Quantum-inspired seagull optimised deep belief network approach for cardiovascular disease prediction

D. Banumathy[1], T. Vetriselvi[2], K. Venkatachalam[3] and Jaehyuk Cho[4]

[1] Department of Computer Science and Engineering, Paavai Engineering College, Namakkal, India
[2] School of Computer science and Engineering, VIT University, Vellore, India
[3] Department of Software Engineering, Jeonbuk National University, Jeonju-si, Republic of South Korea
[4] Department of Software Engineering & Division of Electronics and Information Engineering, Jeonbuk National University, Jeonju-si, Republic of South Korea

ABSTRACT

The early detection and accurate diagnosis of cardiovascular diseases is vital to reduce global morbidity and death rates. In this work, the quantum-inspired seagull optimization algorithm (QISOA) combined with a deep belief network (DBN) is proposed to improve the identification of cardiovascular disorders. As part of preprocessing, cleaning, transformation, and standardization are performed to eliminate noise, inconsistencies, and scaling issues in the data. QISOA is used to optimize the weights and biases of the DBN model, enhancing its prediction efficiency. The algorithm incorporates quantum mechanics concepts to develop its exploration potential further, leading to faster convergence and increased global search efficiency. Optimized DBN provides efficient acquisition of hierarchical representations of the data, which results in improved feature learning and classification accuracy. The publicly accessible Cleveland Heart Disease dataset is used to assess the performance of the suggested model. Extensive experiments are conducted to demonstrate the superior performance of the QISOA-optimized DBN model compared to traditional machine learning and other metaheuristic-based models. Initially, machine learning models such as support vector machines, decision trees, Random Forests, multi-layer perceptrons, and fully connected networks were considered for comparison with the cardiovascular predictive performance of the DBN model. Further, meta-heuristic optimization algorithms such as particle swarm optimization, genetic algorithm, grey wolf optimization, cuckoo search optimization and crow search algorithm are combined with the machine learning models and the classification efficiency is evaluated. Additionally, few state-of-the-art techniques proposed in the existing literature are investigated and compared against the proposed model. It was evident from the comprehensive performance assessment of the proposed model that it yields a higher accuracy of 98.6% with precision, recall, and F1-scores of 97.6%, 96.8%, and 97.1%, respectively, compared to other traditional and existing models for cardiovascular disease prediction.

Subjects Artificial Intelligence, Data Mining and Machine Learning, Data Science, Optimization Theory and Computation, Social Computing
Keywords Quantum computing, Deep belief networks, Seagull optimization, Heart disease prediction, Deep learning

Submitted 28 August 2024
Accepted 18 October 2024
Published 13 December 2024

Corresponding author
Jaehyuk Cho, chojh@jbnu.ac.kr

Academic editor
Bilal Alatas

Additional Information and Declarations can be found on page 26

DOI **10.7717/peerj-cs.2498**

How to cite this article Banumathy D, Vetriselvi T, Venkatachalam K, Cho J. 2024. Quantum-inspired seagull optimised deep belief network approach for cardiovascular disease prediction. **PeerJ Comput. Sci. 10:e2498 DOI 10.7717/peerj-cs.2498**

INTRODUCTION

As heart-related diseases have an elevated death rate worldwide, it has emerged as an important medical priority for many individuals. Though conventional therapeutic analysis of information poses considerable obstacles in diagnosing cardiac disorders such as coronary artery disease, ischaemic attacks, cardiac arrest, chest pain, *etc*., early detection of cardiac ailments has the potential to save many lives (*Akella & Akella, 2021*). Adult males are more likely than females to develop cardiovascular disease, particularly in middle-aged or older individuals. This rapid spike in cardiovascular illness puts a substantial risk to community wellness and puts a significant strain on globally connected healthcare systems. In addition, the likelihood of cardiac disease is rising among youngsters in wealthy nations. The threat of heart attacks and strokes is increased by several improper lifestyle choices, which leads to high blood sugar levels, high blood pressure, overweight and elevated fatty acids (*Al-Tashi, Rais & Jadid, 2019*).

In addition, there are further signs, including problems in respiration, inflammation of limbs, insomnia, erratic heart rate, and, in rare cases, excessive obesity with continuously increasing weight. Each of the above signs is symptomatic of many conditions, including disorders that influence older adults, and they are more challenging to identify in the earliest phases of the condition (*Chicco & Jurman, 2020*). A physiological evaluation of the major arteries is utilised to identify heart diseases using pharmaceutical and echocardiographic stress assessment. Yet another kind of imaging assessment for cardiac disease is coronary artery imaging, which offers details on the underlying anatomy of the circulatory system (*García-Ordás et al., 2023*). Nonetheless, the intricacy of the plaque's construction and the complicated structure of the heart make it possible for the illness to be misdiagnosed. Furthermore, using contrast fluid and regional sedation employed during the diagnostic process harms the patient's health.

Accurately assessing the possibility of cardiovascular disease helps prevent severe myocardial catastrophes and safeguard individuals (*Reddy et al., 2021*). The exponential advancement in artificial intelligence and machine learning have provided researchers access to vast datasets to develop the most accurate prediction model. Machine learning (ML) and deep learning (DL) algorithms are capable of making precise medical diagnoses when trained on pertinent data (*Kishor & Chakraborty, 2022*). Therefore, research into the effects of danger indicators, how frequently they occur in people of all ages, their significant influence on specific heart conditions, and their therapeutic potential to lower the incidence of heart diseases is imperative.

Moreover, when using these methods for forecasting a particular disease, it is crucial to remember the optimal characteristics. The input dataset, in general, will be initially unreliable and include irrelevant and duplicated information with overlapping entries and inaccurate data (*Enad & Mohammed, 2023*). Selecting the most vital and significant characteristics that could be applied to identifying or categorising the targeted disease is crucial for predictive modelling. The essential attribute pairings that maximise the predictive potential and aid in improving the accuracy of the machine learning model should be selected to create reliable estimations (*Sarra et al., 2022*).

Depending on their feature selection methods, many researchers consider different probabilistic variables or features as input for their models when modelling the predictors for such diseases (*Nandy et al., 2023*). Several factors are considered when implementing cardiovascular disease prediction models, including lifespan, elevated cholesterol levels, blood sugar levels, heartburn, arterial pressure, anxiety, heart rate variability, genetic background, inadequate nutrition, weight gain, usage of alcohol, and lack of physical activity. These elements are crucial for identifying and categorising diseases (*Alsubai et al., 2023*).

The influence of feature selection algorithms goes beyond making learning algorithms more capable of classifying data and making dataset reduction in dimension possible. As a result, selecting characteristics involves solving a multi-objective optimisation task that maximises classification effectiveness while minimising the number of features (*Ghosh et al., 2021*). Optimisation is a vital and significant field in scientific research and engineering applications. Multi-objective optimisation problems typically involve optimising many objectives simultaneously in practical settings. Multi-objective optimisation tasks, in contrast to single-objective ones, seek to identify the optimum solutions or the most suitable paths in the choice space (*Ozpolat & Karabatak, 2023*).

Multi-objective optimisation is among the trickiest and most well-liked challenges in current evolutionary computing research. When a multi-objective optimisation issue is solved, equitable solutions are typically produced. To produce a solution set that covers the complete outcomes as much as possible, the optimisation algorithm must discover solutions that are as close to as feasible (*Ullah et al., 2022*). As a result, multi-objective optimisation algorithms must strike a balance between the distribution of optimal solutions and the algorithm's convergence. Nevertheless, local optimization is a common difficulty with multi-objective optimization methods, which can result in uneven convergence and distribution issues.

In the past 10 years, a new topic of study called quantum computing (QC) has attracted much attention. It deals with studies on quantum machines and quantum computations. The foundational ideas of quantum mechanics, such as qubit encoding and dynamic superposition, are used in quantum computing (*Maheshwari et al., 2023*). Large numbers of quantum states can be processed concurrently using quantum computing. Because of its fundamental notions, this quantum computing technology gives optimization a new philosophical perspective. Research combining quantum computing and evolutionary computation has recently garnered increasing attention conceptually and operationally (*Kavitha & Kaulgud, 2023*). The goal is to use the power of quantum computing to improve the speed and efficiency of traditional evolutionary algorithms. As a result, numerous algorithms such as genetic (*Saad et al., 2021*), particle swarm (*Agrawal, Kaur & Agarwal, 2021*), binary bat (*Sharma & Sharma, 2022*), and differential evolution (*Wang & Wang, 2021*) algorithms have been developed and influenced by quantum mechanics. Quantum-inspired algorithms do not require a quantum machine to function, in contrast to classical quantum computing. Numerous combinatorial optimization issues have been effectively resolved using algorithms with quantum inspiration (*Kavitha & Kaulgud, 2023*).

By utilizing the ideas of quantum computing, metaheuristic optimization algorithms can perform even better. It has always been challenging to create an appropriate equilibrium in metaheuristic algorithms between exploration, which is the global search, and exploitation, which is considered the local search, as concentrating on one undermines the other (*Priya, VinilaJinny & Mate, 2020*). The primary benefit of applying quantum computing techniques in metaheuristics is strengthening the global search capability without degrading the algorithmic exploitation phase.

This study offers a quantum-based improved seagull optimization algorithm (QISOA) to optimize the convergence and distribution of solutions in multi-objective optimization problems. The QISOA is an integrated algorithm that combines the seagull optimization algorithm with a search method motivated by quantum mechanics. The concept of superposition, quantization gates, and waves that remain stationary from quantum mechanics has been incorporated into quantum-inspired search algorithms, which are simple to locate in localized convergence of global searches. Furthermore, the seagull optimization does well in local searches, preserving the equilibrium between exploitation and exploration.

In this work, deep belief networks is employed to classify cardiovascular disease. A deep belief network (DBN), an established algorithm used in deep learning, combines low-level data to find distributed feature representations, generating a more abstract high-level model. Using aggressive learning, it can directly extract fundamental features from minimal information layer by layer. The DBN can improve detection ability by avoiding the added ambiguity and intricacy of conventional feature discovery. Consequently, DBN is frequently used to classify problems in the healthcare domain for disease diagnosis. The QISOA with superior optimisation performance is utilised to choose the the parameters in Deep Belief networks to address the subjective identification problem. By removing human factor interference and adaptively choosing the best DBN settings, the QISOA-DBN approach can increase the classification accuracy of cardiovascular diseases.

Research questions

The present research is carried out to find answers to the following questions,

a. How successful is the quantum inspired seagull optimisation algorithm (QISOA) at optimising the deep belief network's parameters?
b. Does the QISOA algorithm enhance the DBN's ultimate performance and rate of convergence?
c. What are the possible outcomes of using the QISOA-DBN model for the early detection of cardiovascular diseases?

Research contributions

The major contributions of this research are,

1. To propose a quantum-inspired seagull optimization algorithm that incorporates quantum principles into the conventional seagull optimization algorithm to improve

the balance between exploration and exploitation, which leads to better convergence towards optimal solutions.

2. To implement deep learning-based methods such as deep belief networks algorithm optimized by QISOA technique to perform cardiovascular disease prediction with enhanced efficiency and reliability.
3. To conduct a comprehensive performance analysis of the proposed QISOA-DBN model to demonstrate its effectiveness in diagnosing cardiovascular disorders by comparing it to the state-of-the-art methods described in the literature.

