Received: 31 December 2024 | Revised: 12 May 2025 | Accepted: 30 May 2025 | Published online: 1 September 2025

RESEARCH ARTICLE

Journal of Computational and Cognitive Engineering
2025, Vol. 00(00) 1–12
DOI: 10.47852/bonviewJCCE52025138

BON VIEW PUBLISHING

QFM-BioPred: Quantum Fusion Model for Bioactivity Prediction in Cardiovascular Disease Drug Discovery

Gundala Pallavi[1], Ali Altalbe[2], and Prasanna Kumar Rangarajan[1,*]

[1] *Department of Computer Science and Engineering, Amrita Vishwa Vidyapeetham-Chennai, India*

[2] *Faculty of Computing and Information Technology, King Abdulaziz University, Saudi Arabia*

Abstract: Cardiovascular diseases (CVDs) are the leading cause of mortality worldwide, highlighting the urgent need for more effective treatments. The conventional drug discovery process is time-consuming and expensive; therefore, new approaches are required. Quantum machine learning in compound bioactivity prediction has been demonstrated in drug discovery, but its application in cardiovascular medicine remains limited. Therefore, this work proposed the quantum fusion model (QFM) to enhance the bioactivity predictions for heart disease treatments. The proposed model encoded molecular data into quantum states using the quantum random forest method on the ChEMBL dataset. Logistic regression classifiers were then trained on these encoded data. The QFM, which integrates quantum-inspired algorithms with classical machine learning, achieved an accuracy of 92.7% in classifying bioactive compounds, outperforming individual models and existing methods. It also demonstrated strong precision (0.92), recall (0.93), and F1 score of 0.92, with receiver operating characteristic area under the curve (AUC) and precision-recall AUC values of 0.961 and 0.959, respectively. These results indicate the model's ability to identify complex molecular structures accurately. This work advances bioactivity prediction to aid drug development for CVDs and aligns with the United Nations Sustainable Development Goal 3: Good Health and Well-being. Future research will apply this approach to other diseases and incorporate more complex quantum circuits to enhance accuracy further.

Keywords: quantum computing, machine learning, drug discovery, bioactivity prediction, cardiovascular disease, heart targets

1. Introduction

Cardiovascular diseases (CVDs) stand as the second most common reason for global deaths because they cause 17.9 million annual fatalities and contribute to 31% of worldwide mortality. The advanced state of medical research faces significant hurdles from the intricate genetic and lifestyle, and environmental relationships of CVDs. The medical industry spends over 10 years and large finances to develop a single therapeutic product through its current drug discovery approaches for CVDs. Research on innovative drug discovery methods has become essential because it enables to creation of potential treatments for CVDs [1].

Drug discovery includes a prediction of bioactivity as its fundamental step, because the assessment of molecular features helps predict how compounds will impact biological systems. The prediction of bioactivity has incorporated machine learning (ML) approaches in recent times because this method allows processing extensive data sets to identify patterns for drug-like compound detection. However, traditional ML methods struggle with the nonlinear and complex molecular interactions inherent to drug discovery [2]. These situations limit the efficacy of prediction models and their ability to scale up because of the large-scale and complex molecular data.

Quantum computing (QC), therefore, has been developed as a solution to these challenges. Superposition and entanglement of quantum bits in quantum ML (QML) enable more efficient searching of high-dimensional spaces than in the case of classical algorithms. This capability makes QML an attractive approach for dealing with the difficulties of drug discovery. Although QML is now in its infancy with the current implementations that are limited by noisy intermediate-scale quantum (NISQ) devices, it has already evidenced the ability to eliminate computational barriers in bioinformatics and drug design.

This work proposes the quantum fusion model (QFM) as a new approach to integrate quantum and classical ML into bioactivity prediction for cardiovascular drugs. This proposed model applies quantum random forest (QRF) to translate molecular data into quantum states, followed by the use of classical logistic regression classifiers. It is this hybrid approach that allows the capturing of more complex interactions between atoms within the molecule, thereby increasing the accuracy of bioactivity prediction. In testing, the QFM produced a classification accuracy of 92.7% and thus provided stronger differentiation of bioactive from inactive compounds as compared to conventional ML methods.

1.1. Research contributions

This work contributes to computational drug discovery in the following major ways:

1. This work introduces QFM-BioPred, a novel framework that combines quantum-inspired and ML models to perform a computation on CVD targets.
2. This study presents a novel quantum circuit, unlike regular encoding, which uses an angle encoding and entanglement-based feature interaction to encode molecular descriptors in high high-dimensional space.

***Corresponding author:** Prasanna Kumar Rangarajan, Department of Computer Science and Engineering, Amrita Vishwa Vidyapeetham-Chennai, India. Email: r_prasannakumar@ch.amrita.edu

3. The amplified quantum attributes in combination with ensemble model prediction in a stacked framework with a logistic regression meta-learner contribute to the improvement in prediction accuracy and generalization capacity, which demonstrates 92.7% accuracy and 0.961 receiver operating characteristic (ROC) area under the curve (AUC), which indicates its robust performance on cardiovascular datasets despite their imbalance.
4. The study supports United Nations Sustainable Development Goal 3: Good Health and Well-being through its development of computational works that fasten and decrease drug creation expenses for necessary diseases such as CVDs.

QML technology provides high-dimensional feature analysis and nonlinear molecular modeling that process bioactivity prediction structures. The proposed model uses Hilbert space to analyze molecular descriptors, including LogP, molecular weight (MW), and hydrogen bond interaction, for structure-activity relationship discovery by the model. Quantum entanglement mechanisms within QML models establish better relationship detection, which leads to outstanding prediction outcomes [3]. The system regulates complex data patterns by using circuits and angle embedding, which eliminates the requirement for conventional kernel function development. Quantum states provide an effective solution for controlling intricate molecular data structures, which become crucial during medical drug development research [4].

The application of QML surpasses traditional ML for cardiovascular drug discovery because target interactions show multifactorial behavior while being sensitive to small alterations in molecular structure [5]. The method helps discover active compounds during early stages at higher accuracy levels, thus reducing experimental expenses and optimizing lead compounds [6].

The remainder of this study is structured as follows: Section 2 presents the state of the art of computational drug discovery and the use of ML and QML for bioactivity prediction. Section 3 of the study provides details on the method and experimental design of the QFM's creation. The experiments and the comparison with traditional methods are provided in Section 4 of the paper. Section 5 contains the conclusion, limitations, and recommendations for future research.

2. Literature Review

The conventional process of drug discovery is slow and expensive, taking over 10 years and several billion dollars to get one drug to the market [7]. These methods are largely based on high-throughput screening and experimental testing, which are time-consuming and liable to contain errors. To deal with these drawbacks, more and more researchers are using computational methods to enhance and facilitate the drug discovery process [8].

2.1. Traditional drug discovery and ML in bioactivity prediction

The method of discovering new drugs has advanced over the last few years, and one of the most useful techniques is using ML to identify new drug molecules. Molecular informatics is useful for target identification, lead optimization, and the optimization of drug-like properties [9]. Some of the conventional methods [10], like random forest (RF) and support vector machine (SVM), are found to be effective in explaining large chemical and biological data sets [11]. But, all these methods fail when the molecular structures are complex and the datasets are high-dimensional, in which case feature extraction and representation become challenging.

Recent work by Rodríguez-Pérez and Bajorath [12] proposed a C-SVM algorithm that successfully identifies dual inhibitors targeting cancer kinase targets. It performs better in terms of false positive detection and equivalent true positive detection than other ML approaches. Support vector regression (SVR) models, which stem from SVM, have a known weakness in that they underestimate the potency of very potent compounds, which can distort the bioactivity landscape. The C-SVM models proved capable of identifying 429 compounds that interact with 24 different kinases for drug discovery chemoinformatics applications.

