ORIGINAL RESEARCH

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

Renu Narain[1]
Sanjai Saxena[1]
Achal Kumar Goyal[2]

[1]Department of Biotechnology, Thapar University, Punjab, India; [2]University 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 The commonly used Framingham risk score (FRS) for CVD prediction is outdated for the modern population, so FRS may not be accurate enough.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 sanjaibiotech@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 |

**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.

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:** [a]Systolic/diastolic blood pressure range (mmHg).

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 Indrayan[32] 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
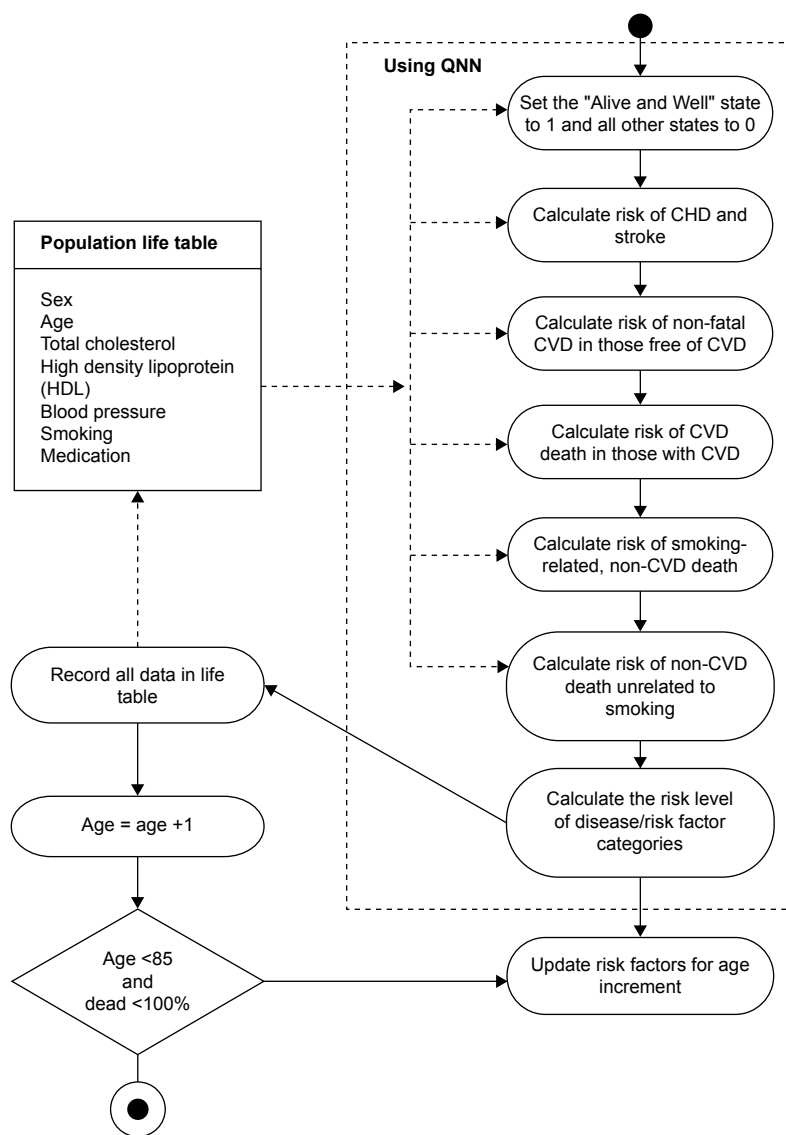
**Figure 1** Activity diagram of proposed system.
**Abbreviations:** QNN, quantum neural network; CHD, coronary heart disease; CVD, cardiovascular disease.

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 $\theta^r$ 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, $\eta$ is learning rate which is a small random value, $\delta_k$ is error rate of output layer, and $\delta_j$ 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 × 1) is input and $t_i$ (K × 1) is target values for given inputs, as shown in Figure 2.