Paper organization

The remainder of the article is organized as follows: "Related Works" describes the state-of-the-art work on cardiovascular disease identification using ML/DL techniques and optimization algorithms. "Preliminaries" provides a theoretical background on quantum computing concepts and conventional seagull optimisation techniques. "Proposed Methodology" discusses the proposed methodology addressing the core concepts of the suggested QISOA-DBN approach. "Results and Discussion" presents the results obtained by applying various ML/DL techniques on the Cleveland Heart Disease prediction dataset and compares the existing and proposed systems. "Conclusion" concludes the present research.

RELATED WORKS

The subsequent section expounds on the diverse approaches and formulations employed in prognosticating cardiac anomalies and the former methods of evaluating the diagnosis of heart illness to mitigate its advancement. The proposed method in *Wang et al. (2023a)* uses convolutional neural networks bi-directional long short-term memory to build an integrated approach based on deep learning determining whether an individual has been adversely affected by heart disease. Using data processing methodology has resolved issues with erroneous and inconsistent data values in a dataset. Additionally, it has employed more tree-based learners in the feature extraction and categorization process. Many patient records have been successfully used to improve the system's performance under consideration, and the dynamic detection framework has been handled with multiple data points collected from different periods. To generate a vector of characteristics based on time for medical diagnosis, the experimental evaluations have demonstrated classification outcomes with the smoothing-based erratic recurring time throughout different stages. With the integration of intelligent medical care systems, deep learning architecture has been shown to advance as a diagnostic instrument for improved clinical treatment and decreased detection expenditure.

Inaccurate identification and premature disease detection have been the main contributing factors to morbidity from heart conditions. Accurate classification of heart failure requires efficient algorithmic parameter tweaking and training (*Tharwat & Hassanien, 2019*). The over-fitting problem has affected several sophisticated mechanisms designed for the automatic identification of cardiac failures, which significantly negatively impacts the system's accuracy. The recommended methods have weakened the accuracy performance of testing and training data. The diagnostic procedure employed a random search technique to find relevant features for disease prediction (*Yildirim et al., 2021*). Two different types of assessments have been used to evaluate the accuracy of the suggested system. The Random Forest model was used in the initial experiment, and the random search procedure was experimented with as the second model. Compared to the conventional method, the accuracy has been observed to be higher and the complexity to be lower. Eleven distinct algorithms have been compared to assess the system's effectiveness (*Tharwat & Hassanien, 2018*).

Different kinds of patient data are collected by healthcare organisations, which has aided physicians in accurately diagnosing patients' medical conditions. The symptoms that patients have reported, the initial diagnosis made by the medical professionals, and comprehensive test results that may help forecast cardiac abnormalities are all included in this health information (*Munshi et al., 2024*). Professional physicians have used this data to analyse and confirm the illness. With artificial intelligence (AI) based algorithms like artificial neural networks to determine whether or not the patient has been affected, accurate identification and classification have been completed by the recommended study (*Abdulsalam, Meshoul & Shaiba, 2023*).

To improve the learners, the UCI heart disease dataset and a voting-based method were used in the suggested study (*Asif et al., 2023*). While machine learning technologies help evaluate large amounts of data, they sometimes include extraneous and irrelevant elements that negatively impact classification accuracy and processing performance. A feature selection procedure has been used in the examined research to remove undesirable and irrelevant data features. The ideal subset for diagnosing heart-related disorders has been identified by selecting features using the wrapper technique. A quantum-inspired genetic algorithm (*Sarra et al., 2022*) was used to choose the best features from the dataset. The support vector machine (SVM) classifier has also assessed the algorithm's fitness.

When it comes to predicting heart problems, the quantum particle swarm optimisation combined with the SVM model has the highest level of accuracy. *Rao & Prasad (2021)* uses QPSO to adjust the ideal SVM parameter to predict heart diseases accurately. Future heart disease prognostic implementations should incorporate various techniques to increase detection efficiency and precision for real-time use. According to one study (*AvinashPrabhu & Parthipan, 2022*), the suggested method outperforms existing classification techniques regarding precise prediction and economical computing time. According to *Bharti et al. (2021)*, using quantum-based technology shortens computing times, which improves system performance.

The application of the quantum paradigm shows promise for accelerating the clustering process. Using a quantum circuit method on the heart disease dataset, the quantum K-means clustering technique was used in *Kumar et al. (2021)* to find instances of heart disease. The study results show that the quantum mechanical clustering technique performs noticeably faster than the conventional clustering method at a time complexity. In addition, a comparison of performance metrics between quantum and classical processing (*Nagarajan et al., 2022*).

Fibrous bands of tissue can form between the heart and surrounding structures, such as the pericardium (the protective sac around the heart), following surgical procedures like open-heart surgery (*Wang et al., 2023b*; *Zhang et al., 2021*). This phenomenon occurs through various pathways, including the induction of inflammation, fibrosis, and the promotion of the transition of valve interstitial cells (VICs) into osteoblast-like cells, which are responsible for depositing calcium in the valve (*Huang et al., 2024*; *Huang, Shu & Liang, 2024*). Counterfactual analysis is valuable in enhancing the robustness of AI systems by enabling them to check their predictions against hypothetical changes, thereby improving accuracy and reducing error (*Cai et al., 2024*; *Sun et al., 2023*). Additionally, in real-time diagnostic settings, cytopathologists can use a technique called rapid on-site evaluation (ROSE) to assess cells collected during a biopsy or surgery while the procedure is still ongoing (*Lahori et al., 2023*; *Zhao et al., 2023*). At the molecular level, the enzyme ALKBH5 normally removes m6A modifications from mRNA, which plays a key role in regulating how long the mRNA remains intact and available for translation into proteins (*Zhao et al., 2021*; *Li et al., 2023*). Traditional AI approaches simulate decision-making processes using algorithms and computational models (*Chen et al., 2023*; *Duan et al., 2023*). However, deep learning has emerged as a powerful tool to enhance the resolution and speed up image reconstruction, making super-resolution imaging faster and more accessible (*Luan et al., 2023*; *Yu et al., 2023*). In another biological process, long non-coding RNAs (lncRNAs) act as "sponges" by binding to microRNAs (miRNAs) and preventing them from interacting with their target mRNAs, thus influencing gene expression (*Zhou et al., 2022*; *Bing et al., 2022*). Similarly, reduced blood flow to the brain, often due to atherosclerotic blockages in the carotid arteries, can negatively affect brain function, leading to cognitive decline, including impairments in memory, attention, and processing abilities (*Huo et al., 2022*). Quantum approaches were found to be excellent choices for early cardiac disease prediction. These quantum machine learning techniques classify heart disease datasets reported areas under the curve value of 97.42%, respectively (*Mehmood et al., 2021*). The utilisation of the complete dataset was impacted by the constraints placed by the study resources. The application of quantitative structure-activity connection machine learning approaches produced a 94.73% predictive accuracy for heart disease, according to the findings of this study. Table 1 compares the existing works on cardiovascular disease detection.

Motivations of current research

The cardiovascular disease diagnosis techniques currently in operation frequently have low diagnosis rates, and thus, improvement in detection rates usually necessitates several iterations with poor convergence. This shortcoming emphasizes the need for a quicker and more precise diagnosis technique that might improve the general efficacy of heart disease prediction. Using sophisticated machine learning models—especially those influenced by quantum algorithms—offers a viable way to improve heart disease diagnostic performance. By increasing the precision of disease prediction, utilizing approaches like QISOA and DBN may be able to advance beyond the drawbacks of the present methods. Machine learning models such as DBNs must fine-tune their parameters for better classification

Table 1 **Comparison of existing works on cardiovascular disease detection**

References	Technique used	Inferences	Limitations
Maheshwari et al. (2023)	Quantum-based swarm optimization	Demonstrates higher values of sensitivity, specificity, accuracy, and precision	The proposed system needs to be implemented in real-time data to improve the performance further
Kavitha & Kaulgud (2023)	Quantum K-means clustering	Higher accuracy levels of 95.87%	Appropriate feature selection and optimization methods not implemented
Wang et al. (2023a)	Hybrid random forest and CNN	Minimal false alarms are produced by this method	The model needs to be enhanced by applying it to larger datasets
Tharwat & Hassanien (2019)	Quantum-inspired particle swarm optimization	Provides efficient solutions for the identification of affected individuals with heart diseases	The model needs to be fine-tuned with more real-time data to maximize the predictive analysis
Yildirim et al. (2021)	Deep neural networks	Maximized performance is achieved with increased accuracy and precision values	The computational cost of the recommended approach is high
Tharwat & Hassanien (2018)	Quantum SVM	This suggested technique exhibits maximum accuracy value for label prediction	Feature fusion techniques are not implemented
Munshi et al. (2024)	Quantum genetic algorithm	Extensive evaluation of the proposed technique is performed, including and excluding irregularities in data	Deep learning algorithms can be combined with the proposed techniques to improve accuracy
Abdulsalam, Meshoul & Shaiba (2023)	Quantum ensemble model	Process overhead is the minimum	Training time is high

accuracy. QISOA offers a valuable mechanism to strike a balance between exploration and exploitation in the search space, thus it is implemented to improve the efficiency of the DBN model.

PRELIMINARIES

This section elaborates on the fundamental concepts in the quantum computing domain and briefs on the conventional seagull optimization technique.

Basics of quantum computing

Quantum computing's foundations are quantum physics, with the Schrodinger equation as the underlying principle. Unlike the traditional approach, quantum computing demonstrates its potential to operate more quickly, particularly when tackling complicated, challenging problems requiring a wide range of possible solutions. This section describes some of the basic concepts in quantum computing.

Qubit

In quantum computing, the tiniest information is denoted as the qubit. The qubit vectors in quantum computing exist in two possible states, which are $|0\rangle$ and $|1\rangle$ and it is represented as shown in Eq. (1) as:

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad and \quad |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{1}$$

In the above equation, each term is considered as the intricate conjugate transposition of the other.

Superposition

Superposition is the ordered combination of the vectors involved in the computation. The quantum superposition for the vectors represented in Eq. (1) is as shown in Eq. (2),

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle. \tag{2}$$

The values of the qubit product and angle between vectors are determined using the representations given in Eq. (3),

$$\rho = \alpha\cdot\beta \quad \mathit{and} \quad \delta = \mathit{arc}\,\tan(\beta/\alpha). \tag{3}$$

Gates

Quantum gates are the primary components essential to processing information. The four major types of gates involved in data processing are rotation gates, NOT gates, Hadamard gates, and controlled-NOT gates.

Coherence

Coherence for each qubit state is calculated using the Hilbert spectrum of contributing wave processes, $|\psi\rangle$. A consistent phase association is measured and recorded during this procedure. Decoherence is the reverse of a coherent occurrence. It happens when the phase connection that links the wave processes is forced to be broken or when the quantum superposition of the involved basic states fails.

Conventional seagull optimization

An innovative swarm optimization algorithm mimics seagulls' relocating and assaulting activity is called the seagull optimization algorithm (SOA) (*Asif et al., 2023*). For multi-objective optimisation problems, an SOA extension has been created that incorporates variable network, leader-selection, and conserving techniques. When migrating, they move in clusters. To avoid crashes, seagulls adjust their initial points of interest. The gulls will flock together and fly towards the fittest gull, who has the most chance of surviving. The other seagulls will adjust their starting places based on the most fitting gull. When birds travel from one spot to another over the sea, gulls attack them. When they attack, they move in a spiral pattern. The migration operation uses the SOA's exploratory power to replicate how a flock of seagulls moves from one location to another. The attack operation uses the algorithm's exploitative power to mimic how flocks of seagulls pursue their target.

