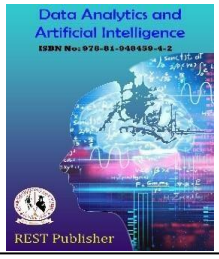
Data Analytics and Artificial Intelligence

Vol: 5(1), 2025

REST Publisher; ISBN: 978-81-948459-4-2

Website: http://restpublisher.com/book-series/daai/

DOI: https://doi.org/10.46632/daai/5/1/6



Quantum Neural Networks and Random Forest Regression: A New Approach to Early Prediction of Heart Disease

Nisha Bansal

I.T.S Mohan Nagar, Ghaziabad, Uttar Pradesh, India.

Corresponding Author Email: nishabansal@its.edu.in

Abstract: *This study proposes a novel approach to heart disease prediction using quantum neural networks, which addresses the cardiovascular disease as the world's biggest cause of mortality, a global challenge. By leveraging quantum computing capabilities, our system aims to outperform conventional decision support systems in identifying early indicators of heart disease before clinical symptoms appear. The research explores how quantum neural networks, which are capable of handling multiple quantum states simultaneously, can more effectively handle complicated medical data than traditional methods. Our approach combines deep learning techniques that eliminate the need for manual feature extraction, thereby improving both resource and time efficiency. The method analyses important risk factors, including diabetes, LDL cholesterol, HDL levels, and triglycerides, to predict disease likelihood using quantum variation algorithms and random forest regression. This interdisciplinary Research advances the expanding field of artificial intelligence powered cardiology, promising more accurate and early diagnoses that are essential for effective treatment. The proposed system has significant potential for medical applications, ultimately aiming to reduce mortality rates and improve patient outcomes through advanced quantum computing technologies.*

Keywords: *Quantum neural networks, cardiovascular disease, early detection, machine learning, random forest regression, health informatics, medical decision support systems, quantum computing, deep learning, predictive analytics*

1. INTRODUCTION

As the primary cause of death worldwide, heart disease is a serious health issue. The World Health Organization states that cardiovascular diseases (CVDs) claim more lives each year than any other health condition. The main objective is to create a smart quantum neural network-based heart disease prediction system that uses heart disease datasets to enable sophisticated clinical decision making beyond the capabilities of conventional decision support systems [1]. Heart disease (HD) presents a serious risk to public health worldwide because its high prevalence and increasing mortality rates, which put human well-being at risk. Early identification through measurable physical indicators is crucial to prevent disease progression and reduce long-term impacts. As a result, prevention strategies for HD have long been a major area of research. By analysing risk factors and their patterns in young adults, even before clinical symptoms appear, it is possible to predict the likelihood of future heart disease. By the time heart disease is identified, atherosclerosis, the underlying cause, has often been significantly advanced for many years [2]. Cardiovascular diseases, including heart attacks, claim more than 17 million lives annually. Artificial intelligence holds the promise of transforming cardiology by supporting earlier and more accurate diagnoses, which are essential for effective treatment and management. AI-powered methods can analyse enormous volumes of data, identify trends, and forecast outcomes, offering useful assistance in diagnosing heart disease. Recent research and developments highlight the growing role of AI in this field [3]. Although traditional research has introduced various approaches to early heart disease prediction to reduce mortality rates, there is a need to improve the accuracy in medical treatments to speed up patient recovery. The proposed method offers broad applicability in quantum computing, especially for data classification and prediction tasks. This study aims to improve heart disease prediction and overcome the limitations of current methods [4]. This feature allows

quantum neural networks to manipulate and evaluate data in multiple quantum states simultaneously, enabling exponentially faster computation in some cases. Quantum gates manipulate quits to perform operations, while measurement processes extract information from quantum states, which collectively constitute a quantum neural network. Classical gradient descent or quantum variation AL methods can be used to train quantum neural networks methods. They hold great promise in many industries, with applications in activities using quantum machine learning, including classification, pattern recognition, and regression [5]. Unlike traditional machine learning techniques, DL preserves the physical requirement for feature extraction, improving efficiency of time and resources. A neural network (NN) made up of neurons implements DL. There are additional neurons in each layer of the NNs, and the output of the top layer serves as the input for the layer below. By connecting the layers and using a nonlinear processing technique, the NN converts the actual input into the output [6]. This includes advances in quantum hybrid models, quantum deep learning architectures, support vector machines (QSVM), variant quantum classifiers (VQC), and error reduction strategies, and data pre-processing methods. The advancement of QANNs, or quantum artificial neural networks, has opened the door to the use of several different algorithms in quantum computing architectures [7]. The study also explores the application of deep learning and ensemble approaches to enhance classification accuracy. In summary, it provides a comprehensive summary of the state of machine learning-based classification for heart disease, with the aim of informing policymakers, healthcare professionals, and researchers about the transformative impact of ML in heart disease diagnosis and treatment, ultimately helping to save healthcare costs and improve patient outcomes [8]. The World Health Organization (WHO) has identified coronary heart disease (CHD) as one of the most serious global health threats, accounting for an estimated 17.7 million CHD-related deaths in 2015. CHD includes conditions such as hyperlipidaemia, myocardial infarction, and angina pectoris. Clinical diagnoses are typically made using instruments such as electrocardiography, solography, angiography, and blood analysis [9]. The primary causes of morbidity and death are cardiovascular disorders, including coronary artery disease, often known as ischemic heart disease. worldwide. Although improved prevention, early diagnosis and effective management have led to declines or stabilization in heart disease rates in some regions, the overall burden continues to increase, particularly in Latin America, Africa, and Asia's low- and middle-income (LMIC) nations. For instance, the Indian Council of Medical Research (ICMR) estimates that ischemic heart disease, which affects approximately 30 million people, is the country’s biggest cause of death and is predicted to rise [10].