Mao et al. [13] provided an iterative methodology that combines wet-lab work with MD simulations, combining ML and deep learning to boost QSAR models. This framework demonstrates model training through unstructured data combination with structured data, although it fails to disclose the exact datasets utilized. This approach utilizes a system that improves prediction accuracy and accelerates drug development by resolving QSAR method shortcomings in versatility and accuracy.

A study by Siddiqui et al. [14] used a DPP-4 inhibitor dataset for a proposed RF-based predictive model, which gave the highest accuracy than the other models. This research analysis revealed that Murcko scaffold-based data splitting provided better reliability while reducing chemical bias compared to random split training approaches. These are the studies that provide the information regarding the previous studies [15].

2.2. QC and QML in drug discovery

QC presents solutions to the problems that classical ML faces by incorporating elements of quantum mechanics, including superposition and entanglement [16]. These principles enable quantum algorithms to perform computations on high-dimensional spaces that cannot be analyzed by classical systems. QC has much potential for drug discovery since it targets two major issues associated with bioactivity prediction, namely the optimization and patterns [17]. QML is a combination of QC and ML, and its application is in enhancing the predictive models of drug discovery. QML can transform classical molecular descriptors into quantum states and thus explain the relationships that classical models fail to explain.

However, current quantum devices, known for their NISQ systems, face challenges with noise and error rates [18]. However, there are some limitations of QC, and still, it has been applied in certain domains, including protein folding, molecular docking, and QSAR modeling [19]. Variational quantum eigensolver and quantum approximate optimization algorithm have been used to improve the model's performance [20, 21].

2.3. Hybrid quantum-classical approaches in bioinformatics

The use of quantum and classical algorithms is intended to provide the benefits of both approaches while avoiding the weaknesses inherent in each. To realize this, data encoding is done via quantum methods, while the fine-tuning is performed by classical ML techniques, allowing the models to search large feature spaces to improve predictive power. Such combinations have been helpful in drug discovery, especially when it comes to deriving interaction models for drugs and proteins and estimating their bioactivity. QML is a process that accelerates the generation and screening of drugs cost-effectively. For instance, Bhatia et al. [22] developed a framework of combining classical support vector classifiers with quantum kernel-based methods that yielded higher AUC ROC values of 0.80–0.95 in the simulation studies. A framework proposed by Mensa et al. [23] used SVC with a quantum kernel using a COVID-19 dataset for ligand-based virtual screening, proving quantum advantage for target-specific datasets. A hybrid quantum neural network

utilized the Genomics of Drug Sensitivity in Cancer dataset for building a prediction model to determine drug responses in cancer patients. The combination of deep learning with ML models gained a 15% increase in the predictive performance. The research did not present detailed, explicit limitations; however, it implied challenges stemming from quantum hardware constraints together with variations in cancer data [24].

Quantum SVMs and quantum kernel-based neural networks, as part of the more complex quantum hybrid models [25], are now proving to have better bioactivity prediction as compared to the traditional methods. Most of them are still in the proof-of-concept phase, but they show a great deal of potential for future drug discovery projects. However, Table 1 provides previous studies of drug discovery, but the application of hybrid quantum-classical methods is not toward cardiovascular drug discovery in particular, which is the gap that this paper seeks to fill.

2.4. Research gap and motivation

QML to drug discovery. Nevertheless, the use of QML for bioactivity prediction of CVDs is still negligible. The majority of the past studies investigated less complex multi-pathway situations or diseases that are based on fewer pathways [26]. Because CVDs are inherently multiscale, they offer a perfect chance for synergistically combining quantum and classical ML approaches to enhance bioactivity predictions and speed up the identification of therapeutic compounds. QML technique in cardiovascular drug discovery is still in its early stage; however, the methodologies and approaches discussed in the recent studies are generic. Due to the complexity and massive data processing of the CVDs, the field is ripe for quantum and quantum-inspired ML. These approaches can manage the structural-activity relationships, which are complex and not taken into account by the basic ML models.

To this end, the QFM is proposed to fill this gap. The QFM combines the quantum-based data encoding with the classical classification methods and the QRF for feature encoding and logistic regression for the prediction. Therefore, the QFM, which encodes molecular descriptors as quantum states, achieves a high level of structural-activity modeling of cardiovascular-targeted compounds and creates a basis for the subsequent development of quantum-driven drug discovery tools.

3. Data Acquisition

In the current study, we employed data from the ChEMBL database [27], which is a large bioinformatics database containing information on the bioactivities of billions of compounds tested against numerous targets. ChEMBL was selected due to its relevance and accuracy in providing data on compounds that target cardiac proteins that are critical for cardiovascular drug discovery. This guarantees that the dataset employed in our model is accurate and relevant for bioactivity prediction in therapeutic compounds for cardiac diseases.

3.1. Dataset characteristics

The dataset includes molecular descriptors, which serve as features for the predictive model are given in Table 2.

Table 1
Previous studies

Study	Methodology	Application	Dataset used	Performance metrics	Key findings	Limitations
Ashraf et al. [5]	ML with neural networks	SARS-CoV-2 bioactivity prediction	Custom SARS-CoV-2 data	Accuracy, precision, F1 score	Neural networks are effective in identifying bioactive compounds for SARS-CoV-2	Limited in capturing nonlinear dependencies in molecular structures
Chenthamarakshan et al. [9]	Deep generative models	Drug target inhibitor discovery	Drug inhibitor dataset	Recall, F1 score	Generated novel candidate molecules; effective in deep generative modeling	High computational cost, challenging to interpret generated features
Vamathevan et al. [10]	Random Forest and Deep Learning	Drug discovery (various applications)	Large biological and chemical datasets	Accuracy, Precision, Recall	Demonstrated deep learning and RF effectiveness across large-scale datasets	Limited generalizability without extensive feature engineering
Shi [11]	Support vector regression (SVR)	QSAR modeling for antioxidant activity	Custom antioxidant compounds	R^2, Mean absolute error	Effective at capturing high-dimensional data patterns, SVR is effective for QSAR	Requires careful hyperparameter tuning, sensitive to data variability
Rodríguez-Pérez and Bajorath [12]	Support vector machine and regression	Chemoinformatics and drug discovery	Molecular descriptors datasets	Accuracy, precision	Robust predictive modeling of bioactive compounds in large datasets	Limited scalability for highly nonlinear bioactivity relationships
Bhatia et al. [22]	Quantum ML (QML) for ADME-Tox prediction	ADME-Tox properties in drug screening	SMILES-based ADME-Tox	AUC ROC, accuracy	QML-based quantum kernel outperformed classical models in predicting ADME-Tox.	NISQ device limitations restrict scalability and real-world applications
Lau et al. [26]	Quantum-Machine Learning	General drug-protein interactions	Simulated drug-protein data	Accuracy, recall	Showed potential for quantum methods to explore high-dimensional feature spaces	Quantum methods are still largely theoretical, limited by current hardware.

Table 2
Dataset Features

Feature	Description
Molecule ChEMBL ID	Unique identifier for each compound.
SMILES	Text representation of molecular structures.
Bioactivity Class	Categorized as "active" or "inactive."
Molecular Weight (MW)	Indicates solubility and bioavailability.
Logarithm of Partition Coefficient (LogP)	Represents hydrophobicity and membrane permeability.
Number of Hydrogen Bond Donors (NumHDonors)	Impacts binding with biological targets.
Number of Hydrogen Bond Acceptors (NumHAcceptors)	Affects molecular interactions with targets.