PROPOSED METHODOLOGY

The proposed research aims to provide an efficient technique for cardiovascular disease classification by integrating the concepts of quantum computing, deep learning, and optimization algorithms. This work modifies the conventional seagull optimization by

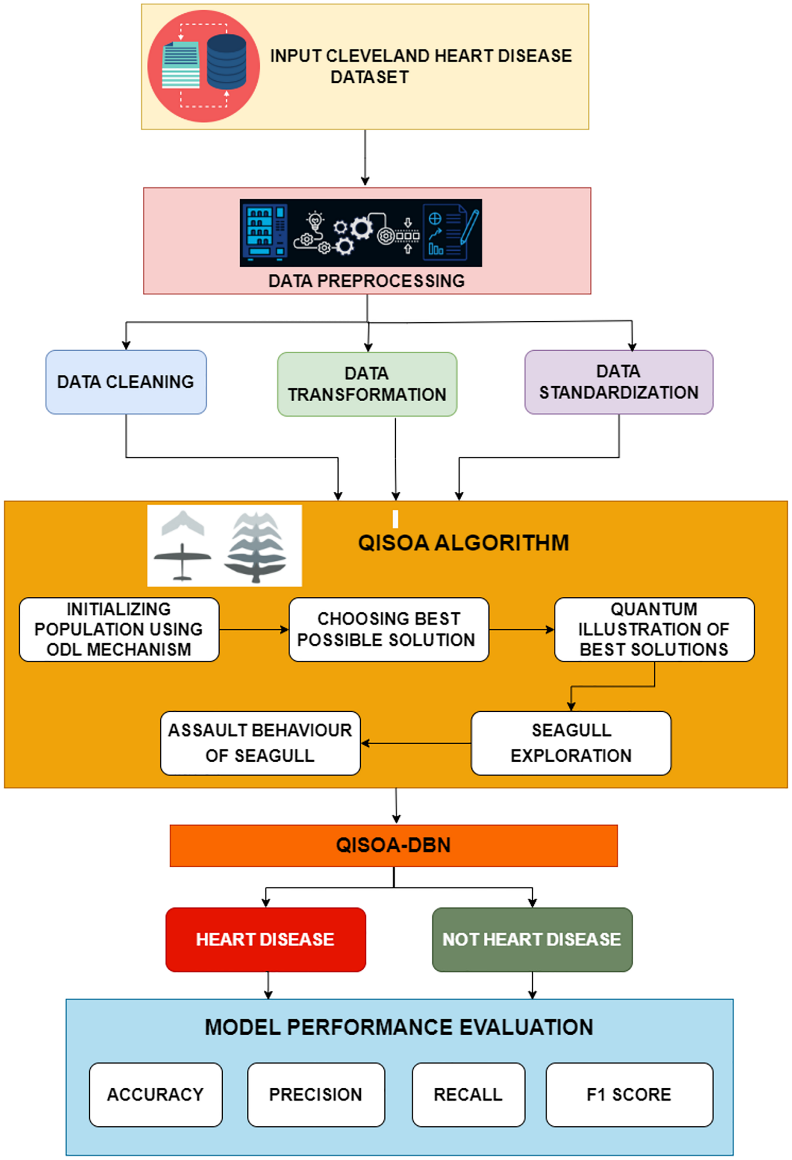


Figure 1 **Proposed workflow.** Full-size DOI: 10.7717/peerj-cs.2498/fig-1

introducing quantum principles to make it a multi-objective optimization algorithm. This quantum-inspired seagull optimization algorithm is further used to choose the parameters of the DBN algorithm, which is employed as the classification technique for cardiovascular disease diagnosis. This section first describes the data preprocessing techniques applied to the dataset. Secondly, the QISOA technique is explained in detail, highlighting its various steps. Thirdly, the deep belief network algorithm, which is optimized by the QISOA technique to enhance performance efficiency, is described. The workflow of the proposed methodology is presented in Fig. 1.

Data preprocessing

The data preprocessing in this work involves three techniques: cleaning, transformation, and standardization. Data cleaning is essential to identify and eliminate redundant, erroneous, or unavailable values. Data transformation is needed to remove compatibility issues based on the nature of the data. Data standardization employs statistical approaches to normalize the data values within a particular range.

Data cleaning

An essential first step is data cleaning, carefully reviewing the dataset to look for any irregularities, such as errors, outliers, or missing numbers. Missing values have to be approximated or eradicated from the dataset, and inconsistencies and abnormalities are corrected or discarded when found. Missing values should be appropriately addressed during data preparation for accurate and trustworthy analysis. In the Cleveland Heart Disease dataset, 293 redundant rows have been altered and removed. As a result, 292 unique cases—158 matched patients with heart disease and 134 matched patients without heart disease—were identified. Furthermore, 262 instances that were similar to each other have been identified, which resulted in the removal of repetitions.

Data transformation

The learning performance may be low when specific data properties have large input values that conflict with other features. As a result, a record investigation was done to examine and choose relationships visually to fit different aspects. Features such as thal and slope, which have larger values, are split further into three values, and one encoding is applied to the dataset. These three features are separated into thal_00, thal_01, thal_02 and thal_03, and slope_00, slope_01, slope_02 and slope_03 before they are merged with the initial data sets. The data are scaled according to the findings of the statistical assessment before being processed further.

Data standardization

The range of values in the distribution is taken, the mean is subtracted, and the data are then divided by the standard deviation to normalize it. Scaling is performed by dividing every value in the data by the standard deviation and removing the average from the data values. The mathematical representation given in Eq. (4) is used for standardization,

$$a' = \frac{a - \mu}{\sigma}. \tag{4}$$

The value in standardised form for original data a is computed as $a\prime$ using standard deviation and mean values such as σ and μ respectively. The values of σ and μ are determined by using the formulas in Eqs. (5) and (6),

$$\sigma = \sqrt{\frac{1}{K}\sum_{n=1}^{K} aK\ (A_n - \mu)^2}. \tag{5}$$

$$\mu = \frac{\sum_{n=1}^{K} A.K}{K}. \tag{6}$$

The data is normalised by taking the highest and lowest values in the data range using the formula in Eq. (7),

$$b' = \frac{B - B_{low}}{B_{high} - B_{low}}. \tag{7}$$

QISOA

The QISOA approach incorporates quantum computing superposition concepts, layout ranking, seagull mobility and assault behaviour, and opposition-driven learning (ODL). The ODL process initialises the seagull group to produce an improved setup population. The real-coded quantum representation of the existing ideal approach and quantum rotation gate have been adjusted to more effectively preserve an equilibrium between exploration and exploitation to attain the best possible solutions. A layout mechanism that combined the layout density ordering and the global layout ordering additionally offered an indicator for the maintenance and selection of leaders.

Step 1: Initializing population using ODL mechanism

The likelihood of sampling better regions in population-based algorithms is decreased by initialisation using random roulette in conventional evolutionary algorithms when beforehand data is not available. On the other hand, without previously acquired knowledge, ODL can identify more appropriate initial candidate regions, raising the likelihood of identifying superior regions and holding out hope for enhancing fitness.

Step 2: Choosing the best possible solution

This step employs layout density ordering and global layout ordering mechanisms to characterize the most optimal solutions. This technique can display the differences in objective values between optimal solutions and alternative solutions and highlight the superiority and inferiority of each option. This work uses layout density ordering (LDO) to increase the spread of the solutions and global layout ordering (GLO) to improve the algorithm's efficiency.

The total number of unique layout positions in each objective that are better than those of other points can be calculated by the GLO technique as represented in Eq. (8):

$$GLO(a_k) = \sum_{h=1}^{P_k} count(L_m(a_k) < L_m(a_l)). \tag{8}$$

In Eq. (8), a_k *and* a_l are two individuals chosen from the population created using Algorithm 1 in which k and l are different. $L_m(a_k)$ denotes the layout positions of a_k for one objective from the total set of objectives represented as P_k. Furthermore, the value of LDO is computed as given in Eq. (9):

$$LDO(a_k) = count\left(\sum_{h=1}^{P_k} |L_m(a_k) - L_m(a_l)\,| < G\right). \tag{9}$$

Algorithm 1 Initializing population using ODL.

```
Input: $N$ ($a_1, a_2, \ldots, a_n$) are points in a $T$-dimensional space where $a_k \in [x_k, y_k]$ and $k = 1, 2, 3, \ldots, T$
Output: Final population $N^*$
Step 1: Initialize the start values $N_k$ and categorize into two groups: $N_1$ and $N_2$
    Generate individuals arbitrarily in group $N_1$
Step 2: Produce opposite points using Orthogonal Design Layout (ODL)
    $\tilde{a}_k = x_k + y_k - a_k$
  and insert these into group $N_2$
Step 3: Merge the populations of $N_1$ and $N_2$
Step 4: Eliminate overlapped individuals in both groups by applying intersection operation
Step 5: Apply arbitrary functions to balance the eliminated individuals
    Generate the final population $N^*$
```

In the above equation, G represents the present set of objectives in the population. This mechanism prefers the candidate solution with a higher GLO value among those in the collection. Furthermore, when multiple solutions have identical GLO values, the solution with the lowest value is favored as the best option. A roulette wheel chooses a solution as the optimal solution among several solutions with the same GLO values.

Step 3: Quantum illustration of best solutions

This step is comprised of the quantum illustration of the best possible solution and the quantum rotation gate, which updates the likelihood magnitudes of two states. A combined function of dimensions and duration, known as the wave function, is needed to fully characterize a quantum state in the theory of quantum mechanics. The magnitude coefficient denotes the likelihood that the quantum state will occur at a suitable place and time. To compute the observed values of the quantum illustration, a typical wave function is implemented as shown in Eq. (10):

$$|W(a_k)|^2 = \frac{1}{\sqrt{2\pi\theta_k}} exp\left[-\frac{(a_k - \mu_k)^2}{2\theta_k^2}\right]. \tag{10}$$

In the above equation, μ_k denotes the average location of the individual and θ_k represents the dispersal spectrum of the individuals with respect to the value of μ_k. The quantum illustration of the best possible solution (R_c^i) is determined using Eq. (11) as:

$$R_c^i \triangleq \begin{bmatrix} a_{c,1} & a_{c,2} & \ldots & a_{c,k} \\ \gamma_1 & \gamma_2 & \ldots & \gamma_k \end{bmatrix}. \tag{11}$$

In Eq. (11), i represents the present iteration for a set of values, γ_k represents the coefficient of magnitude for each optimal solution.