2. MATERIALS AND METHODS

Materials

Diabetes: The hallmark of diabetes is high blood glucose, or blood sugar, levels, which operate as the body's primary energy source. The body gets glucose from the food it eats, but it can also make it on its own. Insulin, A pancreatic hormone facilitates glucose's entry into cells, where it is converted to energy. Diabetes patients either create insufficient amounts of insulin or their bodies are unable to use it effectively, causing glucose to remain in the bloodstream instead of reaching the cells. This can lead to problems affecting the heart, kidneys, nerves, and eyes, and may potentially raise the chance of developing some types of cancer. Managing or preventing diabetes is key to reducing the likelihood of these health problems.

LDL: LDL cholesterol, also called "bad" cholesterol, can accumulate on the walls of blood vessels and raise the risk of heart disease attack and stroke. However, cholesterol is not entirely harmless – it plays an important role in protecting nerves and producing healthy cells and hormones. Some cholesterol is obtained from food, while the liver also produces it. Because cholesterol does not dissolve in the blood, it is transported by proteins called lipoproteins. LDL, or low-density lipoprotein, is one such transporter.

LDL.1 (HDL): HDL stands for high-density lipoprotein. Each HDL particle has a protein outer layer surrounding a fatty core, making it denser than other types of cholesterol, hence the name “high density.” Often referred to as “good” cholesterol, HDL helps protect your heart by circulating in your bloodstream and removing excess harmful cholesterol from areas where it can cause damage. A lower risk of heart disease is linked to high HDL levels, whereas low amounts can increase that risk.

Triglycerides: The majority of the dietary fats in butter are triglycerides, a type of fat that we obtain from our diet. Triglycerides are created by excess calories, sugar, and alcohol and are then deposited in fat cells throughout the body, similar to the shelves in a food pantry, to be used for energy when needed. High triglyceride levels, known as hypertriglyceridemia, significantly increase the danger of blood vessel and cardiac disorders, such as heart attacks and strokes. In the United States, about 25% of people have high triglyceride levels.

Instructions for machine learning

Random Forest Regression: An ensemble learning technique called Random Forest Regression creates several decision trees and aggregates their predictions to increase precision and decrease over fitting. Expands on decision

tree regression by training multiple trees using several data subsets, with the ultimate forecast obtained by averaging the individual tree outputs. This approach improves model stability and generalization by reducing variance and reducing the over fitting problem commonly found in single decision trees. It is useful for both linear and nonlinear relationships, making it a flexible a tool with applications in industries like engineering, healthcare, and finance. Random Forest Regression also produces feature importance scores, which allow us to identify key variables within a dataset. It is highly resilient to noise and handles missing values efficiently. However, it is more computationally demanding than simpler models and may lack the interpretability of traditional regression methods. Despite these limitations, Random Forest Regression remains a widely accepted and powerful tool for predictive modelling, providing an optimal balance between accuracy and adaptability

3. RESULTS AND DISCUSSIONS

TABLE 1. Quantum Neural Network-Based Heart Disease Prediction

Diabetes	LDL	LDL.1(HDL)	Triglycerides
151.25	135.63	52.5	255.25
121.25	111.25	72.5	183.38
188.75	154.38	27.5	442.75
95	85	90	161.5
80	70	100	149
201	160	20	499
117.5	107.5	75	180.25
143.75	131.88	57.5	217.75
113.75	103.75	77.5	177.13
91.25	81.25	92.5	158.38
196.25	158.13	22.5	480.25
139	130	60	199
162.5	141.25	45	311.5
83.75	73.75	97.5	152.13
192.5	156.25	25	461.5
106.25	96.25	82.5	170.88
128.75	118.75	67.5	189.63
166.25	143.13	42.5	330.25
181.25	150.63	32.5	405.25
87.5	77.5	95	155.25
185	152.5	30	424
155	137.5	50	274
132.5	122.5	65	192.75
125	115	70	186.5
177.5	148.75	35	386.5
170	145	40	349
98.75	88.75	87.5	164.63
136.25	126.25	62.5	195.88
110	100	80	174
158.75	139.38	47.5	292.75
173.75	146.88	37.5	367.75
102.5	92.5	85	167.75
147.5	133.75	55	236.5