The error signal term of the output $(\delta_k)$ and hidden layers $(\delta_j)$ are written as,

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

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

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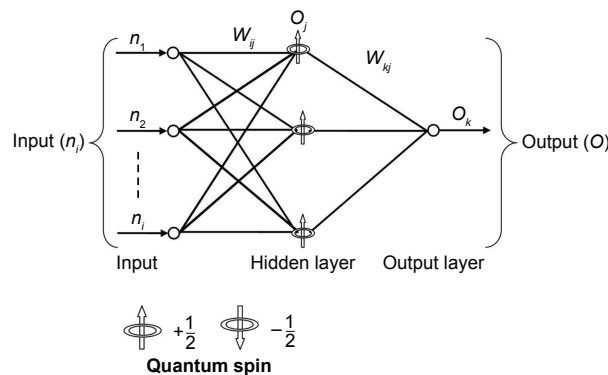


**Figure 2** Architecture of quantum neural network.
**Notes:** $n_i$ denotes the input to the input layer; $O_j$ and $O_k$ denote the output of hidden and output layer, respectively. The weights between input and hidden layers are denoted by $W_{ij}$ and the weights between hidden and output layers are denoted by $W_{kj}$.

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

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

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

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

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:

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

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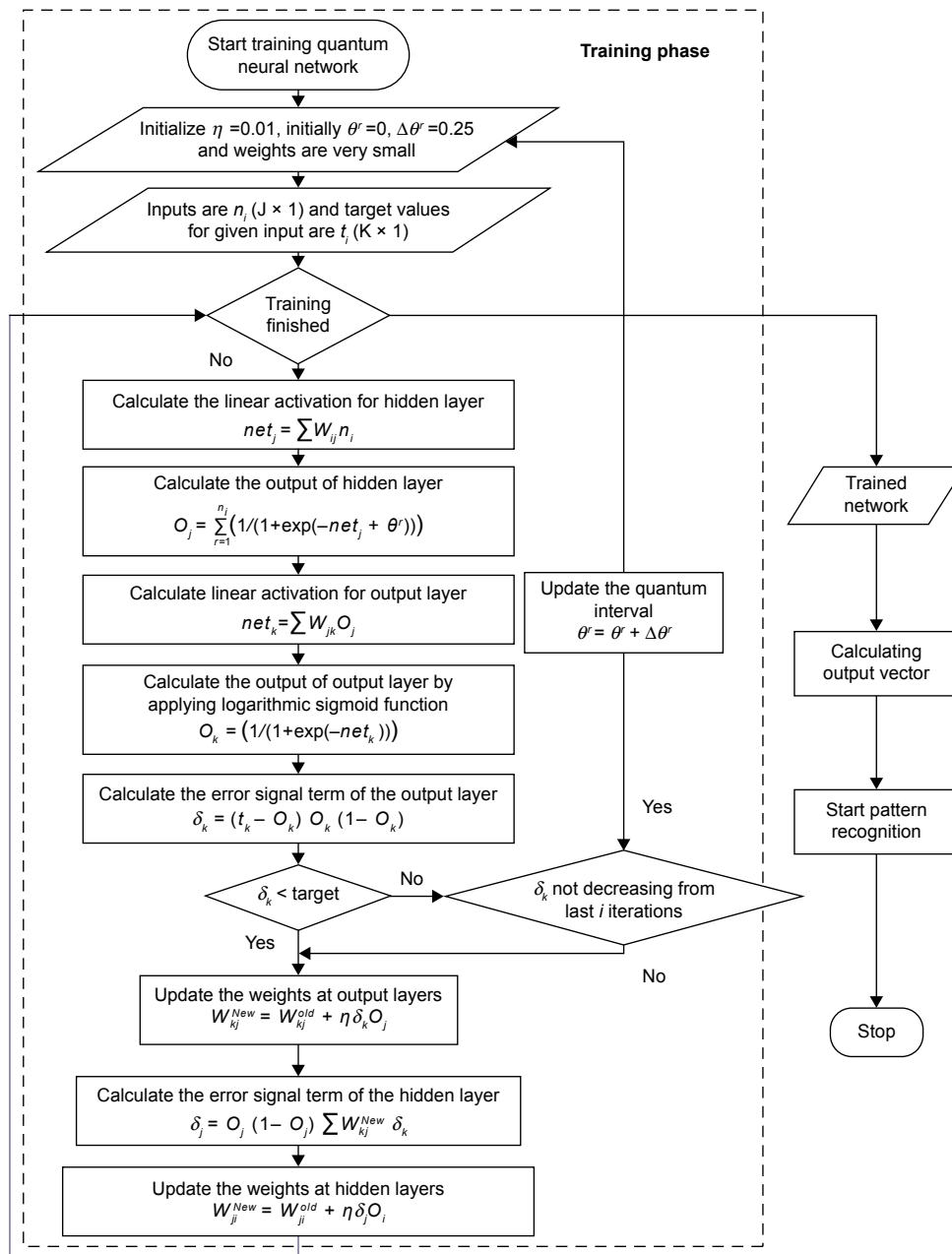
**Figure 3** Flowchart of quantum neural network for heart disease prediction system.