It includes the molecular descriptors of the dataset, such as MW, LogP, NumHDonors, and NumHAcceptors are prominent in determining the level of bioactivity of the compounds. For instance, the LogP measures the hydrophobicity characteristic, which is fundamental to the bioactivity of a substance, and the MW affects the solubility and bioavailability of the substance. The dependent variable, Bioactivity Class, distinguishes compounds between "active" and "inactive."

3.2. Data collection process

The data collection process involves systematic steps, which include:

1. Accessing ChEMBL: Bioactivity data mining through the ChEMBL query tool.
2. Defining search criteria: Concentrating on compounds with certain IC50 values associated with cardiac targets and high assay reliability.
3. Filtering and quality control: The first step in cleaning the data is to delete any records that contain such values as missing or low-quality data.

4. Methodology

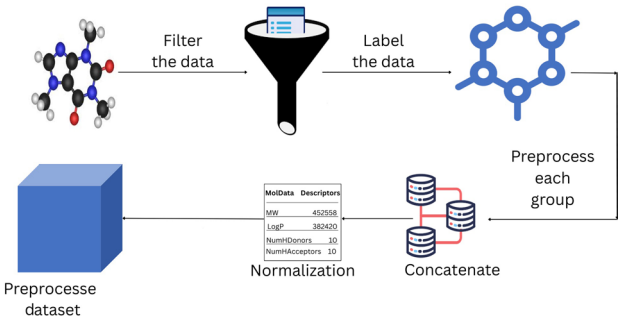
This research develops the QFM to integrate quantum-inspired and conventional ML models for better bioactivity prediction in cardiovascular drugs. This work proceeds from data preprocessing to quantum encoding of data and model development before implementing stacking classifiers for conventional and quantum models. The proposed model integrates classical and QC to build an operational model through these sequential steps

4.1. Data preprocessing

The QFM model requires good-quality input data, which preprocessing ensures according to Figure 1. The following data preparation procedure converted the ChEMBL dataset into a form suitable for bioactivity prediction:

1. Eliminating incomplete records: The dataset underwent cleaning procedures to remove entries that contained errors or data omissions, such as improper SMILES strings or bioactivities, because of their impact on data integrity. Computational software depends on SMILES strings as its fundamental element to process and handle chemical structures.
2. Filtration based on molecular descriptors: Compounds having MW or logP values outside permitted ranges were excluded. The selection process aims to eliminate compounds that would not perform well during target interaction with heart-related proteins.
3. Feature normalization: The model required all molecular descriptors to maintain proper weighting, so we normalized all numerical features, including MW, LogP, NumHDonors, and NumHAcceptors, between 0 and 1 values. The process intends to keep features uniform for model training purposes.

Figure 1
Data Preprocessing



4.1.1. Molecular descriptors calculation

The following are the molecular descriptors calculation:

1) Molecular Weight (MW)

The total mass of all atoms in a compound is calculated as:

$$MW = \sum_{i=1}^{n} M_i \times N_i \tag{1}$$

where M_i is the atomic mass and N_i is the number of atoms of each element.

2) LogP

The logarithm of the partition coefficient between octanol and water, defined as:

$$Log\ P = log_{10}\left(\frac{C_{Octanol}}{C_{water}}\right) \tag{2}$$

3) NumHDonors and NumHAcceptors

- The number of hydrogen atoms capable of forming hydrogen bonds.
- The number of atoms (typically nitrogen, oxygen, or fluorine) that can form hydrogen bonds.
- All descriptors were calculated from SMILES strings using RDKit, a cheminformatics toolkit. These descriptors are essential for identifying bioactivity trends in the molecular dataset

4.1.2. Class balancing with synthetic minority oversampling technique

A model trained on unbalanced datasets will make predictions toward the majority class due to the unequal distribution of "active" and "inactive" compounds. The synthetic minority oversampling technique (SMOTE) served as a solution to handle this problem. Through SMOTE, the minority class receives synthetic samples that arise from

interpolating existing feature space data points, thus achieving dataset balance.

The model's dominance by the majority class is avoided through this method, which enhances its capability to handle imbalanced datasets. By balancing the instances of two classes in the dataset, SMOTE enables more accurate predictions of compound bioactivity between active and inactive categories.

4.2. Model development

QFM is a novel model that combines classical ML and quantum-inspired strategies to improve the prediction performance of bioactivity in CVD drug candidates.

1. Initially, the core factor of the proposed model is where the QRF maps the molecular descriptors to quantum feature vectors in a high-dimensional Hilbert space to learn nonlinear dependencies and entangled relationships where the classical model fails.
2. The proposed model has multiple layers can be observed in Figure 2, where the base layer and meta layer are. The first layer, where the classifiers are trained on quantum features in parallel with classical models trained on the same feature space. Bioactivity benefits from the ML approaches, which utilize distinct biases to discover different aspects of bioactivity data.
3. During the second stage, the logistic regression model functions as a meta-learner that unifies prediction probabilities received from both quantum and classical base learners. The stacking technique serves three purposes, including error correction, weighted consensus formation, and strong generalizability to unseen molecular data.

This meta learner fuses both models, which helps to align model complexity with classical stability.

4.2.1. Quantum data encoding

The major part of this research is the use of quantum data encoding [28] within the QFM. This approach uses angle embedding as the quantum encoding method because this approach is efficient in the encoding of classical data into quantum states. Quantitative descriptors such as MW, LogP, NumHDonors, and NumHAcceptors are represented as rotation angles of quantum circuits. This transformation enables the model to work in a high feature space, identifying complex and nonlinear features that classical algorithms find difficult to capture.

1) Angle embedding layer

In this layer, it allows taking numeric values from molecular descriptors and inputs them into the state of qubits. Imagine this is similar to turning the dial on a device with input numerals. This process brings standard data into the appropriate format that the quantum circuit will be able to work with.

The quantum state $(|\Psi(\theta)\rangle)$ after angle embedding is mathematically represented as follows:

$$|\Psi(\theta)\rangle = U(\theta)|0\rangle^{\otimes n} \tag{3}$$

where U (θ) denotes the unitary operation parameterized by θ, and $0^{\otimes n}$ represents the initial state of n qubits.

Mathematically, this process transforms the default state of the qubits by applying a specific operation known as a unitary transformation based on input values. All qubits begin in a default state, essentially blank, and this layer configures them with the molecular data.

2) Strongly entangled layer

This layer initiates the qubit states first and then connects the qubits. It uses a mix of two operations: CNOT gates that act as a connection between qubits so that they affect one another, and rotation gates (RX(π/5)) that further modify the state of each qubit.

These two steps enable the circuit to capture more sophisticated relationships between features. The embedding and entangling layers, in conjunction with each other, convert the input data into a new and more informative format better suited to learn patterns of bioactivity.

4.2.2. Quantum embedding circuit

As shown in Figure 3, the QFM uses a quantum embedding circuit in its design. The circuit comprises four quantum bits, q[0] through q[3], and four control classical bits, c[0] through c[3].

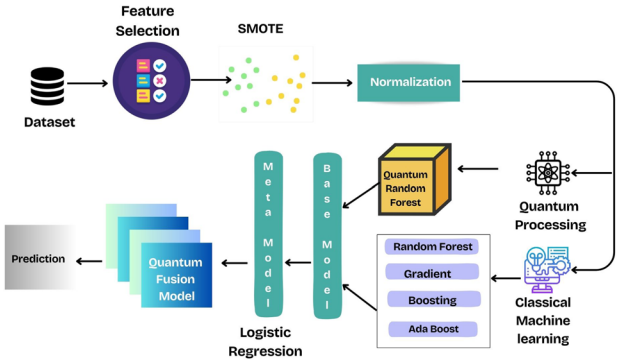
The key steps include the following:

1. Every qubit is then subjected to an RX(π/2), where molecular features of heart-targeted drugs are encoded into the qubit states.
2. A control gate connects q[0] with q[1], with an RZ(π/5) performed on q[1].
3. q[1] and q[2] are entangled, and then an RZ (π) is done on q[2].
4. q[2]and q[3]are now entangled and then an RZ(π/5) is performed on q[3].