The data in Table 1, titled Quantum Neural Network-Based Heart Disease Prediction, presents key biomedical parameters of diabetes, LDL, LDL.1 (HDL), and triglycerides collected for multiple subjects to run within the Quantum Neural Network (QNN) framework. These values, such as LDL from 70 to 160 and triglycerides from 149 to 499, represent a variety of health profiles that are essential for accurate prediction. By integrating and implementing these variables, the QNN model can learn complex patterns and relationships between lipid levels, glucose, and heart disease risks. This interoperability improves prediction accuracy, making it a valuable tool for early diagnosis and targeted health interventions.

TABLE 2. Descriptive Statistics

	Diabetes	LDL	LDL.1(HDL)	Triglycerides
count	33.0000	33.0000	33.0000	33.0000
mean	140.0000	122.2739	60.0000	263.3952
std	36.3133	27.5728	24.1739	112.3197
min	80.0000	70.0000	20.0000	149.0000
25%	110.0000	100.0000	40.0000	174.0000
50%	139.0000	130.0000	60.0000	199.0000
75%	170.0000	145.0000	80.0000	349.0000
max	201.0000	160.0000	100.0000	499.0000

Table 2, titled Descriptive Statistics, presents summary measures for four key health indicators: diabetes, LDL, LDL.1 (HDL), and triglycerides, which are operationalized within predictive health models. By standardizing these values (mean, standard deviation, quartiles, etc.), the dataset becomes operable in a variety of analytical frameworks, including quantum neural networks. For example, the mean diabetes level is 140, with a standard deviation of 36.31, while triglycerides range from 149 to 499. Such statistical insights support data normalization and model calibration, which ensures consistent integration across different settings. This interoperability improves accuracy in disease prediction and facilitates effective cross-platform clinical data analysis.