**Notes:** $\theta^r$ denotes quantum interval between sub-states, with the difference of quantum level $r$. $n_s$ denotes the number of grades or excitation levels, $\eta$ is learning rate, $\delta_k$ is error rate of output layer, and $\delta_j$ 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

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

| Age | Sex | TC | HDL | | SBP | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 30 | Female | 4.74 | 1.94 | No | 115 | No | 31.25% | 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 | 93.73% | Highest risk | 99.98 | | 40.08 | 42.75 |
| 35 | Female | 5.28 | 1.68 | No | 120 | Yes | 43.75% | Low risk | 43.75% | Low risk | 100.00 | | 0.91 | 2.07 |
| 60 | Female | 4.45 | 2.07 | No | 130 | Yes | 25.00% | Normal | 25.01% | Normal | 100.04 | 99.99 | 5.17 | 20.69 |
| 55 | Female | 3.91 | 2.33 | No | 125 | Yes | 12.50% | Normal | 12.50% | Normal | 100.00 | | 2.52 | 20.14 |
| 60 | Male | 6.92 | 0.91 | Yes | 160 | Yes | 81.25% | 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 | 31.21% | Low risk | 99.87 | | 1.57 | 5.02 |
| 50 | Male | 4.45 | 2.07 | No | 123 | Yes | 25.00% | Normal | 25.00% | Normal | 100.00 | | 3.93 | 15.71 |
| 35 | Female | 3.65 | 2.46 | No | 110 | Yes | 6.25% | Normal | 6.25% | Normal | 100.00 | | 0.17 | 2.75 |
| 25 | Female | 4.20 | 2.20 | No | 110 | Yes | 18.75% | Normal | 18.75% | Normal | 100.00 | | 0.04 | 0.24 |
| 20 | Female | 4.20 | 2.20 | No | 110 | Yes | 18.75% | Normal | 18.75% | Normal | 100.00 | 99.28 | 0.01 | 0.06 |
| 50 | Female | 4.45 | 2.07 | No | 125 | Yes | 25.00% | Normal | 24.01% | Normal | 96.04 | | 2.47 | 9.90 |
| 40 | Female | 6.37 | 1.17 | No | 129 | Yes | 68.75% | 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 | 81.26% | Highest risk | 100.01 | | 36.06 | 44.38 |
| 55 | Male | 3.91 | 2.33 | No | 121 | Yes | 12.50% | Normal | 12.50% | Normal | 100.00 | | 4.10 | 32.78 |
| 25 | Female | 6.37 | 1.17 | No | 125 | Yes | 68.75% | Intermediate risk | 68.00% | Intermediate risk | 98.91 | | 0.60 | 0.87 |
| 40 | Female | 4.45 | 2.07 | No | 115 | Yes | 25.00% | Normal | 25.00% | Normal | 100.00 | | 0.75 | 2.99 |
| 30 | Female | 5.83 | 1.42 | No | 120 | Yes | 56.25% | Intermediate risk | 56.25% | Intermediate risk | 100.00 | 99.97 | 0.71 | 1.26 |
| 60 | Male | 5.54 | 1.55 | No | 134 | Yes | 50.00% | 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 | 81.26% | Highest risk | 100.01 | | 30.24 | 37.22 |
| 50 | Female | 6.63 | 1.04 | No | 180 | Yes | 75.00% | Intermediate risk | 75.00% | Intermediate risk | 100.00 | | 11.20 | 14.94 |
| 35 | Male | 4.20 | 2.20 | No | 116 | Yes | 18.75% | Normal | 18.71% | Normal | 99.79 | | 0.46 | 2.46 |
| 55 | Female | 4.45 | 2.07 | No | 124 | Yes | 25.00% | Normal | 25.00% | Normal | 100.00 | | 3.36 | 13.46 |
| 35 | Female | 6.63 | 1.04 | No | 160 | Yes | 75.00% | Intermediate risk | 75.00% | Intermediate risk | 100.00 | 99.98 | 2.87 | 3.82 |
| 20 | Male | 5.83 | 1.42 | Yes | 122 | Yes | 56.25% | Intermediate risk | 56.22% | Intermediate risk | 99.95 | | 0.28 | 0.49 |
| 55 | Female | 6.63 | 1.04 | No | 140 | No | 75.00% | 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 | 81.26% | Highest risk | 100.01 | | 56.33 | 69.33 |
| 40 | Female | 5.54 | 1.55 | No | 126 | Yes | 50.00% | 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 | 31.23% | Low risk | 99.94 | | 1.72 | 5.50 |
| 25 | Female | 4.74 | 1.94 | No | 115 | Yes | 31.25% | Low risk | 31.25% | Low risk | 100.00 | | 0.09 | 0.30 |
| 30 | Male | 3.65 | 2.46 | Yes | 112 | No | 6.25% | Normal | 6.25% | Normal | 100.00 | | 0.36 | 5.75 |