This circuit creates dependencies between qubits, which correspond to different molecular descriptors. It demonstrates how the partition coefficient (LogP) may affect bioactivity if the MW is taken into consideration. The outcome of this quantum embedding circuit is to expand the number of features that QRF can look at by an exponential factor, improving its capability to classify bioactivities.

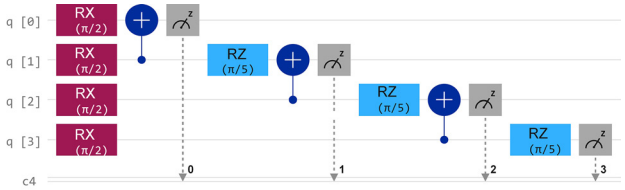
In this study, we used angle embedding for quantum data encoding because it provides good computational efficiency and NISQ-era hardware compatibility. Angle embedding converts traditional molecular attributes like MW and LogP together with hydrogen bond donors and acceptors into quantum gate rotation angles, mostly consisting of RX, RY, or RZ types that alter the quantum state. The implementation of Amplitude Embedding depends on vector normalization followed by

Figure 2
Quantum Fusion Model



Note. SMOTE = synthetic minority oversampling technique.

Figure 3
Quantum Embedding Circuit



data encoding into quantum states through amplitude modification [29]. The unique nature of amplitude encoding requires deep quantum circuits and exact state preparation, but it remains incompatible with both simulation work and hardware that needs precision. Basic encoding techniques use binary inputs, which makes them incapable of representing continuous molecular descriptors. This requires high computational expenses and optimization complications. Angle embedding provides linear depth in circuits and easier training and simulation capabilities, and direct links between quantum parameters and input features to boost interpretability.

4.3. Quantum random forest

The QRF classifier is the central component of the QFM model, as it implements quantum states to boost the bioactivity prediction. After the circuit transformation of every data point into one with the quantum embedding circuit, the intricate patterns are more recognizable. QRF works with parallel "quantum trees" in a way that resembles the decision tree approach of classical RFs. A different section of the data set reaches each model, which generates a prediction. By using quantum operations, the quantum trees contain data processing operations, including rotations and entanglements. The operations within the model enable it to uncover relationships between features that classical trees cannot usually discover. The quantitative predictions come from individual quantum trees. Majority voting serves as the basis to produce the final output, just as traditional forests do. The voting procedure strengthens both the accuracy and stability of the model.

4.3.1. Comparison with other quantum techniques

The development of the QFM involved studying multiple QML techniques, where QRF emerged as the selected quantum component. The following explanation describes the selection of QRF over other quantum models during the QFM development process.

1) Quantum Support Vector Machine (QSVM)

The QSVM detects complex data relationships through quantum kernel functions. The theoretical power of QSVM [30] requires substantial computational effort because it must generate extensive kernel matrices that expand proportionally to dataset expansion. Quantum models demonstrate poor performance and slowness during dataset processing on current quantum platforms.

2) Variational Quantum Classifier

Variational circuits incorporate trainable parameters as the basis for modeling nonlinear data in variational quantum classifier. The expressive nature of these models comes with two main challenges that must be addressed. The deep circuit architecture becomes incompatible with NISQ devices due to their unreliable operation. Training optimization turns extremely difficult as barren plateaus create barren regions, which frequently appear in these models.

The use of QRF enables superiority because it merges QC advantages with classical computing advantages. This helps to train different models to produce predictions that combine into a single output. This technique demonstrates both effective noise resistance and protection against overfitting. The system allows parallel prediction processing that suits molecular data sets with large dimensions and unbalanced characteristics. QRF functions effectively as part of QFM stacked architecture with traditional models like RF and gradient boosting (GB). The method uses angle-encoded molecular descriptors with logistic regression classifiers in each tree to generate results that run on current hardware systems.

The prediction through QRF uses aggregated outputs from multiple logistic regression classifiers that analyze quantum-encoded features. The system reaches its final decision by applying majority voting to predictions from individual classifiers, which evaluate each input.

4.3.2. QRF prediction mechanism

QRF performs predictions by combining the results from multiple logistic regression classifiers that work with features encoded in quantum mechanics. This makes use of majority voting to determine its final decision based on the class predictions provided by individual classifiers. Mathematically, the final decision is made through major voting by the following:

$$\hat{y}_i = arg\ max_y \left(\sum_{j=1}^{N_{estimators}} \boldsymbol{I}[h_j(x_i)] \right) \tag{4}$$

where $N_{estimators}$ are the number of estimators in the QRF, $h_j(x_i)$ is the prediction of the jth logistic regression model for input data xi, and $I(\cdot)$ is an indicator function.

4.4. Quantum fusion model

QFM is an ensemble framework that combines logistic regression learners that use quantum-level encoding with traditional machine learning classifiers, which apply RF and logistic regression as the meta-learner. This combination takes the feature expressiveness of the quantum circuit and the classical stability and uses them to enhance the prediction capability.

4.4.1. QFM training

At the beginning of the QFM pipeline, the molecular descriptor vectors enter a feature space through the quantum circuit. Simple rotations and entangling gates in this method generate the set of quantum measurements, which serve as quantum features. The quantum measurements produce data that contains high-order correlation information through mechanisms that avoid deep or noisy circuit operations. Once they are precomputed, quantum feature vectors are trained on the following base learners in parallel:

1) Random forest (RF)

This model constructs numerous decision trees using RF from randomly selected molecular descriptor subsets that contain both physicochemical properties and fingerprint counts. This employs bagging techniques that protect against noise while avoiding overfitting of related chemical characteristics.

2) Gradient boosting (GB)

The same descriptor space is used by GB to build multiple shallow decision trees in sequence, while each new tree aims to address remaining errors from previous trees. The model reveals hidden structure activity relationships that exist within the dataset through this method.

3) AdaBoost (AB)

The class imbalance between active and inactive bioactivity labels leads AB to adaptively adjust the weights of misclassified molecules during training. The ensemble model focuses exclusively on testing molecules that are situated close to the activity threshold during training.

4) Quantum random forest (QRF)

The quantum measurements of molecules are the input for QRF through the Pauli-Z expectation values from the embedding circuit, while the model trains multiple logistic regression nodes on different bootstrap samples. QRF uses the quantum-encoded features to identify the complex interactions.

These base learners, consisting of three classical and one quantum-enhanced approach, deliver distinct perspectives about the cardiovascular bioactivity data. The stacked meta-learner receives predictions from each of the four learners to generate the final activity classification.