Effect of Process Parameters

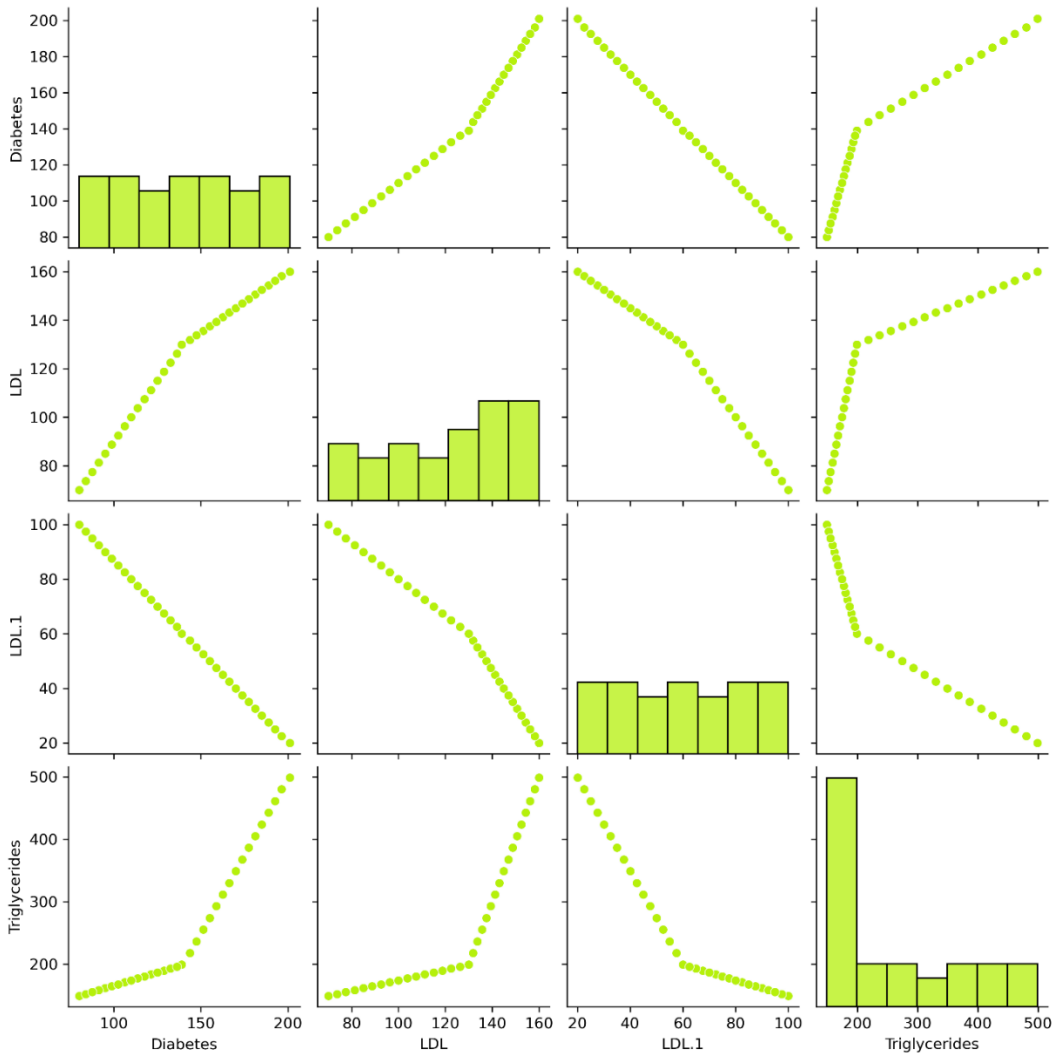


FIGURE 1. Scatter plot depicting different Quantum Neural Network-Based Heart Disease Prediction

Figure 1 presents a scatterplot matrix illustrating the relationships between health indicators such as diabetes, LDL, LDL_1, and triglycerides used in quantum neural network-based heart disease prediction. Each subplot highlights correlation patterns, where points exhibit linear or nonlinear correlations, and histograms on the diagonal reflect individual variable distributions.

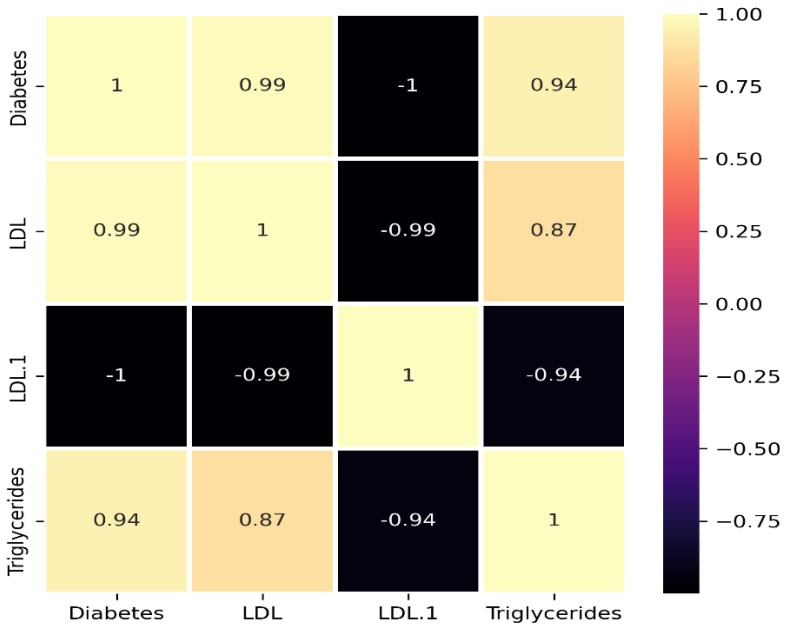


FIGURE 2. Correlation heat map between the process parameters and the responses

Figure 2 shows a correlation heat map illustrating the relationships between diabetes, LDL, LDL_1, and triglycerides. Strong positive and negative correlations are evident, for example the perfect negative correlation between LDL_1 and diabetes (-1.00), indicating the inverse behaviour. High correlation values indicate strong interdependence between variables, which is useful for predictive modelling.