Step 4: Seagull exploration

During this stage, the seagulls relocate themselves from one position to another so they do not collide. The orientation of relocation without interference is determined using the movement nature of the seagull in the search space (F) as shown in Eq. (12):

$$D_z = F \times C_z(i). \tag{12}$$

In the above equation, C_z represents the present location of individual and D_z denotes the relocated position such that it does not collide with any other individual in the search space. The value of F is computed in a non-linear fashion in order to improve the efficiency of the algorithm using Eq. (13):

$$F = m_s(2^\tau - 1) \tag{13}$$

$$\tau = e^1 - \frac{i_{max}}{i_{max} - i}. \tag{14}$$

The relocation direction of each individual in quantum representation towards the best possible solution is given using Eq. (15) as:

$$E_z = G \times (R_c(i) - C_z(i)) \tag{15}$$

$$G = 2 \times F^2 \times rand(0, 1). \tag{16}$$

The modified position of the relocated seagull is calculated by combining the values of the orientation of relocation without interference and relocation direction towards the best possible solution as shown in Eq. (17):

$$P_z = |D_z + E_z|. \tag{17}$$

Step 5: Assault behaviour of seagull

During the process of relocating from one position to another, seagulls implement an assault against each other in a 3-D space. The movement of seagulls in 3-D space is described in Eqs. (18) to (21):

$$u = rd \times \cos(l) \tag{18}$$

$$v = rd \times \sin(l) \tag{19}$$

$$w = rd \times l \tag{20}$$

$$rd = m \times e^{ln}. \tag{21}$$

In the above equations, l is an arbitrary value between 0 and 2π and rd denotes the radius with which the seagull makes a spiral motion in the orientation of m and n. The updated position of the seagull is an aggregation of the relocation and assault behaviour, and the final most optimal position is determined as shown in Eq. (22):

$$C_z(i) = (P_z \times u \times v \times w) + R_c(i). \tag{22}$$

QISOA-DBN

One neural network that is stacked with numerous supervised back propagation networks as well as restricted Boltzmann machine networks is the DBN. The DBN model is extensively used in the healthcare domain for disease prediction and classification, due to its great approximation capability to the nonlinear function. The arbitrary initialisation of the DBN connectivity weights as well as biases results in a lack of global optimisation potential and final descent into the local optimum. As a result, the DBN's connection weights and biases have a significant impact on the reliability of disease classification.

The QISOA algorithm handles the functions of quick convergence, population variability, and global search potential. To maximise each node's connection weights, the QISOA with global optimization potential is incorporated into the DBN. To achieve an optimised DBN such as the QISOA-DBN model, a parametric optimization approach based on QISOA is proposed. This method helps to prevent a lack of clarity in parameter selection, minimise its effect on modelling precision, and improve the efficiency of the model for forecasting and classifying cardiovascular diseases.

A transparent layer and a concealed layer constitute every restricted Boltzmann machine. For any neuron x in the transparent layer and y in the concealed layer denoted as (t,c), the energy parameter is represented as in Eq. (23):

$$EG(t,c|\delta) = -\sum_{x=1}^{X} m_x t_x - \sum_{j=1}^{Y} n_y c_y - \sum_{x=1}^{X}\sum_{y=1}^{Y} t_x w_{xy} c_y. \tag{23}$$

In the above equation, $\delta = (w_{xy}, m_x, n_y)$ in which w_{xy} denotes the weight parameter between the nodes in the transparent and concealed layers t and c, respectively. m_x *and* n_y are the bias values for t and c layers. The joint probability distribution is computed based on the energy parameter calculated using Eq. (23) as given in Eq. (24):

$$JP(t,c|\delta) = e^{-EG(t,c|\delta)}/[R(\delta)]. \tag{24}$$

In Eq. (24), the standardization parameter is determined as $R(\delta) = \sum_t \sum_c e^{-EG(t,c|\delta)}$. The actuation states of the nodes in the concealed layer are independent since there is no connection between the layers in the restricted Boltzamnn when the state of the nodes in the transparent layer is determined. The likelihood of actuating a node from the concealed layer y is given by the Eq. (25):

$$JP(c_y = 1|t,\delta) = \sigma\left(n_y + \sum_{x=1}^{X} t_x w_{yx}\right). \tag{25}$$

In the above equation, the sigmoid function $\sigma(a)$ is given by $\sigma(a) = \frac{1}{1+e^{-a}}$. Similarly, the likelihood of actuating neurons from the transparent layer is given by the Eq. (26) as:

$$JP(t_x = 1|c,\delta) = \sigma\left(m_x + \sum_{y=1}^{Y} c_x w_{yx}\right). \tag{26}$$

The sensitivity (φ_x) in the final output layer (f_x) is calculated from the actual value (v_x) as in Eq. (27):

$$\varphi_x = f_x(1-f_x)(v_x - f_x). \tag{27}$$

Similarly, the sensitivity for ith layer can be expressed as shown in Eq. (28):

$$\varphi_x^i = g_x^i\left(1 - g_x^i\right)\sum\nolimits_y w_{xy}^i\ \varphi_y^{i+1}. \tag{28}$$

The modified mathematical formulations for the weight as well as bias values for every layer of the deep belief network as the rate of learning (γ) for every epoch is given by Eqs. (29) to (31):

$$w_{xy}^i = w_{xy}^i + \gamma * g_x^i\ \ \varphi_y^{i+1} \tag{29}$$

$$m_{xy}^i = m_y^i + \gamma * \varphi_y^{i+1} \tag{30}$$

$$n_{xy}^i = n_y^i + \gamma * \varphi_y^{i+1}. \tag{31}$$

RESULTS AND DISCUSSION

This section presents the dataset used in assessing the proposed system's performance. It compares the experimental results obtained with traditional machine learning and deep learning algorithms, optimized ML/DL techniques, and a few existing works from the literature.

Dataset description

The Cleveland benchmark dataset from the UCI machine learning repository is used in this study. The Cleveland dataset has fourteen characteristics and 303 records. The classification of the target attribute is represented with two class values labeled 1 and 0, denoting patients with and without heart disease, respectively. The assortment is approximately balanced because 165 records represent individuals having heart disease and 138 records represent individuals without heart disease. The dataset used in this research can be accessed using the link: https://archive.ics.uci.edu/dataset/45/heart+disease. This dataset is a combination of numerical and categorical features. The numerical features in the dataset include age, cholesterol, resting blood pressure, maximum heart rate, ST depression, and number of major vessels. The categorical features in this dataset are chest pain types, sex, fasting blood sugar, resting electrocardiogram, exercise-induced angina, ST slope, and Thalassemia. The data distribution of numerical and categorical features is analyzed and presented in Figs. 2 and 3.

Experimental setup

The experimental setup used in this research includes an NVIDIA GTX 1080 Ti as a dedicated GPU with CUDA compatibility with 16GB RAM and Windows 11 Operating System. Python 3.7 version is utilized with the Jupyter Notebook environment for code execution. Libraries such as TensorFlow, pandas, numpy, scikit-learn and matplotlib were used for the algorithm implementation.

The parameter settings for the QISOA-DBN model are described in this section. The population size of QISOA, which specifies the number of seagull agents, is 50, with maximum optimization iterations set as 200. The quantum parameter, which balances exploration and exploitation, is assigned a value of 0.5. The inertia weight, which controls

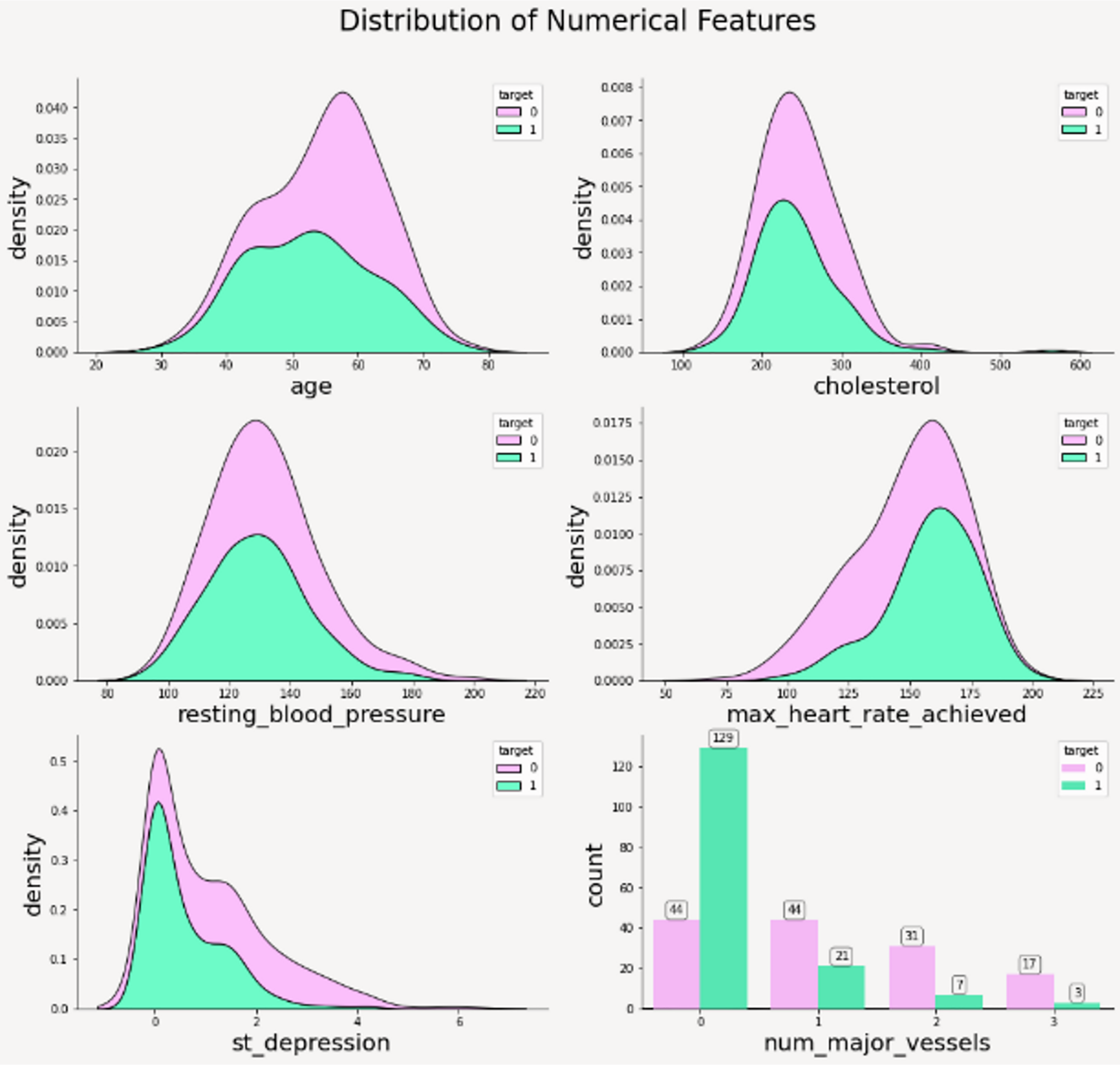


Figure 2 **Distribution analysis of numerical features in Cleveland Heart Disease dataset.** Full-size DOI: 10.7717/peerj-cs.2498/fig-2

the impact of previous positions on the current iteration, is set to decrease from 0.9 to 0.4. The adaptation speed of seagull agents is 0.7. The probability of mutation to maintain diversity is assigned as 0.05. The DBN model comprises five layers with four stacked Restricted Boltzmann Machine layers. The model is executed for 300 iterations with a maximum number of nodes set as 20 and a minimum number of nodes set as 2,000. The

Figure 3 Distribution analysis of categorical features in Cleveland Heart Disease dataset. Full-size DOI: 10.7717/peerj-cs.2498/fig-3

learning rate for the first, second, and third layers is 0.015. The learning rate for the final fourth layer of RBM is set to be 0.001. The value of the initial momentum is initialized with 0.5, and the regularization parameter, which prevents overfitting by penalizing large weights, is assigned as 0.001.