Algorithm 1: Quantum Fusion Model (QFM)

Input: A Dataset (D) with a set of feature vectors X and target labels Y

Output: Model evaluation

Metrics: accuracy, classification report, confusion matrix

1: Initialize

2: Load dataset $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}\}$ from a CSV file

3. Split D into training set D_{train} and test set D_{test}

4: Process Data:

5: Standardize features using StandardScaler:

$$\tilde{x}_i = \frac{x_i - \mu_X}{\sigma_X} \quad (5)$$

6: Obtain scaled feature matrices $\tilde{X}_{\text{train}}$ and $\tilde{X}_{\text{test}}$

7: Quantum Embedding:

8: for each feature vector x ∈ X do

9: Encode features using quantum embedding circuit E(x):

$$\phi(x) = [\langle \text{PauliZ}_1 \rangle, \langle \text{PauliZ}_2 \rangle, \ldots, \langle \text{PauliZ}_n \rangle] \quad (6)$$

10: end for

11: Train Models:

12: **for** each quantum node Q_k in Quantum Random Forest **do**

13: Resample D_{train} to obtain $D_{\text{train},k}$

14: Encode resampled features $\tilde{X}_{\text{train},k}$:

$$\tilde{X}_{\text{train},k} = \{\phi(x_1), \phi(x_2), \ldots, \phi(x_{N_k})\} \quad (7)$$

15: Train logistic regression model L_k on resampled data $(\tilde{X}_{\text{train},k}, y_{\text{train},k})$

16: end for

17: *Stacking Classifier:*

18: Combine Quantum Random Forest and Random Forest as base models

19: Fit stacking classifier with logistic regression as meta-model:

$$\text{StackingClassifier} = \text{fit}(\tilde{X}_{\text{train}}, y_{\text{train}}) \quad (8)$$

20: Prediction:

21: for each test sample x_{test} do

22: Encode features using $E(x_{\text{test}})$:

$$\tilde{x}_{\text{test}} = \phi(x_{\text{test}}) \quad (9)$$

23. Predict using a stacking classifier:

$$\hat{y}_{\text{test}} = \text{StackingClassifier}(\tilde{x}_{\text{test}}) \quad (10)$$

24: end for

25: *Evaluation Metrics:*

26: Calculate accuracy

27: where $\delta(a, b) = 1$ if $a = b$, otherwise 0

28: Generate classification report

16: return

4.4.2. QFM stacking procedure

The stacking classifier functions through the following two-step process:

1. Training each base model requires separate execution using 80% of the dataset with optimized hyperparameters for achieving maximum performance. The models produce meta-features that serve as inputs for the subsequent stage of the process.
2. The training of a logistic regression classifier uses base model-derived meta-features to produce its combined output. The stacking classifier produces its final output through an integration of base model predictions that harmonizes QC elements with classical approaches.

The meta-classifier is defined as follows:

$$h_{meta}(x) = \sigma(W \cdot y + b) \quad (12)$$

The softmax function σ applies to the weight matrix W together with the bias vector b, which were learned during training.

4.4.3. Cross-validation and training

A five-fold cross-validation approach is applied to the model during the training methodology.

1. For each fold k = 1, . . . , 5:

The training of base models hi occurred on four folds while predictions were generated for the reserved fold. The out-of-fold predictions get aggregated to create new features X_{meta}. The training of h_{meta} occurs on the generated meta-features through a process.

2. The entire dataset undergoes model training with the optimized configuration after final model development.

This method employs two distinct levels to predict accurate performance, along with avoiding overfitting through prediction validation across different dataset sections.

4.5. Hyperparameter tuning and cross-validation

Hyperparameter tuning is an essential step for both classical and quantum models to enhance model performance. The grid search methodology was used to determine parameter values, which included mtry for RF and alpha for GB, and max_depth together with n_estimators for AB from predefined ranges. The proposed model adjusted the QRF model parameters to find the best point between performance effectiveness and quantum computational requirements.

The training phase employed five-fold cross-validation to validate the proposed models against overtraining while performing the training process. The evaluation process provided independent assessments of all components, ensuring high validity and transferability for the stacked QFM.

4.6. Computational resources and constraints

The training of high-power computing resources with GPU capabilities was used for the development and evaluation of the model, as well as to train the classical models for the QFM. Data encoding and circuit emulation were performed with classical hardware, using the PennyLane library to simulate an NISQ environment. All the quantum operations were performed in the emulated quantum platform, which is realistic considering existing limitations to quantum technology and available quantum hardware.

5. Model Evaluation and Results

This section analyzes the performance results of the proposed QFM with the classification report, F1 score, recall, precision, and ROC curves. The QFM, together with QRF, RF, GB, and AB models, showed high accuracy in bioactive compounds classification. Table 3 shows that

Table 3
Classification Report

Class	Precision	Recall	F1 score
Active	0.92	0.93	0.92
Inactive	0.94	0.93	0.93
Accuracy: 0.93			

the QFM attained an accuracy of 92.7%, which is higher than that of the individual models. High accuracy is further backed up by the balanced precision, recall, and F1 score of 0.91 in both the active and inactive classes in the model.

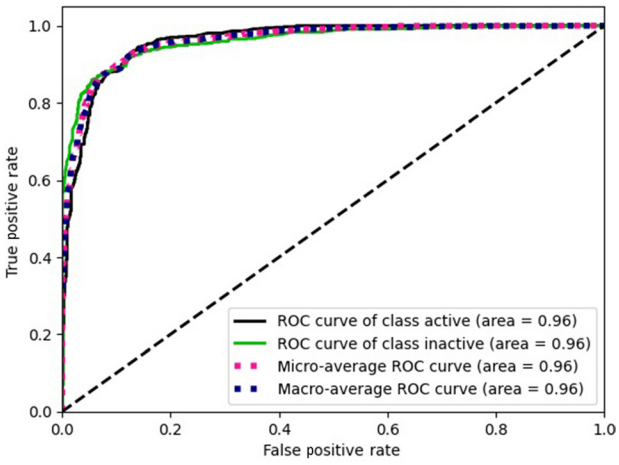
A detailed analysis shows a balanced class-wise performance. For the active class, the model was 0.92 for accuracy, 0.93 for recall, and 0.92 for the F1 score. At the same time, for the inactive class, it obtained the precision of 0.94, the recall of 0.93, and the F1 score of 0.93. This balance is significant in bioactivity classification because misclassification in either class can be costly.

Figures 4 and 5 represent the ROC and precision-recall curves of QFM, which achieves high performance in bioactivity prediction. QFM demonstrates reliable active-compound identification by showing a steep rising ROC curve and a large AUC value, along with a high PR AUC value for maintaining precision in rare active detection. It combines multiple perspective views through majority voting, which detects weak structure–activity relationship patterns that classical models cannot detect, while using partial vote dilution to minimize bias and variance in the process. This makes the QFM optimal for imbalanced noisy_filted datasets found in cardiovascular drug discovery.

5.1. Comparative analysis

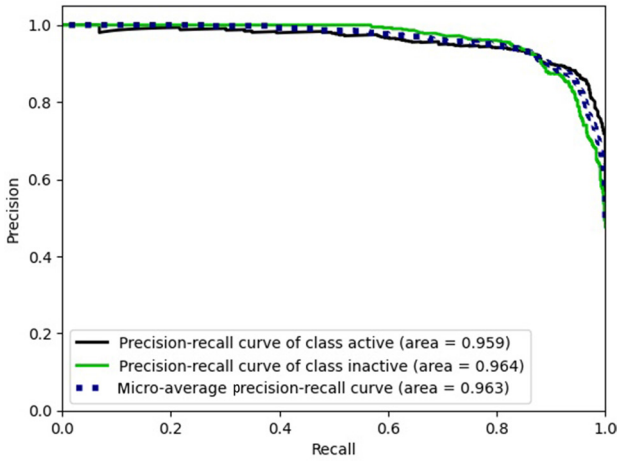
This section provides a comparative analysis of the proposed QFM with the individual base models and classical ML methods. The evaluation metrics include accuracy, precision, recall, F1 score, confusion matrix, and ROC-AUC. This section highlights the values of the quantum-classical integration in improving the predictive capability and model stability of a given problem.

Figure 4
QFM ROC Curve



Note. QFM = quantum fusion model and ROC = receiver operating characteristic.