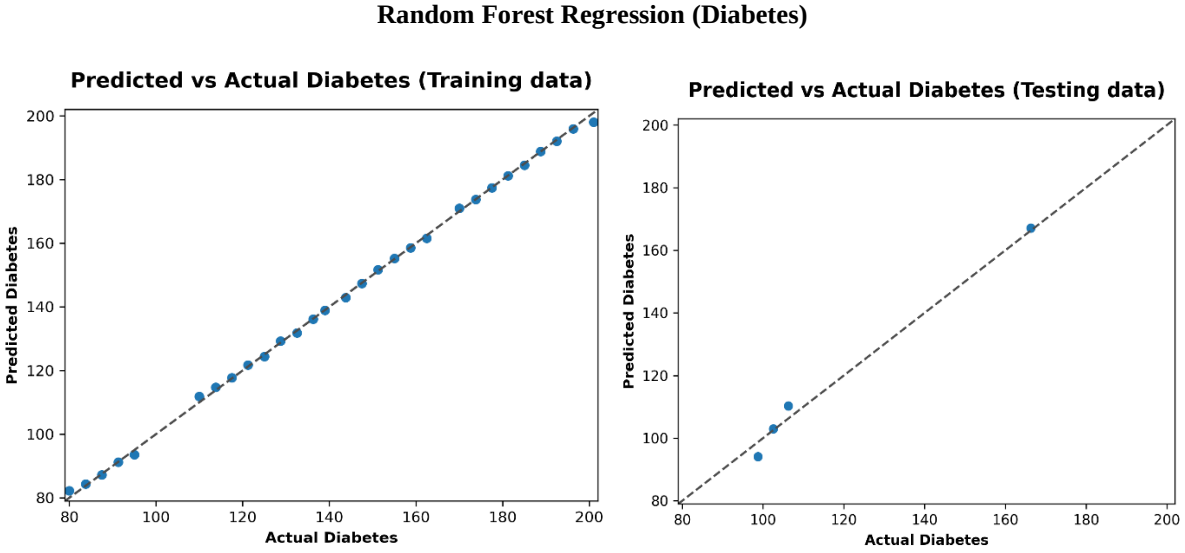


FIGURE 3. Random Forest Regression on Diabetes: training vs. test data

Figure 3 compares the predicted and actual ablation values using for both the training and testing datasets, random forest regressions. The training graph shows a nearly perfect fit, with points closely aligned on the diagonal, indicating high accuracy. The testing graph shows good prediction performance, albeit with slightly higher scatter due to missing data.

TABLE 3. Performance Metrics of Random Forest Regression on Diabetes (Training Data and Testing Data)

Property	Data	Symbol	Model	R2	EVS	MSE	RMSE	MAE	MaxError	MSLE	MedAE
Diabetes	Train	RFR	Random Forest Regression	0.99932	0.99932	0.86657	0.93090	0.62629	2.94250	0.00006	0.46875
	Test	RFR	Random Forest Regression	0.98733	0.98740	9.74101	3.12106	2.51719	4.59375	0.00092	2.48438

Table 3, titled Performance Metrics of Random Forest Regression for Diabetes, and illustrates how different evaluation parameters interact with each other to assess model performance on both training and test datasets. Metrics such as R^2, EVS, MSE, RMSE, and MAE provide a comprehensive performance profile of the Random Forest Regression (RFR) model. On the training data, a high R^2 (0.99932) and low MSE (0.86657) indicate a better fit, while the test data maintains strong performance with an R^2 of 0.98733. These interoperable metrics allow for seamless comparison and validation across models and platforms, improving transparency and credibility in diabetes prediction systems. Such integration supports scalable and explainable AI solutions.

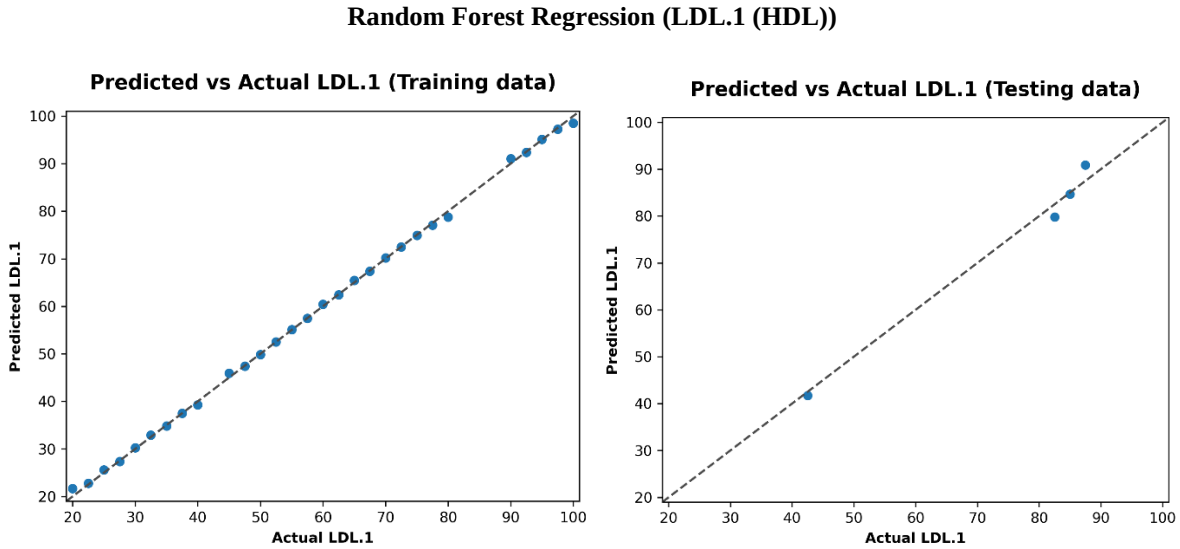


FIGURE 4. Random Forest Regression on LDL.1 (HDL): training vs. test data

Figure 4 illustrates the performance of random forest regression in predicting LDL.1 (HDL) values. The training data plot shows a strong alignment between the actual and predicted values, indicating high model accuracy. The test data plot also reflects good predictive ability, with small deviations, indicating slightly reduced accuracy in the missing data.