Experimental evaluation

The correlation matrix representing the relationship between each of the variables in the dataset is presented in Fig. 4. The inferences interpreted from the details in Fig. 5 are presented in Table 2.

From the correlation matrix obtained it can be inferred that chest pain type, maximum heart rate, ST slope, number of major vessels and thalassemia are the most important features that can predict the presence of heart diseases. These features have strong positive correlations with the target indicating that the individual is affected by heart disease. The features such as exercise-induced angina and oldpeak can also be considered as significant contributors as they are correlated with several other features. The performance of the existing and proposed techniques is assessed using metrics such as accuracy, precision, recall and F1-score.

For this research, accuracy, precision, recall, and F1-score are critical performance indicators since they offer a thorough assessment of the suggested QISOA-DBN model's efficacy in classifying heart diseases. While precision concentrates on the consistency of positive diagnoses while reducing false positives, accuracy assesses the overall correctness of forecasts. By minimising false negatives, recall evaluates the model's sensitivity in detecting actual cases of heart disease. Because false positives and false negatives can both have serious repercussions in medical applications, the F1-score, which is the harmonic mean of precision and recall, offers a balanced measurement.

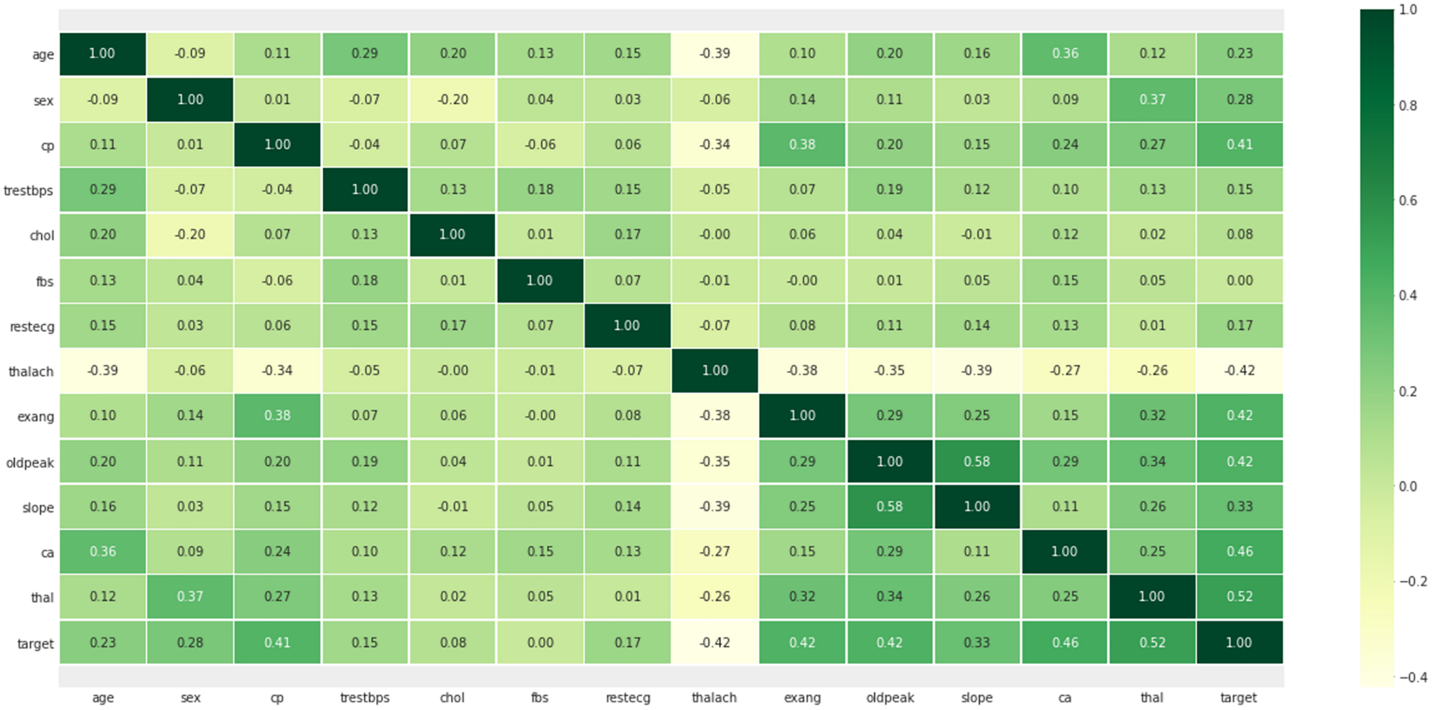
The confusion matrix is built using the predictions made by various models based on true positive/negative and false positive/negative values.

Initially, the performance of the DBN model is compared with the baseline machine learning and deep learning models such as support vector machines (SVM), Random Forest (RF), decision trees (DT), multilayer perceptrons (MLP), and fully convolutional networks (FCN).

Support vector machines (SVM): SVM is a supervised learning approach that divides observations into different categories by determining the best hyperplane. SVM is useful in heart disease research for binary classification and performs well with high-dimensional data, assisting in the discovery of correlations that point to the presence of heart disease.

Random Forest (RF): Several decision trees are constructed using the Random Forest ensemble learning technique, combining them to increase the model's predictability and reduce overfitting. RF handles large datasets with numerous attributes in heart disease research, yielding reliable insights into the significance of various risk indicators for heart disease.

Decision trees (DT): A decision tree is a straightforward yet effective method that divides data into branches depending on feature values to make predictions. Because of its



	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
age	1.00	-0.09	0.11	0.29	0.20	0.13	0.15	-0.39	0.10	0.20	0.16	0.36	0.12	0.23
sex	-0.09	1.00	0.01	-0.07	-0.20	0.04	0.03	-0.06	0.14	0.11	0.03	0.09	0.37	0.28
cp	0.11	0.01	1.00	-0.04	0.07	-0.06	0.06	-0.34	0.38	0.20	0.15	0.24	0.27	0.41
trestbps	0.29	-0.07	-0.04	1.00	0.13	0.18	0.15	-0.05	0.07	0.19	0.12	0.10	0.13	0.15
chol	0.20	-0.20	0.07	0.13	1.00	0.01	0.17	-0.00	0.06	0.04	-0.01	0.12	0.02	0.08
fbs	0.13	0.04	-0.06	0.18	0.01	1.00	0.07	-0.01	-0.00	0.01	0.05	0.15	0.05	0.00
restecg	0.15	0.03	0.06	0.15	0.17	0.07	1.00	-0.07	0.08	0.11	0.14	0.13	0.01	0.17
thalach	-0.39	-0.06	-0.34	-0.05	-0.00	-0.01	-0.07	1.00	-0.38	-0.35	-0.39	-0.27	-0.26	-0.42
exang	0.10	0.14	0.38	0.07	0.06	-0.00	0.08	-0.38	1.00	0.29	0.25	0.15	0.32	0.42
oldpeak	0.20	0.11	0.20	0.19	0.04	0.01	0.11	-0.35	0.29	1.00	0.58	0.29	0.34	0.42
slope	0.16	0.03	0.15	0.12	-0.01	0.05	0.14	-0.39	0.25	0.58	1.00	0.11	0.26	0.33
ca	0.36	0.09	0.24	0.10	0.12	0.15	0.13	-0.27	0.15	0.29	0.11	1.00	0.25	0.46
thal	0.12	0.37	0.27	0.13	0.02	0.05	0.01	-0.26	0.32	0.34	0.26	0.25	1.00	0.52
target	0.23	0.28	0.41	0.15	0.08	0.00	0.17	-0.42	0.42	0.42	0.33	0.46	0.52	1.00

Figure 4 **Correlation analysis of features in Cleveland Heart Disease dataset.** Full-size DOI: 10.7717/peerj-cs.2498/fig-4

interpretability, DTs are employed in heart disease research to determine which features—such as age and blood pressure—are most important in predicting the risk of heart disease.

Multilayer perceptrons (MLP): This model is a composition of multiple layers of neurons that are capable of learning intricate, non-linear correlations from data. This technique is effective in forecasting the existence or severity of heart disease, complex patterns, and interactions between different clinical and demographic data related to cardiovascular research.

Fully convolutional networks (FCN): This is a deep learning architecture mainly used to analyze sequential data in heart disease research. It is a valuable tool for identifying patterns in heart disease data because it can automatically extract features from raw data without requiring manual feature extraction.

The confusion matrix obtained based on the predictions made by each model is presented in Fig. 5.

Table 3 presents the performance metric values obtained according to the efficiency of the models in classifying heart diseases. It can be observed that the SVM model produces an accuracy of 92.35%, precision of 91.32%, recall of 90.53% and F1-score of 91.75%. RF model produces a lower accuracy of 83.21% compared to the SVM model. The precision, recall and F1 score values are also less, with 82.36%, 81.65% and 82.95%, respectively. The DT model shows the least performance with poor accuracy of 81.52% compared to the other baseline models. MLP model shows a better performance compared to SVM, RF and DT models with an average accuracy of 93.25%, precision of 92.45%, recall of 91.36% and