Figure 5
QFM Precision-Recall Curve



5.1.1. Confusion matrix analysis

The confusion matrix shows detailed qualitative model classification performance results by recording TP, TN, FP, and FN counts for active and inactive class classifications. The classification errors demonstrated by the QFM appear less frequent in Figure 6 when compared to classical model performance levels.

The QFM demonstrates strong generalization capabilities for unseen data because it balances true positive and true negative predictions effectively for bioactivity prediction. The diminished number of false positives and false negatives demonstrates improved reliability of active compound identification through error reduction in classification.

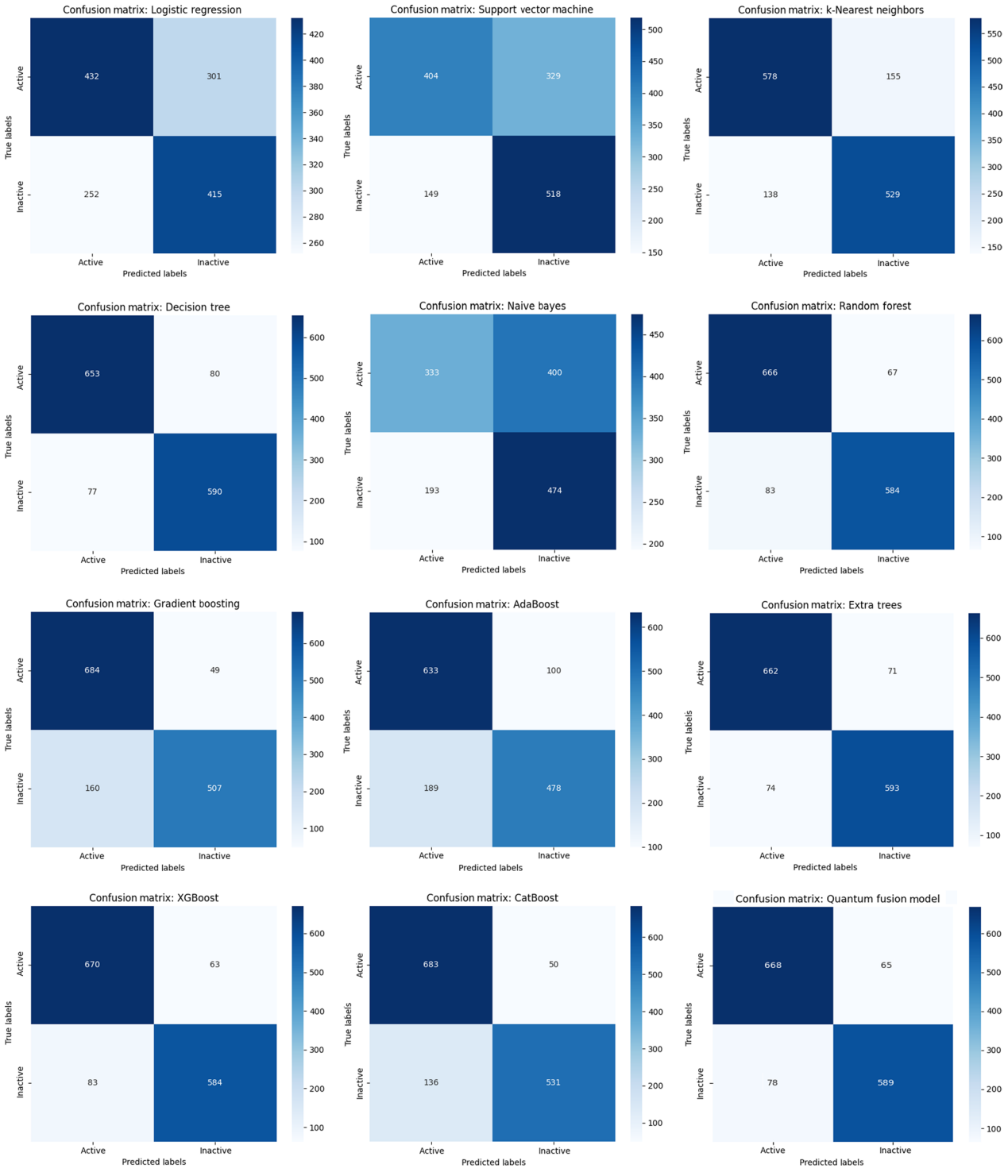
5.1.2. ROC-AUC analysis

The AUC acts as a performance indicator for the ROC curve, which controls model classification abilities. Figure 7 shows the ROC curves for the QFM and classical models. QFM model with an AUC score of 0.961 as its curve extends toward the top-left corner of the plot for excellent discrimination ability. Quantum data encoding in the QFM enables strong discriminative power because it successfully detects complex molecular patterns, resulting in a high AUC score. The AUC scores from classical models fall between 0.82 and 0.87, indicating moderate performance levels. This represents that the proposed model performs well compared with the classical models.

5.1.3. Precision-recall curve analysis

The precision-recall curve acts bioactivity prediction evaluation because it demonstrates how well models predict “active” class instances among minority datasets. The QFM demonstrates superior precision performance for all models. The QFM demonstrates the best precision-recall area performance in Figure 8 among all evaluated models, which proves its effectiveness in working with imbalanced datasets and keeping precision at elevated recall points.

Figure 6
Confusion Matrix for Comparison with Traditional Models



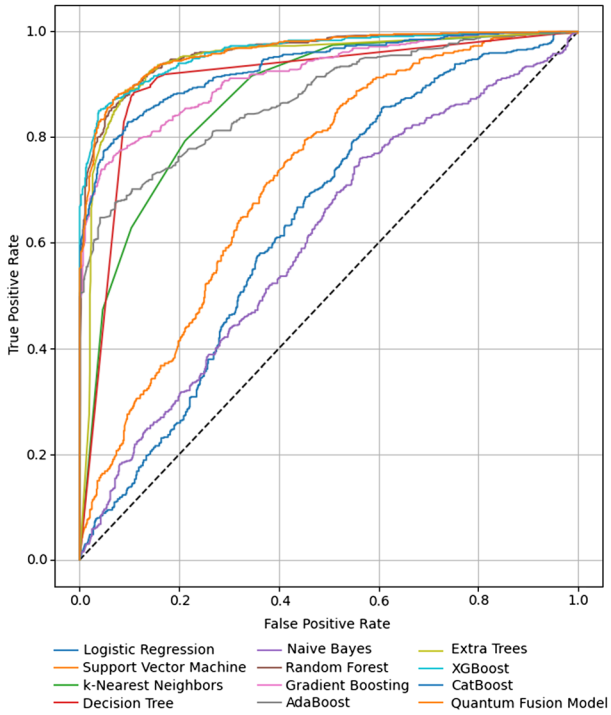
5.1.4. Performance metrics summary

The results in Figure 9 show that the QFM provided superior performance than classical approaches across all metrics, including accuracy, precision, recall, and F1-score. In Table 4, QFM demonstrated 92.7% accuracy and precision levels of 0.94 and recall levels of 0.92, resulting in an F1-score of 0.93.

Tables 5, 6, and 7 provide the key improvements about the QFM approach. First, the comparison with Classical models, Tree-based models, and Boosting models is addressed below.

The QFM shows better accuracy, recall, and F1 score outcomes when compared to conventional ML methods, as well as classical learners and tree-based models, and advanced ensemble techniques. QFM demonstrates remarkable performance gains because it achieves

Figure 7
Comparison of ROC Curves



Note. ROC = receiver operating characteristic.

