MW, Miller JD, Kraemer DF, D'Agostino RB, Ellison RC. Cost-effectiveness of statins. Am. J. Cardiol. 1998;82:1357-63.
23. Nyman JA, Martinson MS, Nelson D, Nugent S, Collins D, Wittes J, et al. Cost-effectiveness of gemfibrozil for coronary heart disease patients with low levels of high-density lipoprotein cholesterol: The Department of Veterans Affairs High-Density Lipoprotein Cholesterol Intervention Trial. Arch. Intern. Med. 2002; 162:177-82.
24. Weinstein MC, Coxson PG, Williams LW, Pass TM, Stason WB, Goldman L. Forecasting coronary heart disease incidence, mortality, and cost: The Coronary Heart Disease Policy Model. Am J Public Health. 1987;77:1417-1426.
25. Parthiban Latha, Subramanian R. Intelligent heart disease prediction system using CANFIS and genetic algorithm. International Journal of Biological, Biomedical and Medical Sciences. 2007; 3(3):157- 160.
26. Zhou J, Gan Q, Krzyzak A. Recognition of handwritten numerals by quantum neural network with fuzzy features. Intl. J. on document analysis and recogn. 1999;30-36.
27. Karayiannis NB, Mukherjee A, Glover JR. An evaluation of quantum neural networks in the detection of epileptic seizures in the neonatal electroencephalogram. Soft Computing. 2006;382-396.
28. Karayiannis NB, Mukherjee A, Glover JR. Detection of pseudosinusoidal epileptic segments in the neonatal EEG by cascading a rule – based algorithm with a neural network. IEEE Trans. on Biomed. Engg. 2006;633-641.
29. Narayan R, Chakraverty S, Singh VP. Quantum neural network based machine translator for Hindi to English. The scientific world Journal.2014;8. Article ID 485737. DOI: 10.1155/2014/485737
30. Chakraverty S, Gupta P, Sharma S. Neural network-based simulation for response identification of two-storey shear building subject to earthquake motion. J. of Neural Computing and App. 2010;3(19):367-375.
31. Narayan R, Chakraverty S, Singh VP. Machine translation using quantum neural network for simple sentences. Intl. J. of Info and Computation Techno. 2013;3(7) 683-690.

© 2017 Narain et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:
The peer review history for this paper can be accessed here:
<http://sciencedomain.org/review-history/19878>

APPENDICES

Patient Record

Name.....

Age.....Years

Gender: Male ☐ Female ☐

Address.....

Height.....cm

Weight.....kg

.....

Alcoholic: Yes ☐ No ☐

Smoking : Yes ☐ No ☐

Family History: Yes ☐ No ☐

Number of Family Members having Heart Disease :

History

.....
.....

ECG Finding.....

Blood Pressure.....

Triglycerides

Blood Glucose.....

Electrolytes

Complete Blood Count

Sodium

Red Blood Cell Count

Potassium

Hemoglobin

Bicarbonate

Hematocrit

Chloride

Mean Corpuscular Vol

Calcium

Mean Corpuscular HGB

SGOT (Aspartate transaminase)

Mean Corpuscular Hemoglobin-

CPK (MB)

-Concentration

hs- CRP

Red Cell Distribution Width

Diagnosis

Platelet Count

.....
.....

Mean Platelet Volume

.....
.....

White Blood Cell Count

.....
.....

Lipid Profile

Percentage of Chances of Heart Attack

Total Cholesterol

(Prediction).....

HDL

Any Other Comment

LDL

.....
.....
.....