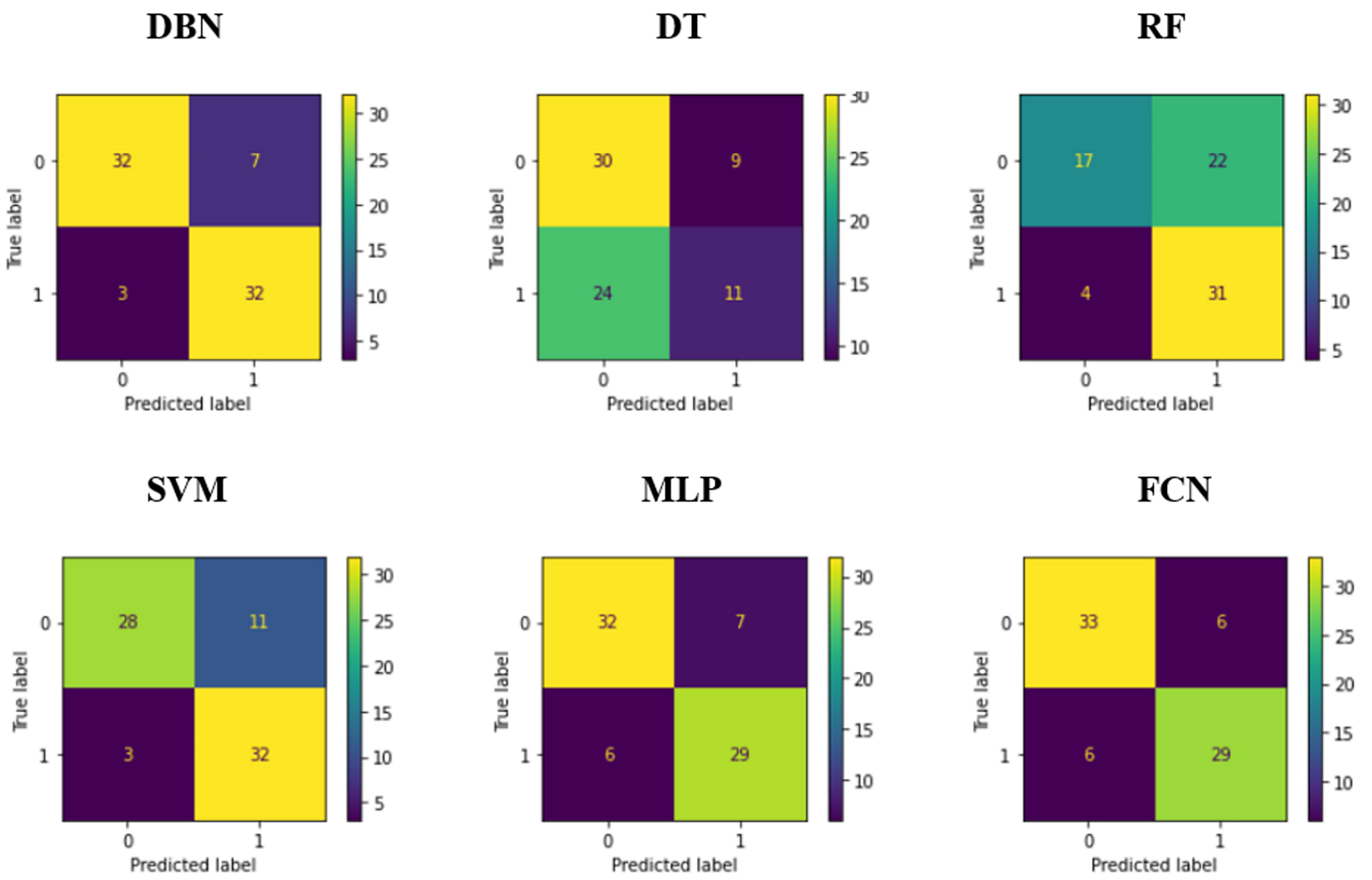


Figure 5 **Confusion matrix of ML and DL techniques.** Full-size DOI: 10.7717/peerj-cs.2498/fig-5

F1 score of 92.86%. The FCN model is 94.51% accurate in classifying cardiovascular disease with 93.31%, 91.64% and 92.98% of precision, recall and F1 score correspondingly. The DBN model is more precise and reliable than other traditional ML/DL models, with an accuracy level of 96.78% in predicting individuals with and without heart diseases. Figure 6 depicts the performance comparison of the baseline ML/DL models.

Furthermore, the performance of the baseline models, along with various optimization algorithms, is analyzed, and the corresponding confusion matrix is presented in Fig. 7. The optimization algorithms such as particle swarn optimization (PSO), genetic algorithm (GA), grey wolf optimization (GWO), cuckoo search optimization (CSO), crow search algorithm (CSA) and seagull optimization algorithm (SOA) are employed to optimize the parameters of the traditional ML/DL algorithms. The obtained values of various performance metrics are shown in Table 4. It can be seen that the performance of the SVM model optimized by PSO, the RF model optimized by GA and the DT model optimized by GWO algorithm are quite close to each other with accuracy levels of 93.89%, 94.12% and 94.36% respectively. This is because the total true positives and true negatives predicted by

Table 2 Attribute correlation description

Attribute	Correlation level	Correlated attribute
Age	Positively correlated	trestbps (0.29), chol (0.20), and ca (0.36)
	Negatively correlated	thalach (−0.39)
Sex	Negatively correlated	thalach (−0.34)
	Slight positive correlations	cp (0.11) and ca (0.09)
Chest pain	Positively correlated	thalach (0.38), exang (0.38), oldpeak (0.29), slope (0.25), and thal (0.37), target (0.41)
Resting blood pressure	Positively correlated	age (0.29) and chol (0.13)
	Negatively correlated	thalach (−0.07)
Cholesterol	Positively correlated	age (0.20), trestbps (0.13), and ca (0.15)
	Slightly negatively correlated	fbs (−0.20)
Fasting blood sugar	Negatively correlated	chol (−0.20) and slope (−0.16)
	Slight positive correlation	restecg (0.06)
Resting ECG	Slightly negatively correlated	chol (−0.07) and oldpeak (−0.05)
Maximum heart rate	Strongly negatively correlated	age (−0.39) and trestbps (−0.34)
	Positively correlated	cp (0.38) and slope (0.25)
Exercise induced angina	Positively correlated	cp (0.38), oldpeak (0.29), and ca (0.15)
	Strong negative correlation	thalach (−0.38)
ST depression	Positively correlated	cp (0.29), exang (0.29), and slope (0.58)
	Negatively correlated	thalach (−0.35)
ST slope	Positively correlated	cp (0.25), oldpeak (0.58), and target (0.33)
Number of major vessels	Positively correlated	age (0.36), chol (0.15), thal (0.37) and target (0.46)
Thalassemia	Positively correlated	cp (0.37), exang (0.26), ca (0.37) and target (0.52)

Table 3 Performance comparison of ML and DL techniques for cardiovascular disease prediction.

Techniques	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
SVM	92.35	91.32	90.53	91.75
RF	83.21	82.36	81.65	82.95
DT	81.52	79.63	78.35	80.62
MLP	93.25	92.45	91.36	92.86
FCN	94.51	93.31	91.64	92.98
DBN	96.78	95.46	93.98	95.12

these models are the same, with slight variations in the number of false positives as well as false negatives forecasted by these models.

It is also evident that the FCN and CSA combination produces the lowest accuracy of 92.78% with a precision of 91.23%, recall of 90.45%, and F1 score of 91.15%. MLP model optimized by the CSA technique produced a better accuracy of 95.24% with precision, recall, and F1 score values of 94.52%, 93.14%, and 94.16%, respectively.

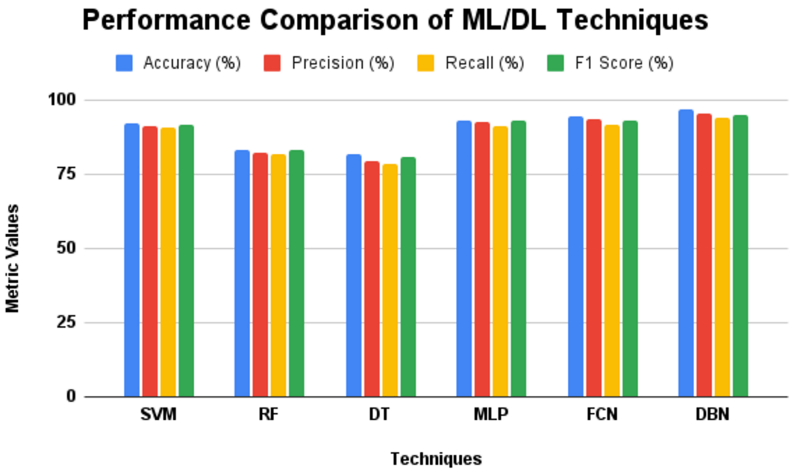


Figure 6 Performance comparison of ML/DL techniques.

Full-size DOI: 10.7717/peerj-cs.2498/fig-6

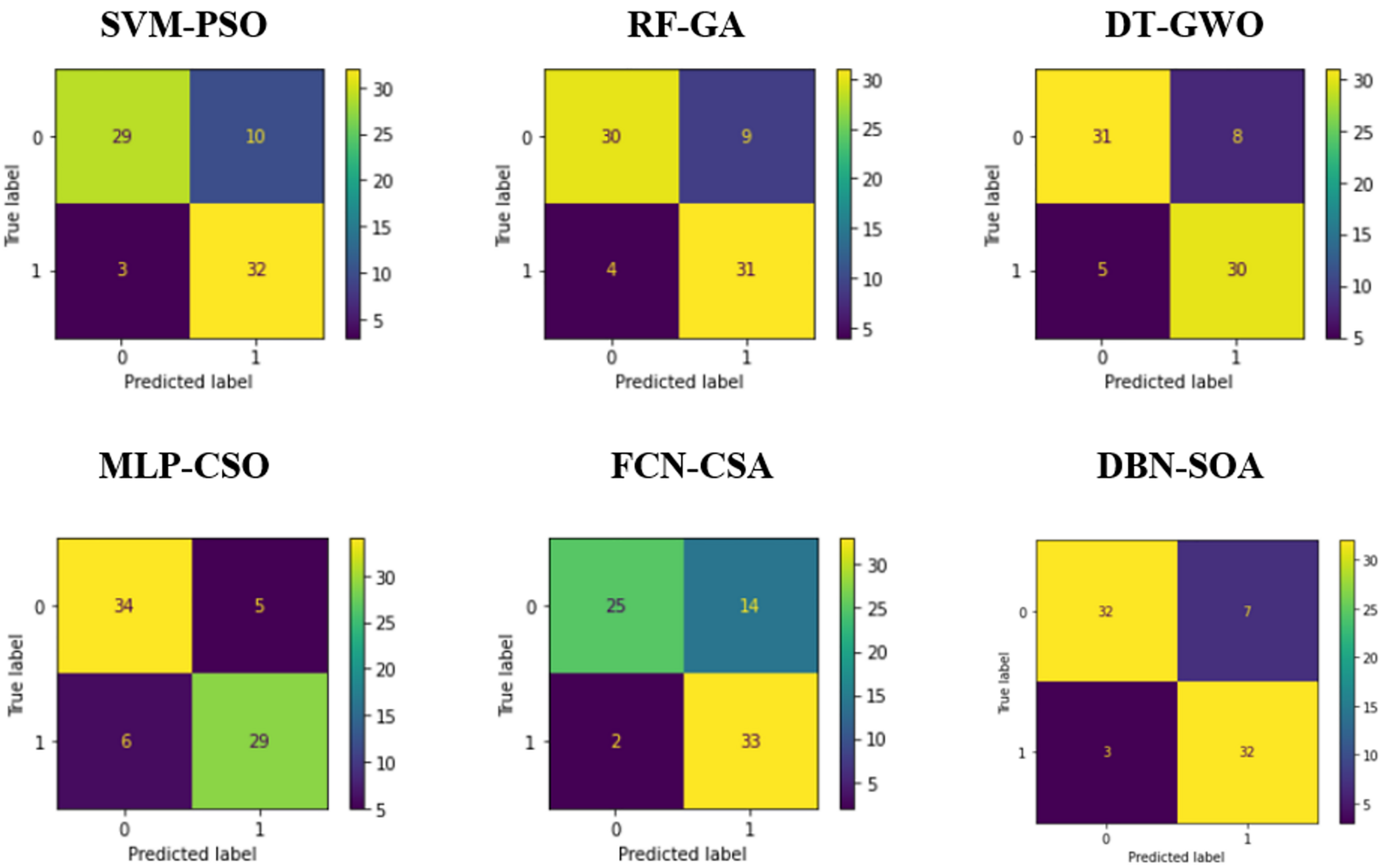


Figure 7 Confusion matrix of optimized ML and DL techniques.