Figure 8
Comparison of the Precision-Recall Curve

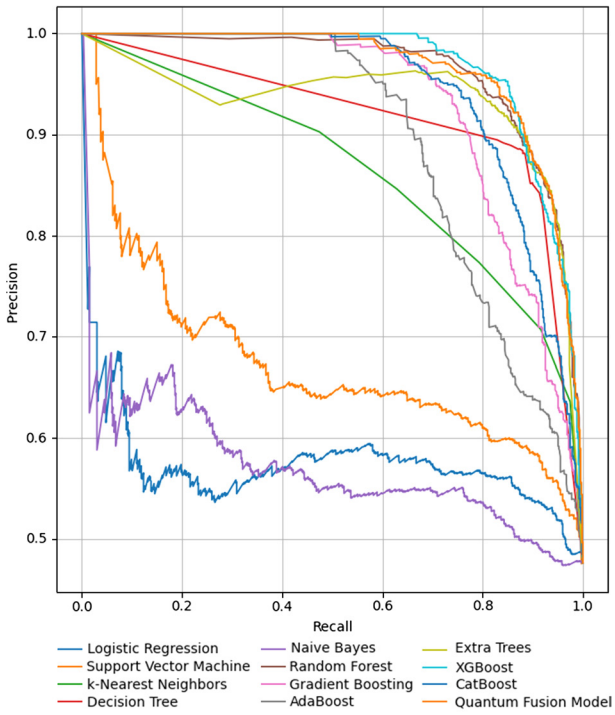


Figure 9
Model Performance

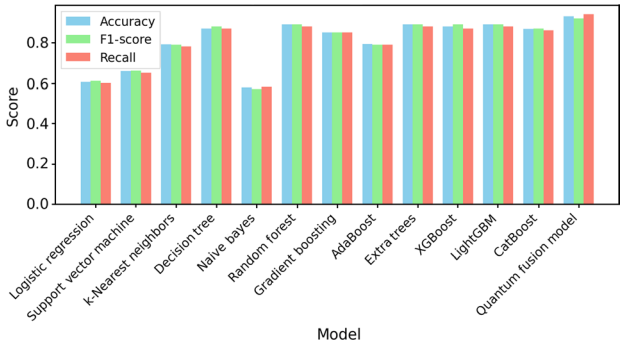


Table 4
Comparative Analysis

Model	Accuracy	F1 score	Recall
Logistic Regression	0.60	0.61	0.60
Support Vector Machine	0.65	0.66	0.65
k-Nearest Neighbors	0.79	0.79	0.78
Decision Tree	0.85	0.87	0.86
Naive Bayes	0.57	0.57	0.58
Random Forest	0.88	0.88	0.87
Gradient Boosting	0.85	0.85	0.85
AdaBoost	0.79	0.79	0.79
Extra Trees	0.86	0.88	0.86
XGBoost	0.87	0.87	0.85
CatBoost	0.86	0.87	0.86
Quantum Fusion Model	0.93	0.92	0.94

Table 5
Comparison with Classical Models

Model	Accuracy	F1 score	Recall
Logistic regression	+0.33	+0.31	+0.34
Support vector machine	+0.28	+0.26	+0.29
k-nearest neighbors	+0.14	+0.13	+0.16

Table 6
Comparison with Tree-Based Models

Model	Accuracy	F1 score	Recall
Decision tree	+0.08	+0.05	+0.08
Random forest	+0.05	+0.04	+0.07
Extra trees	+0.07	+0.04	+0.08

Table 7
Comparison with Boosting Models

Model	Accuracy	F1 score	Recall
AdaBoost	+0.14	+0.13	+0.15
XGBoost	+0.06	+0.05	+0.09
CatBoost	+0.07	+0.05	+0.08

33% more accuracy than logistic regression and 29% better recall than SVM. The fixed feature spaces, together with linear separability and distance-based models within continuous neighborhoods, fail to function properly when molecular features interact through hierarchical methods. Tree-based models perform better due to their ability to model nonlinearity, but they depend on discrete decision boundaries and cannot represent subtle interdependencies among molecular descriptors. QFM focuses on these limitations by introducing quantum angle embedding and entangling layers that send molecular descriptors into a high-dimensional Hilbert space, which allows for capturing intricate correlations between features such as LogP, MW, and hydrogen bond interactions. The hybrid architecture combines a quantum-enhanced model with diverse classical models through a stacked ensemble. This stacked fusion enables to representation of discrete decisions and continuous entangled patterns, resulting in higher accuracy and better generalization to unseen compounds. Particularly, the improved recall across all comparisons indicates QFM's strength in minimizing false negatives in bioactivity prediction.

The high precision and recall values depict the efficiency of the QFM in the bioactivity prediction by assessing the number of true positive activities while at the same time reducing false positive activities. The high F1 score also underlines the fact that for the majority of real-world applications where both precision and recall are equally crucial, the proposed model yields satisfactory results.

6. Conclusion and Future Work

This study provides a novel QFM framework that integrates quantum mechanics with the classical models, which classifies the active and inactive compounds in CVD bioactivity targets. It combines a shallow-circuit QRF alongside RF, GB, along with AB through stacking, which enables the model to detect both quantum correlations of high order and classical decision boundaries. The proposed model demonstrates generalization capacity with high precision and recall values under imbalanced class conditions. The capabilities of QFM produce significant benefits for drug discovery through two channels: it enables the robust selection of better candidate molecules and slashes false positive rates in experiments, and guides more advanced enhancement methods focused on unobserved feature interactions.

This work has some constraints despite recent accomplishments. This examines cardiovascular targets from ChEMBL with standard descriptors, but it does not cover the extensive chemical diversity found in wider therapeutic fields. The performance of quantum models requires optimization of hyperparameters and circuit designs because their impact on quantum hardware will improve with its ongoing development.

The future work may include the following:

1. Testing on new bioactivity repositories, which include various targets from oncology, infectious disease, and metabolic categories, to determine their predictive power across diverse chemical environments.
2. A study can be conducted on deep or variational quantum circuits with error mitigation methods while also testing stacked components, including SVMs and neural networks, to improve performance.

These directions may drive the development of more efficient and cost-effective drug discovery strategies for CVD and other therapeutic areas.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available at https://www.ebi.ac.uk.

Author Contribution Statement

Gundala Pallavi: Methodology, Software, Data curation, Writing – original draft, Visualization, Project administration. **Ali Altalbe:** Investigation, Writing – review & editing. **Prasanna Kumar Rangarajan:** Conceptualization, Validation, Formal analysis, Resources, Writing – review and editing, Supervision.

References

[1] Zaidan, A. M. (2023). The leading global health challenges in the artificial intelligence era. *Frontiers in Public Health*, *11*, 1328918. https://doi.org/10.3389/fpubh.2023.1328918

[2] Schapin, N., Majewski, M., Varela-Rial, A., Arroniz, C., & de Fabritiis, G. (2023). Machine learning small molecule properties in drug discovery. *Artificial Intelligence Chemistry*, *1*(2), 100020. https://doi.org/10.1016/j.aichem.2023.100020

[3] Babu, S. V., Ramya, P., & Gracewell, J. (2024). Revolutionizing heart disease prediction with quantum-enhanced machine learning. *Scientific Reports*, *14*(1), 7453. https://doi.org/10.1038/s41598-024-55991-w

[4] Liang, Z., He, Z., Sun, Y., Herman, D., Jiao, Q., Zhu, Y., ..., & Shi, Y. (2024). Synergizing quantum techniques with machine learning for advancing drug discovery challenge. *Scientific Reports*, *14*(1), 31216. https://doi.org/10.1038/s41598-024-82576-4

[5] Ashraf, F. B., Akter, S., Mumu, S. H., Islam, M. U., & Uddin, J. (2023). Bio-activity prediction of drug candidate compounds targeting SARS-Cov-2 using machine learning approaches. *PLOS ONE*, *18*(9), e0288053. https://doi.org/10.1371/journal.pone.0288053