Accuracy column values: 17.02, 13.85, 14.90, 15.05 (aligned with 99.99, 99.28, 99.97, 99.98 rows respectively).

**Abbreviations:** TC, total cholesterol; HDL, high density lipoprotein; SBP, systolic blood pressure; QNN, quantum neural network; FRS, Framingham risk score.
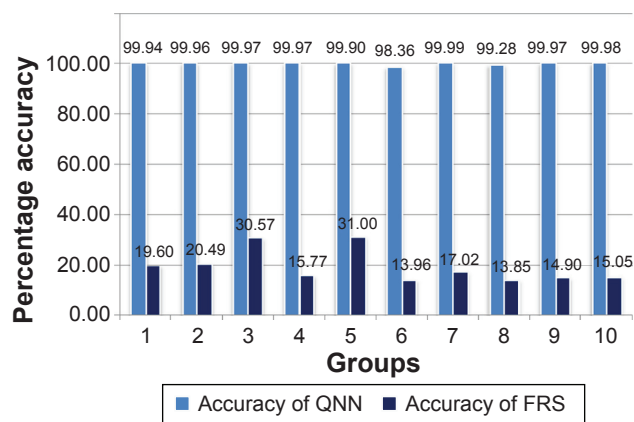
**Figure 4** Graph showing accuracy of proposed system and Framingham risk score (FRS).
**Abbreviation:** QNN, quantum neural network.

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:** [a]Predicted 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

## Disclosure

The authors report no conflicts of interest in this work.

## References

1. 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.
2. 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.
3. Khatibi V, Montazer GA. A fuzzy-evidential hybrid inference engine for coronary heart disease risk assessment. *Expert Syst Appl*. 2010; 37(12):8536–8542.
4. Akay M. Noninvasive diagnosis of coronary artery disease using a neural network algorithm. *Biol Cybern*. 1992;67(4):361–367.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. Lloyd-Jones DM, Larson MG, Beiser A, Levy D. Lifetime risk of developing coronary heart disease. *Lancet*. 1999;353(9147):89–92.
11. 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.
12. Pilote L, Hlatky MA. Attitudes of women toward hormone therapy and prevention of heart disease. *Am Heart J*. 1995;129(6):1237–1238.
13. Ezzati M, Hoorn 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.
14. 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.
15. 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.
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(5):707–714.
17. 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.
18. Kannel WB. Blood pressure as a cardiovascular risk factor: prevention and treatment. *JAMA*. 1996;275(20):1571–1576.
19. 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.
20. Anderson KM, Wilson PW, Odell PM, Kannel WB, An updated coronary risk profile. A statement for health professionals. *Circulation*. 1991; 83(1):356–362.
21. 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.
22. American Diabetes Association. Detection and management of lipid disorders in diabetes. *Diabetes Care*. 1993;16(5):828–834.
23. 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.
24. 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.
25. 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.
26. 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.
27. 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.
28. 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.
29. Bartechi CE, MacKenzie TD, Schrier RW. The human costs of tobacco use. *N Engl J Med*. 1994;330:907–912.
30. MacKenzie TD, Bartechi CE, Schrier RW. The human costs of tobacco use. *N Engl J Med*. 1994;330:975–980.
31. Manson JE, Tosteson H, Ridker PM, et al. The primary prevention of myocardial infarction. *N Engl J Med*. 1992;326(21):1406–1416.
32. Indrayan A. Forecasting vascular disease cases and associated mortality in India. *NCMH Background Papers-Burden of disease in India*. 2005;(9): 198–215.
33. 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.