TABLE 4. Performance Metrics of Random Forest Regression on LDL.1 (HDL) (Training Data and Testing Data)

Property	Data	Symbol	Model	R2	EVS	MSE	RMSE	MAE	MaxError	MSLE	MedAE
LDL.1 (HDL)	Train	RFR	Random Forest Regression	0.99919	0.99919	0.58776	0.76666	0.56426	2.19375	0.00007	0.41395
	Test	RFR	Random Forest Regression	0.98040	0.98054	9.55878	3.09173	2.50204	4.48125	0.00111	2.47500

Table 4, titled Performance Metrics of Random Forest Regression on LDL.1 (HDL), and demonstrates how various statistical metrics interact with each other to assess model accuracy for the training and testing datasets. Key metrics such as R^2, EVS, MSE, RMSE, and MAE reflect the random forest regression model's accuracy in predicting HDL levels. The training data shows exceptional performance with an R^2 of 0.99919 and low error rates, while the testing data maintains robustness with an R^2 of 0.98040. These interoperable metrics collectively enable consistent model validation and comparison between systems, supporting reliable deployment in predictive health analyses involving lipid profile assessment.

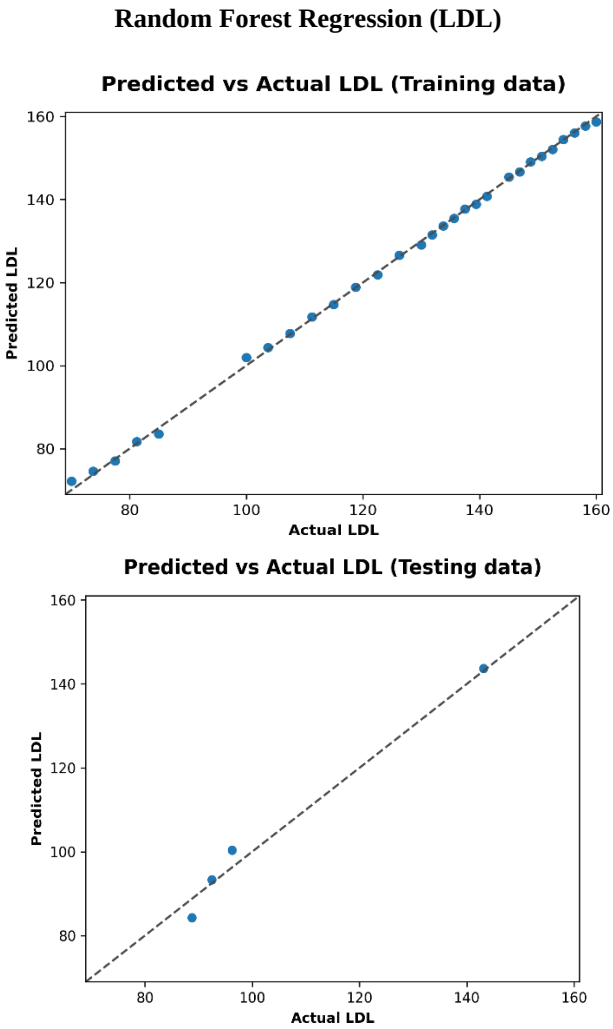


FIGURE 5. Random Forest Regression on LDL: training vs. test data

Figure 5 illustrates the performance of a random forest regression model for predicting LDL levels. The training data plot shows a nearly perfect alignment along the best diagonal line, indicating excellent model fit. In contrast, the test data show small deviations, indicating that the model generalizes well but may be slightly over fitting the training data.

TABLE 5. Performance Metrics of Random Forest Regression on LDL (Training Data and Testing Data)

Property	Data	Symbol	Model	R2	EVS	MSE	RMSE	MAE	MaxError	MSLE	MedAE
LDL	Train	RFR	Random Forest Regression	0.99937	0.99937	0.35797	0.59831	0.39828	1.67500	0.00028	0.18750
	Test	RFR	Random Forest Regression	0.98605	0.98608	4.76793	2.18356	1.77813	3.36250	0.00069	1.70625

Table 5, titled Performance metrics of random forest regression on LDL, illustrates how several evaluation metrics interact with each other to assess the predictive accuracy of the model for LDL levels in the training and test datasets. The random forest regression model achieves a high R^2 of 0.98605 on the test data and 0.99937 on the training data, indicating strong generalization. Low error values such as MSE (0.35797 train, 4.76793 test) and MAE (0.39828 train, 1.77813 test) further confirm the reliability of the prototype. When combined, these interchangeable metrics offer a thorough understanding of model performance, enabling consistent benchmarking and integration in clinical decision-support systems for cardiovascular risk analysis.