Full-size DOI: 10.7717/peerj-cs.2498/fig-7

Table 4 **Performance comparison of optimized ML and DL techniques.**

Techniques	Optimization algorithms	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
SVM	PSO	93.89	92.56	91.35	92.12
RF	GA	94.12	93.45	92.51	93.92
DT	GWO	94.36	93.78	92.85	94.13
MLP	CSO	95.24	94.52	93.14	94.16
FCN	CSA	92.78	91.23	90.45	91.15
DBN	QISOA	98.6	97.6	96.8	97.1

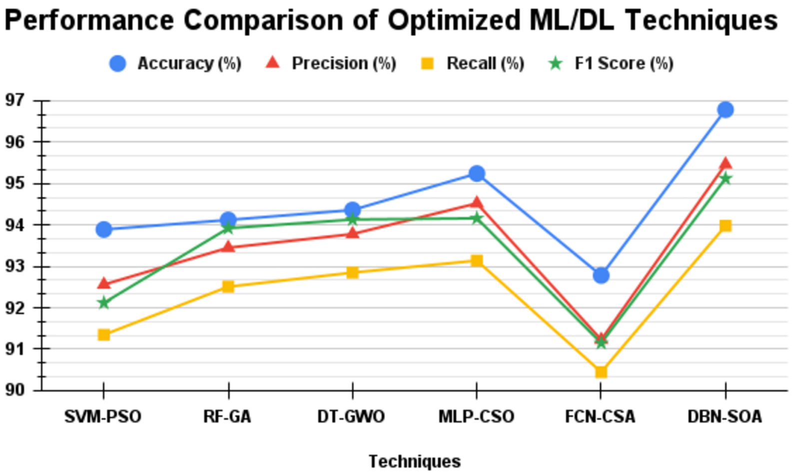


Figure 8 **Performance comparison of optimized ML/DL techniques.**
Full-size DOI: 10.7717/peerj-cs.2498/fig-8

DBN model optimized by QISOA produces an accuracy of 98.6%, precision of 97.6%, recall of 96.8% and F1 score of 97.1%. It has been observed that the optimized DBN model produces performance superior to other baseline models that are optimized using several optimization algorithms. The graphical depictions of the achieved outcomes are presented in Fig. 8.

Table 5 shows the performance of the proposed QISOA-DBN model, which is compared against a few existing works in the literature that focus on cardiovascular disease classification using quantum-based ML/DL or optimization methods. The work proposed in *Nandy et al. (2023)* uses a swarm artificial neural network to predict heart diseases with an accuracy of 91.6%, precision of 90.9%, recall of 90.1% and F1 score of 90.4%. The researchers in *Ullah et al. (2022)* have proposed a fully connected quantum convolutional neural network to classify individuals with and without heart disease and have achieved an accuracy of 95.3%. Quantum K-means clustering was suggested as a quantum machine learning algorithm for heart disease classification in *Kavitha & Kaulgud (2023)*, and the outcomes were found to be 93.5% accurate and precise.

The SVM model was optimized by a quantum-based PSO algorithm in *Tharwat & Hassanien (2019)* to produce a better accuracy of 97.6%, which is higher than the DBN

Table 5 Performance comparison of existing and proposed techniques

References	Techniques	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Nandy et al. (2023)	Swarm artificial neural network	91.6	90.9	90.1	90.4
Ullah et al. (2022)	Fully connected quantum convolutional neural network	95.3	94.2	93.6	94.2
Kavitha & Kaulgud (2023)	Quantum K-means clustering	93.5	92.2	91.7	92.7
Tharwat & Hassanien (2019)	QPSO-SVM	97.6	96.7	95.6	96.2
Abdulsalam, Meshoul & Shaiba (2023)	Ensemble-quantum machine learning	96.8	95.3	94.7	96.2
Proposed	QISOA-DBN	98.6	97.6	96.8	97.1

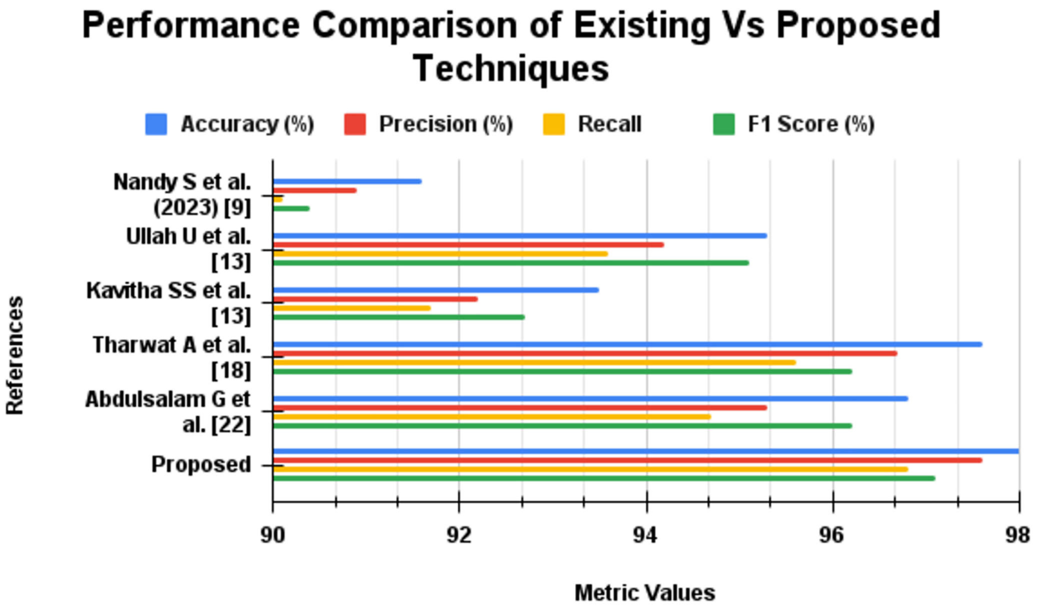


Figure 9 Performance comparison of existing *vs* proposed techniques.
Full-size DOI: 10.7717/peerj-cs.2498/fig-9

model used separately and optimized by the SOA model. However, it was clear that the QISOA-DBN model produced the highest accuracy of 98.6% in identifying and categorizing individuals with cardiovascular issues. The proposed model's precision, recall and F1-scores were also higher, with values such as 97.6%, 96.8% and 97.1%, respectively. The performance comparison existing and proposed methods is shown in Fig. 9.

Limitations of present research

The major limitation of the proposed research is that it can be challenging to deploy resource-intensive models for real-time prediction, particularly in settings with limited resources. Also, elevated computational capabilities result in amplified energy usage, which may constrain extensive or instantaneous applications. In addition to this, DBNs can grow quite complex and prone to overfitting, especially when optimized for high performance.

This is the case if the training data is insufficiently vast or heterogeneous. Depending on the particular disease under investigation, medical data can differ substantially. The patterns and aspects associated with cardiovascular diseases may not be appropriate for diseases like cancer or neurological problems. To obtain comparable performance, the QISOA-DBN model could, therefore, require significant retraining and feature modification. Highly complicated data types, including multi-omics data and histopathology images, are frequently used in cancer diagnosis. Compared to specialized models created especially for multi-modal medical data, the QISOA-DBN technique may find it challenging to integrate and handle these various data kinds, resulting in inadequate performance. Genomic and medical imaging data can include extremely high-dimensional characteristics. Quantum-inspired optimization may not scale well in such cases, resulting in slower processing times and higher computing costs, even though DBNs can manage such data to some extent.

CONCLUSION

Research on cardiovascular disease has advanced significantly in the last several years. While much research has been done on cardiac risk detection, most is inefficient and has numerous shortcomings. The present study proposes using a hybrid model, QISOA-DBN, to evaluate and classify the risk of heart disease. The advantages of QISOA and DBN algorithms are combined to improve the efficiency of cardiovascular disease prediction. The first phase in the proposed methodology is data preprocessing, which includes cleaning, transforming, and standardizing data. The second phase incorporates quantum principles in seagull optimization for a better convergence rate. The third phase employs the QISOA technique to set the parameter values for the DBN algorithm to enhance its efficacy in making classifications. The proposed QISOA-DBN model is applied to the Cleveland Heart Disease dataset, and the outcomes are compared with the existing works published in the literature. The suggested approach outperforms other models with accuracy, precision, recall, and F1 score values of 98.6%, 97.6%, 96.8%, and 97.1%, respectively. Future research on this work can be focused on examining reinforcement learning techniques combined with the proposed model to enhance the training of sequential and dynamic data, enabling it to assess the development of a disease and the effectiveness of treatment. The subsequent studies should also evaluate the QISOA-DBN model's performance on more extensive and varied datasets from various demographics and healthcare settings to confirm the robustness of the model.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the Korea Environmental Industry & Technology Institute (KEITI), with a grant funded by the Korean government, Ministry of Environment (The development of IoT-based technology for collecting and managing big data on environmental hazards and health effects), Grant RE202101551. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Korea Environmental Industry & Technology Institute (KEITI).
Korean Government, Ministry of Environment: RE202101551.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- D. Banumathy conceived and designed the experiments, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- T. Vetriselvi conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- K. Venkatachalam conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Jaehyuk Cho performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The Heart Disease dataset is available at UCI Machine Learning Repository: Janosi, A., Steinbrunn, W., Pfisterer, M., & Detrano, R. (1989). Heart Disease [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C52P4X.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj-cs.2498#supplemental-information.

REFERENCES

Abdulsalam G, Meshoul S, Shaiba H. 2023. Explainable heart disease prediction using ensemble-quantum machine learning approach. *Intelligent Automation & Soft Computing* **36(1)**:761–779 DOI 10.32604/iasc.2023.032262.

Agrawal R, Kaur B, Agarwal P. 2021. Quantum inspired particle swarm optimization with guided exploration for function optimization. *Applied Soft Computing* **102(1)**:107122 DOI 10.1016/j.asoc.2021.107122.

Akella A, Akella S. 2021. Machine learning algorithms for predicting coronary artery disease: efforts toward an open source solution. *Future Science OA* **7(6)**:FSO698 DOI 10.2144/fsoa-2020-0206.

Al-Tashi Q, Rais H, Jadid S. 2019. Feature selection method based on grey wolf optimization for coronary artery disease classification. In: *Recent Trends in Data Science and Soft Computing: Proceedings of the 3rd International Conference of Reliable Information and Communication Technology (IRICT 2018)*. Berlin, Germany: Springer, 257–266.

Alsubai S, Alqahtani A, Binbusayyis A, Sha M, Gumaei A, Wang S. 2023. Heart failure detection using instance quantum circuit approach and traditional predictive analysis. *Mathematics* **11(6)**:1467 DOI 10.3390/math11061467.

Asif D, Bibi M, Arif MS, Mukheimer A. 2023. Enhancing heart disease prediction through ensemble learning techniques with hyperparameter optimization. *Algorithms* **16(6)**:308 DOI 10.3390/a16060308.

AvinashPrabhu S, Parthipan V. 2022. A novel approach for prediction of human disease using symptoms by multilayer perceptron algorithm to improve accuracy and compared with k nearest neighbor algorithm. *Journal of Pharmaceutical Negative Results* **13**:329–335 DOI 10.47750/pnr.2022.13.S04.037.

Bharti R, Khamparia A, Shabaz M, Dhiman G, Pande S, Singh P. 2021. Prediction of heart disease using a combination of machine learning and deep learning. *Computational Intelligence and Neuroscience* **2021(1)**:8387680 DOI 10.1155/2021/8387680.

Bing P, Liu Y, Liu W, Zhou J, Zhu L. 2022. Electrocardiogram classification using TSST-based spectrogram and ConViT. *Frontiers in Cardiovascular Medicine* **9**:983543 DOI 10.3389/fcvm.2022.983543.

Cai L, Fang H, Xu N, Ren B. 2024. Counterfactual causal-effect intervention for interpretable medical visual question answering. *IEEE Transactions on Medical Imaging* **43(12)**:4430–4441 DOI 10.1109/TMI.2024.3425533.