[6] Sadybekov, A. V., & Katritch, V. (2023). Computational approaches streamlining drug discovery. *Nature*, *616*, 673–685. https://doi.org/10.1038/s41586-023-05905-z

[7] Obaido, G., Mienye, I. D., Egbelowo, O. F., Emmanuel, I. D., Ogunleye, A., Ogbuokiri, B., ..., & Aruleba, K. (2024). Supervised machine learning in drug discovery and development: Algorithms, applications, challenges, and prospects. *Machine Learning with Applications*, *17*, 100576. https://doi.org/10.1016/j.mlwa.2024.100576

[8] Pallavi, G., & Prasanna Kumar, R. (2025). Quantum natural language processing and its applications in bioinformatics: A comprehensive review of methodologies, concepts, and future directions. *Frontiers in Computational Science*, *7*, 1464122. https://doi.org/10.3389/fcomp.2025.1464122

[9] Chenthamarakshan, V., Hoffman, S. C., Owen, C. D., Lukacik, P., Strain-Damerell, C., Fearon, D., ..., & Das, P. (2023). Accelerating drug target inhibitor discovery with a deep generative foundation model. *Science Advances*, *9*(25), eadg7865. https://doi.org/10.1126/sciadv.adg7865

[10] Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., ..., & Zhao, S. (2019). Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*, *18*, 463–477. https://doi.org/10.1038/s41573-019-0024-5

[11] Shi, Y. (2021). Support vector regression-based QSAR models for prediction of antioxidant activity of phenolic compounds. *Scientific Reports*, *11*(1), 8806. https://doi.org/10.1038/s41598-021-88341-1

[12] Rodríguez-Pérez, R., & Bajorath, J. (2022). Evolution of support vector machine and regression modeling in chemoinformatics and drug discovery. *Journal of Computer-Aided Molecular Design*, *36*(5), 355–362. https://doi.org/10.1007/s10822-022-00442-9

[13] Mao, J., Akhtar, J., Zhang, X., Sun, L., Guan, S., Li, X., ..., & Wang, G. (2021). Comprehensive strategies of machine-learning-based quantitative structure-activity relationship models. *iScience*, *24*(9), 103052. https://doi.org/10.1016/j.isci.2021.103052

[14] Siddiqui, N. F., Vishwakarma, P., Thakur, S., & Jadhav, H. R. (2025). Bioactivity predictions and virtual screening using machine learning predictive model. *Journal of Biomolecular Structure and Dynamics*, *43*(8), 3909–3928. https://doi.org/10.1080/07391102.2023.2300132

[15] Raschka, S., & Kaufman, B. (2020). Machine learning and AI-based approaches for bioactive ligand discovery and GPCR-ligand recognition. *Methods*, 180, 89–110. https://doi.org/10.1016/j.ymeth.2020.06.016

[16] Yadalam, P. K., Natarajan, P. M., Saeed, M. H., & Ardila, C. M. (2025). Variational approaches for drug-disease-gene links in periodontal inflammation. *International Dental Journal*, *75*(1), 185–194. https://doi.org/10.1016/j.identj.2024.09.025

[17] Pyrkov, A., Aliper, A., Bezrukov, D., Lin, Y. C., Polykovskiy, D., Kamya, P., ..., & Zhavoronkov, A. (2023). Quantum computing for near-term applications in generative chemistry and drug discovery. *Drug Discovery Today*, *28*(8), 103675. https://doi.org/10.1016/j.drudis.2023.103675

[18] Bharti, K., Cervera-Lierta, A., Kyaw, T. H., Haug, T., Alperin-Lea, S., Anand, A., ..., & Aspuru-Guzik, A. (2022). Noisy intermediate-scale quantum algorithms. *Reviews of Modern Physics*, *94*(1), 015004. https://doi.org/10.1103/RevModPhys.94.015004

[19] Wang, P. H., Chen, J. H., Yang, Y. Y., Lee, C., & Tseng, Y. J. (2023). Recent advances in quantum computing for drug discovery and development. *IEEE Nanotechnology Magazine*, *17*(2), 26-30. https://doi.org/10.1109/MNANO.2023.3249499

[20] Ding, Q. M., Huang, Y. M., & Yuan, X. (2024). Molecular docking via quantum approximate optimization algorithm. *Physical Review Applied*, *21*(3), 034036. https://doi.org/10.1103/PhysRevApplied.21.034036

[21] Xie, N., Lee, X., Cai, D., Saito, Y., & Asai, N. (2023). Quantum approximate optimization algorithm parameter prediction using a convolutional neural network. *Journal of Physics: Conference Series*, *2595*(1), 012001. https://doi.org/10.1088/1742-6596/2595/1/012001

[22] Bhatia, A. S., Saggi, M. K., & Kais, S. (2023). Quantum machine learning predicting ADME-Tox properties in drug discovery. *Journal of Chemical Information and Modeling*, *63*(21), 6476–6486. https://doi.org/10.1021/acs.jcim.3c01079

[23] Mensa, S., Sahin, E., Tacchino, F., Kl Barkoutsos, P., & Tavernelli, I. (2023). Quantum machine learning framework for virtual screening in drug discovery: A prospective quantum advantage. *Machine Learning: Science and Technology*, *4*(1), 015023. https://doi.org/10.1088/2632-2153/acb900

[24] Sagingalieva, A., Kordzanganeh, M., Kenbayev, N., Kosichkina, D., Tomashuk, T., & Melnikov, A. (2023). Hybrid quantum neural network for drug response prediction. *Cancers*, 15(10), 2705. https://doi.org/10.3390/cancers15102705

[25] Batra, K., Zorn, K. M., Foil, D. H., Minerali, E., Gawriljuk, V. O., Lane, T. R., & Ekins, S. (2021). Quantum machine learning algorithms for drug discovery applications. *Journal of Chemical Information and Modeling*, *61*(6), 2641-2647. https://doi.org/10.1021/acs.jcim.1c00166

[26] Lau, B., Emani, P. S., Chapman, J., Yao, L., Lam, T., Merrill, P., ..., & Lam, H. Y. K. (2023). Insights from incorporating quantum computing into drug design workflows. *Bioinformatics*, *39*(1), btac789. https://doi.org/10.1093/bioinformatics/btac789

[27] Rath, M., & Date, H. (2024). Quantum data encoding: A comparative analysis of classical-to-quantum mapping techniques and their impact on machine learning accuracy. *EPJ Quantum Technology*, *11*, 72. https://doi.org/10.1140/epjqt/s40507-024-00285-3

[28] Riaz, F., Abdulla, S., Suzuki, H., Ganguly, S., Deo, R. C., & Hopkins, S. (2025). The application of quantum pre-processing filter for binary image classification with small samples. *Journal of Data Science and Intelligent Systems*, *3*(2), 109–116. https://doi.org/10.47852/bonviewJDSIS42024229

[29] Rani, S., Kaur, R., Desai, C., & Ambilwade, R. P. (2024). Quantum machine learning: Leveraging quantum computing for enhanced learning algorithms. *International Journal of Future Management Research*, *6*(5), 1–15. https://doi.org/10.36948/ijfmr.2024.v06i05.27450

[30] Ding, C., Bao, T. Y., & Huang, H. L. (2022). Quantum-inspired support vector machine. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(12), 7210–7222. https://doi.org/10.1109/TNNLS.2021.3084467

How to Cite: Pallavi G., Altalbe A., & Rangarajan, P. K. (2025). QFM-BioPred: Quantum Fusion Model for Bioactivity Prediction in Cardiovascular Disease Drug Discovery. *Journal of Computational and Cognitive Engineering*. https://doi.org/10.47852/bonviewJCCE52025138