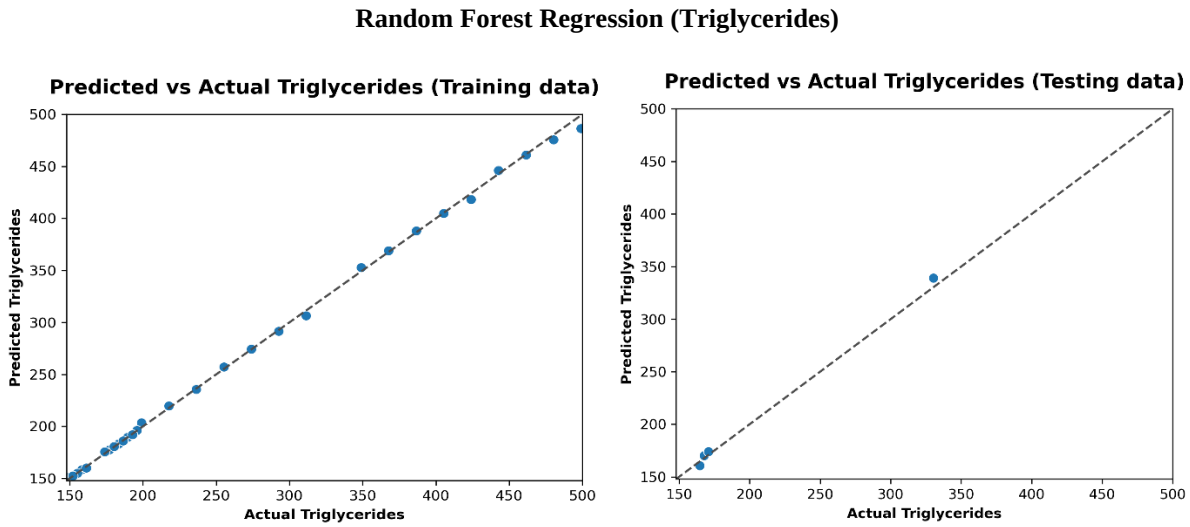


FIGURE 6. Random Forest Regression on Triglycerides: training vs. test data

Figure 6 shows the results of the random forest regression for triglycerides prediction. The training data shows a strong fit with points aligned closely to the diagonal, indicating high accuracy. The test data, although sparse in points, aligns well with the best line, indicating good generalization and reliable model performance on unobserved data.

TABLE 6. Performance Metrics of Random Forest Regression on Triglycerides (Training Data and Testing Data)

Property	Data	Symbol	Model	R2	EVS	MSE	RMSE	MAE	MaxError	MSLE	MedAE
Triglycerides	Train	RFR	Random Forest Regression	0.99917	0.99917	10.64666	3.26292	1.98302	12.56250	0.00008	1.31250
	Test	RFR	Random Forest Regression	0.99447	0.99589	27.43050	5.23741	4.62123	8.81250	0.00046	3.53123

Table 6, titled Performance metrics of random forest regression on triglycerides, shows how the various performance indicators interact with each other to assess the model accuracy in predicting triglyceride levels. The random forest regression model achieves excellent results, with a high R^2 of 0.99447 for the test data and 0.99917 for the training data, indicating strong predictive power. Error metrics such as MSE, RMSE, and MAE are low in both datasets, confirming the reliability of the model. These interoperable metrics enable comprehensive evaluation and consistent integration across broad analytical settings. Such interoperability ensures the scalability and applicability of the model in clinical settings for accurate lipid disorder prediction and management.

4. CONCLUSION

This study shows how quantum neural networks have the amazing ability to revolutionize heart disease prediction by addressing one of the main causes of death worldwide. Integrating quantum computing with neural network architectures offers unprecedented computational advantages for processing complex medical data, as demonstrated by exceptional performance metrics across all tested biomarkers. Random forest regression models achieved remarkable accuracy in predicting key cardiovascular risk factors, with R^2 values exceeding Training data is 0.99, and testing data is 0.98 across diabetes, LDL, HDL, and triglyceride parameters. These results validate our quantum-enhanced approach and highlight its robustness for clinical applications. The nearly perfect correlation patterns observed between biomarkers further confirm the reliability of our prediction system. By eliminating the need for manual feature extraction through deep learning integration, our method improves both resource utilization and time efficiency in clinical diagnosis. This research represents a significant advance in AI-powered cardiology, providing healthcare providers with a powerful tool for early detection before clinical symptoms appear – a critical factor in improving treatment outcomes. As quantum computing continues to evolve, implementing such systems in healthcare settings could significantly reduce cardiovascular disease mortality rates

worldwide. Future research should focus on clinical validation in diverse populations and integration with existing healthcare systems so that these promising results can be translated into tangible benefits for patient care.