Chen Z, Liang Q, Wei Z, Chen X, Shi Q, Yu Z, Sun T. 2023. An overview of in vitro biological neural networks for robot intelligence. *Cyborg and Bionic Systems* **4(8)**:0001 DOI 10.34133/cbsystems.0001.

Chicco D, Jurman G. 2020. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC Medical Informatics and Decision Making* **20(1)**:1–16 DOI 10.1186/s12911-020-1023-5.

Duan X, Xie D, Zhang R, Li X, Sun J, Qian C, Song X, Li C. 2023. A novel robotic bronchoscope system for navigation and biopsy of pulmonary lesions. *Cyborg and Bionic Systems* **4(9)**:0013 DOI 10.34133/cbsystems.0013.

Enad HG, Mohammed MA. 2023. 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 and Applications* **11(1)**:08–25 DOI 10.54216/FPA.110101.

García-Ordás MT, Bayón-Gutiérrez M, Benavides C, Aveleira-Mata J, Benítez-Andrades JA. 2023. Heart disease risk prediction using deep learning techniques with feature augmentation. *Multimedia Tools and Applications* **82(20)**:31759–31773 DOI 10.1007/s11042-023-14817-z.

Ghosh P, Azam S, Jonkman M, Karim A, Shamrat FJM, Ignatious E, Shultana S, Beeravolu AR, De Boer F. 2021. Efficient prediction of cardiovascular disease using machine learning algorithms with relief and lasso feature selection techniques. *IEEE Access* **9**:19304–19326 DOI 10.1109/ACCESS.2021.3053759.

Huang H-H, Shu J, Liang Y. 2024. MUMA: a multi-omics meta-learning algorithm for data interpretation and classification. *IEEE Journal of Biomedical and Health Informatics* **28(4)**:2428–2436 DOI 10.1109/JBHI.2024.3363081.

Huang Y, Wang C, Zhou T, Xie F, Liu Z, Xu H, Liu M, Wang S, Li L, Chi Q. 2024. Lumican promotes calcific aortic valve disease through h3 histone lactylation. *European Heart Journal* **45(37)**:ehae407 DOI 10.1093/eurheartj/ehae407.

Huo R, Liu Y, Xu H, Li J, Xin R, Xing Z, Deng S, Wang T, Yuan H, Zhao X. 2022. Associations between carotid atherosclerotic plaque characteristics determined by magnetic resonance imaging and improvement of cognition in patients undergoing carotid endarterectomy. *Quantitative Imaging in Medicine and Surgery* **12(5)**:2891–2903 DOI 10.21037/qims-21-981.

Kavitha S, Kaulgud N. 2023. Quantum k-means clustering method for detecting heart disease using quantum circuit approach. *Soft Computing* **27(18)**:13255–13268 DOI 10.1007/s00500-022-07200-x.

Kishor A, Chakraborty C. 2022. Artificial intelligence and internet of things based healthcare 4.0 monitoring system. *Wireless Personal Communications* **127(2)**:1615–1631 DOI 10.1007/s11277-021-08708-5.

Kumar Y, Koul A, Sisodia PS, Shafi J, Verma K, Gheisari M, Davoodi MB. 2021. Heart failure detection using quantum-enhanced machine learning and traditional machine learning techniques for internet of artificially intelligent medical things. *Wireless Communications and Mobile Computing* **2021(1)**:1616725 DOI 10.1155/2021/1616725.

Lahori M, Dehghani A, Wilson C, Law W, Agaram N, Murali R, Sigel C. 2023. Cytopathologic features of epithelioid hemangioendothelioma including touch imprints for rapid on-site evaluation. *Cytojournal* **20**:29 DOI 10.25259/Cytojournal.

Li Q, You T, Chen J, Zhang Y, Du C. 2023. LI-EMRSQL: linking information enhanced Text2SQL parsing on complex electronic medical records. *IEEE Transactions on Reliability* **73(2)**:1280–1290 DOI 10.1109/TR.2023.3336330.

Luan S, Yu X, Lei S, Ma C, Wang X, Xue X, Ding Y, Ma T, Zhu B. 2023. Deep learning for fast super-resolution ultrasound microvessel imaging. *Physics in Medicine & Biology* **68(24)**:245023 DOI 10.1088/1361-6560/ad0a5a.

Maheshwari D, Ullah U, Marulanda PAO, Jurado AG-O, Gonzalez ID, Merodio JMO, Garcia-Zapirain B. 2023. Quantum machine learning applied to electronic healthcare records for ischemic heart disease classification. *Human-centric Computing and Information Science* **13(06)**:1–17 DOI 10.22967/HCIS.2023.13.006.

Mehmood A, Iqbal M, Mehmood Z, Irtaza A, Nawaz M, Nazir T, Masood M. 2021. Prediction of heart disease using deep convolutional neural networks. *Arabian Journal for Science and Engineering* **46(4)**:3409–3422 DOI 10.1007/s13369-020-05105-1.

Munshi M, Gupta R, Jadav NK, Polkowski Z, Tanwar S, Alqahtani F, Said W. 2024. Quantum machine learning-based framework to detect heart failures in healthcare 4.0. *Software: Practice and Experience* **54(2)**:168–185 DOI 10.1002/spe.3264.

Nagarajan SM, Muthukumaran V, Murugesan R, Joseph RB, Meram M, Prathik A. 2022. Innovative feature selection and classification model for heart disease prediction. *Journal of Reliable Intelligent Environments* **8(4)**:333–343 DOI 10.1007/s40860-021-00152-3.

Nandy S, Adhikari M, Balasubramanian V, Menon VG, Li X, Zakarya M. 2023. An intelligent heart disease prediction system based on swarm-artificial neural network. *Neural Computing and Applications* **35(20)**:14723–14737 DOI 10.1007/s00521-021-06124-1.

Ozpolat Z, Karabatak M. 2023. Performance evaluation of quantum-based machine learning algorithms for cardiac arrhythmia classification. *Diagnostics* **13(6)**:1099 DOI 10.3390/diagnostics13061099.

Priya L, VinilaJinny S, Mate Y. 2020. Early prediction model for coronary heart disease using genetic algorithms, hyper-parameter optimization and machine learning techniques. *Health Technology* **11**:63–73.

Rao JN, Prasad RS. 2021. An enhanced novel dynamic data processing (enddp) algorithm for predicting heart disease in machine learning. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology* **7**:2456–3307 DOI 10.32628/IJSRCSEIT.

Reddy KVV, Elamvazuthi I, Aziz AA, Paramasivam S, Chua HN, Pranavanand S. 2021. Heart disease risk prediction using machine learning classifiers with attribute evaluators. *Applied Sciences* **11(18)**:8352 DOI 10.3390/app11188352.

Saad HM, Chakrabortty RK, Elsayed S, Ryan MJ. 2021. Quantum-inspired genetic algorithm for resource-constrained project-scheduling. *IEEE Access* **9**:38488–38502 DOI 10.1109/ACCESS.2021.3062790.

Sarra RR, Dinar AM, Mohammed MA, Ghani MKA, Albahar MA. 2022. A robust framework for data generative and heart disease prediction based on efficient deep learning models. *Diagnostics* **12(12)**:2899 DOI 10.3390/diagnostics12122899.

Sharma P, Sharma K. 2022. A novel quantum-inspired binary bat algorithm for leukocytes classification in blood smear. *Expert Systems* **39(3)**:e12813 DOI 10.1111/exsy.12813.

Sun T, Lv J, Zhao X, Li W, Zhang Z, Nie L. 2023. In vivo liver function reserve assessments in alcoholic liver disease by scalable photoacoustic imaging. *Photoacoustics* **34(8)**:100569 DOI 10.1016/j.pacs.2023.100569.

Tharwat A, Hassanien AE. 2018. Chaotic antlion algorithm for parameter optimization of support vector machine. *Applied Intelligence* **48(3)**:670–686 DOI 10.1007/s10489-017-0994-0.

Tharwat A, Hassanien AE. 2019. Quantum-behaved particle swarm optimization for parameter optimization of support vector machine. *Journal of Classification* **36(3)**:576–598 DOI 10.1007/s00357-018-9299-1.

Ullah U, Jurado AGO, Gonzalez ID, Garcia-Zapirain B. 2022. A fully connected quantum convolutional neural network for classifying ischemic cardiopathy. *IEEE Access* **10(2)**:134592–134605 DOI 10.1109/ACCESS.2022.3232307.

Wang J, Rao C, Goh M, Xiao X. 2023a. Risk assessment of coronary heart disease based on cloud-random forest. *Artificial Intelligence Review* **56(1)**:203–232 DOI 10.1007/s10462-022-10170-z.

Wang Y, Wang W. 2021. Quantum-inspired differential evolution with grey wolf optimizer for 0-1 knapsack problem. *Mathematics* **9(11)**:1233 DOI 10.3390/math9111233.

Wang Y, Zhai W, Zhang H, Cheng S, Li J. 2023b. Injectable polyzwitterionic lubricant for complete prevention of cardiac adhesion. *Macromolecular Bioscience* **23(4)**:2200554 DOI 10.1002/mabi.202200554.

Yildirim O, Baloglu UB, Talo M, Ganesan P, Tung JS, Kang G, Tooley J, Alhusseini MI, Baykaner T, Wang PJ. 2021. Deep neural network trained on surface ECG improves diagnostic accuracy of prior myocardial infarction over Q wave analysis. In: *2021 Computing in Cardiology (CinC)*. Piscataway: IEEE, 1–4 DOI 10.23919/CinC53138.2021.9662825.

Yu X, Luan S, Lei S, Huang J, Liu Z, Xue X, Ma T, Ding Y, Zhu B. 2023. Deep learning for fast denoising filtering in ultrasound localization microscopy. *Physics in Medicine & Biology* **68(20)**:205002 DOI 10.1088/1361-6560/acf98f.

Zhang C, Ge H, Zhang S, Liu D, Jiang Z, Lan C, Li L, Feng H, Hu R. 2021. Hematoma evacuation via image-guided para-corticospinal tract approach in patients with spontaneous intracerebral hemorrhage. *Neurology and Therapy* **10(2)**:1001–1013 DOI 10.1007/s40120-021-00279-8.

Zhao Y, Hu J, Sun X, Yang K, Yang L, Kong L, Zhang B, Li F, Li C, Shi B. 2021. Loss of m6A demethylase ALKBH5 promotes post-ischemic angiogenesis via post-transcriptional stabilization of WNT5A. *Clinical and Translational Medicine* **11(5)**:e402 DOI 10.1002/ctm2.402.

Zhao Y, Xiong W, Li C, Zhao R, Lu H, Song S, Zhou Y, Hu Y, Shi B, Ge J. 2023. Hypoxia-induced signaling in the cardiovascular system: pathogenesis and therapeutic targets. *Signal Transduction and Targeted Therapy* **8(1)**:431 DOI 10.1038/s41392-023-01652-9.

Zhou X, Lu J, Wu B, Guo Z. 2022. HOXA11-AS facilitates the proliferation, cell cycle process and migration of keloid fibroblasts through sponging miR-188–5p to regulate VEGFA. *Journal of Dermatological Science* **106(2)**:111–118 DOI 10.1016/j.jdermsci.2022.04.004.