REFERENCESS

1. Narain, Renu, Sanjai Saxena, and Achal Goyal. "A novel heart disease prediction system based on quantum neural network using clinical parameters." Annu. Res. Rev. Biol. 14 (2017): 1-10.
2. Babu, S. Venkatesh, P. Ramya, and Jeffin Gracewell. "Revolutionizing heart disease prediction with quantum-enhanced machine learning." Scientific Reports 14, no. 1 (2024): 7453.
3. Enad, Huda Ghazi, and Mazin Abed Mohammed. "Cloud computing-based framework for heart disease classification using quantum machine learning approach." Journal of Intelligent Systems 33, no. 1 (2024): 20230261.
4. Narain, R., S. Saxena, and A. K. Goyal. "Cardiovascular Disease Prediction based on Physical factors using Quantum Neural Network." International Journal of Information Processing 8, no. 3 (2014): 83-94.
5. Karthigeyan, S., M. Diviya, R. Bhuvaneswari, and T. Sabhanayagam. "QCVNET-Leveraging Quantum Technology in Convolutional Neural Networks for Improved Cardiovascular Disease Prediction." In 2024 5th International Conference on Circuits, Control, Communication and Computing (I4C), pp. 495-502. IEEE, 2024.
6. Enad, Huda Ghazi, and Mazin Abed Mohammed. "A Review on Artificial Intelligence and Quantum Machine Learning for Heart Disease Diagnosis: Current Techniques, Challenges and Issues, Recent Developments, and Future Directions." Fusion: Practice & Applications 11, no. 1 (2023).
7. Alsubai, Shtwai, Abdullah Alqahtani, Adel Binbusayyis, Mohemmed Sha, Abdu Gumaei, and Shuihua Wang. "Heart failure detection using instance quantum circuit approach and traditional predictive analysis." Mathematics 11, no. 6 (2023): 1467.
8. Zhang, Shiyue, Aijuan Wang, and Lusi Li. "Quantum-convolution-based hybrid neural network model for arrhythmia detection." Quantum Machine Intelligence 6, no. 2 (2024): 1-16.
9. Oviesi, Safura, and Mohamad Jafar Tarokh. "Quantum neural network-assisted learning for small medical datasets: a case study in emphysema detection." The Journal of Supercomputing 81, no. 1 (2025): 1-32.
10. Bopardikar, Soham, Manuel Montoya, Pierre Decoodt, Juan G. Lalinde-Pulido, and Daniel Sierra-Sosa. "Quantum Machine Learning and TensorNetworks as an Aid in the Clinical Diagnosis ofCoronary Artery Disease." (2024).
11. Xiang, Qiuyu, Dongfen Li, Zhikang Hu, Yuhang Yuan, Yuchen Sun, Yonghao Zhu, You Fu, Yangyang Jiang, and Xiaoyu Hua. "Quantum classical hybrid convolutional neural networks for breast cancer diagnosis." Scientific Reports 14, no. 1 (2024): 24699.
12. Sha, Mohemmed, and Mohamudha Parveen Rahamathulla. "Quantum deep learning in Parkinson's disease prediction using hybrid quantum–classical convolution neural network." Quantum Information Processing 23, no. 12 (2024): 383.
13. Waskar, Sheral, Santhosh Raminedi, and S. Shridevi. "Heart Failure Risk Assessment with Quantum-Inspired Support Vector Classification: An Empirical Study." In International Health Informatics Conference, pp. 53-68. Singapore: Springer Nature Singapore, 2023.
14. Ullah, Ubaid, and Begonya Garcia-Zapirain. "Quantum machine learning revolution in healthcare: a systematic review of emerging perspectives and applications." IEEE Access 12 (2024): 11423-11450.
15. Sheikh, Fazil, S. Y. Amdani, and Shaini Suraj. "Quantum convolutional neural network for Alzheimer's disease diagnosis." In AIP Conference Proceedings, vol. 3188, no. 1. AIP Publishing, 2024.
16. Dutta, Shivanya Shomir, Sahil Sandeep, S. Sridevi, Mohan Ram Sridhar, and Chaitanya Singh. "Quantum kernel based support vector machine with quantum approximate optimization algorithm embedding for improved lung cancer prediction." In 2024 15th international conference on computing communication and networking technologies (ICCCNT), pp. 1-6. IEEE, 2024.
17. Swathi, Gunti, Ali Altalbe, and R. Prasanna Kumar. "QuCNet: quantum-inspired convolutional neural networks for optimized thyroid nodule classification." IEEE Access (2024).
18. Balakrishnan, D., Umasree Mariappan, Pagadala Geetha Manikanta Raghavendra, Pallela Karthikeya Reddy, Rayavarapu Lakshmi Narasimha Dinesh, and Shaik Bugganapalli Jabiulla. "Quantum neural network for time series forecasting: Harnessing quantum computing's potential in predictive modeling." In 2023 2nd International Conference on Futuristic Technologies (INCOFT), pp. 1-7. IEEE, 2023.
19. Maheshwari, Danyal, Begonya Garcia-Zapirain, and Daniel Sierra-Sosa. "Quantum machine learning applications in the biomedical domain: A systematic review." Ieee Access 10 (2022): 80463-80484.
20. Asadoorian, Nazeh, Shokufeh Yaraghi, and Araeek Tahmasian. "Pre-trained quantum convolutional neural network for COVID-19 disease classification using computed tomography images." PeerJ Computer Science 10 (2024): e2343.

