JCTA Journal of Computing Theories and Applications ISSN:3024-9104

Research Article

Integrating Quantum, Deep, and Classic Features with Attention-Guided AdaBoost for Medical Risk Prediction

Muh Galuh Surya Putra Kusuma [1], De Rosal Ignatius Moses Setiadi [2,*], Wise Herowati [2], T. Sutojo [2], Prajanto Wahyu Adi [3], Pushan Kumar Dutta [4], and Minh T. Nguyen [5]

[1] Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia;
e-mail : muhgaluhspk@gmail.com
[2] Research Group for Quantum Computing and Materials Informatics, Universitas Dian Nuswantoro, Semarang 50131, Indonesia;
e-mail : moses@dsn.dinus.ac.id, wise.herowati@dsn.dinus.ac.id, sutojo@dsn.dinus.ac.id
[3] Department of Informatics, Universitas Diponegoro, Semarang 50275, Indonesia;
e-mail : prajanto@live.undip.ac.id
[4] School of Engineering and Technology, Amity University Kolkata, Kolkata 700135, West Bengal, India;
e-mail : pkdutta@kol.amity.edu
[5] Thai Nguyen University of Technology, Thai Nguyen University, Thai Nguyen 240000, Viet Nam;
e-mail : nguyentuanminh@tnut.edu.vn
* Corresponding Author : De Rosal Ignatius Moses Setiadi

Abstract: Chronic diseases such as chronic kidney disease (CKD), diabetes, and heart disease remain major causes of mortality worldwide, highlighting the need for accurate and interpretable diagnostic models. However, conventional machine learning methods often face challenges of limited generalization, feature redundancy, and class imbalance in medical datasets. This study proposes an integrated classification framework that unifies three complementary feature paradigms: classical tabular attributes, deep latent features extracted through an unsupervised Long Short-Term Memory (LSTM) encoder, and quantum-inspired features derived from a five-qubit circuit implemented in PennyLane. These heterogeneous features are fused using a feature-wise attention mechanism combined with an AdaBoost classifier to dynamically weight feature contributions and enhance decision boundaries. Experiments were conducted on three benchmark medical datasets—CKD, early-stage diabetes, and heart disease—under both balanced and imbalanced configurations using stratified five-fold cross-validation. All preprocessing and feature extraction steps were carefully isolated within each fold to ensure fair evaluation. The proposed hybrid model consistently outperformed conventional and ensemble baselines, achieving peak accuracies of 99.75% (CKD), 96.73% (diabetes), and 91.40% (heart disease) with corresponding ROC AUCs up to 1.00. Ablation analyses confirmed that attention-based fusion substantially improved both accuracy and recall, particularly under imbalanced conditions, while SMOTE contributed minimally once feature-level optimization was applied. Overall, the attention-guided AdaBoost framework provides a robust and interpretable approach for clinical risk prediction, demonstrating that integrating diverse quantum, deep, and classical representations can significantly enhance feature discriminability and model reliability in structured medical data.

Keywords: AdaBoost; Attention Mechanism; Deep Learning; Feature Fusion; Medical Diagnosis; Quantum Machine Learning; Structured Data; Unsupervised LSTM.

Received: September, 12th 2025
Revised: October, 7th 2025
Accepted: October, 10th 2025
Published: October, 11th 2025

1. Introduction

In recent years, machine learning (ML) has played a crucial role in advancing healthcare, particularly in clinical data analysis and medical decision support systems [1]–[3]. Its ability to process large-scale and complex data makes it crucial in supporting the diagnosis of chronic diseases such as chronic kidney disease (CKD), diabetes, and heart disease [4]–[8]. Early detection and accurate diagnosis contribute significantly to the effectiveness of therapy and improve patients' quality of life [9]. However, clinical data often faces challenges such as class imbalance, temporal complexity, and limited feature representation. These conditions often

cause conventional ML models to fail to recognize important patterns or generalize optimally. On the other hand, the advancement of precision medicine demands predictive systems that are not only accurate but also adaptive to diverse patient characteristics [10].

Classical feature-based approaches, such as logistic regression and support vector machines (SVM), remain relevant due to their efficiency and interpretability [11], but are limited in capturing non-linear relationships. Meanwhile, deep learning (DL) is capable of extracting temporal relationships between medical features and producing more expressive non-linear representations [12]. Among various architectures, Long Short Term Memory (LSTM) networks are particularly advantageous for capturing long-range dependencies and mitigating the vanishing gradient problem in sequential data [13]–[15], and have been widely applied in various researches not only in healthcare dataset [16]–[18]. However, these models tend to require large data sets and are susceptible to overfitting.

Quantum machine learning (QML) presents new opportunities by transforming data into a high-dimensional feature space, leveraging the principles of superposition and entanglement, which enable the formation of richer representations [19]–[21]. This approach has shown potential in improving classification performance on small-scale data [22], [23], although its implementation remains largely simulated due to hardware limitations. To achieve a more robust representation, integrating all three feature sources, i.e., classical, deep, and quantum is a promising strategy [20], [21]. However, direct concatenation without a weight selection mechanism often results in redundancy and imbalanced contributions among features. Therefore, this study proposes a fusion attention mechanism that can dynamically adjust feature weights based on their informative contributions.

The feature-wise self-attention mechanism employed consists of two sequential dense layers: the first layer utilizes the tanh activation function to capture non-linearity, and the second layer employs softmax to generate normalized attention weights. These weights are then multiplied element-by-element against the combined feature vector, so that features more relevant to the prediction receive proportional amplification. This approach is suitable for heterogeneous tabular data and does not require a query–key–value structure, unlike sequential models [24]–[26]. Thus, the system can adapt to different feature distributions while remaining computationally efficient.

The classification process is performed using Adaptive Boosting (AdaBoost), which is known to be effective in iteratively amplifying weak learners and remains stable over variations in features [27]. AdaBoost was chosen as the primary classification model for several key reasons. First, AdaBoost can iteratively amplify weak learners, enabling it to combine features with different distributions and scales without losing stability. Second, its adaptive nature automatically adjusts the error weights at each iteration, allowing greater focus on minority samples or difficult cases [28]. Third, compared to other ensemble methods such as bagging, AdaBoost is more efficient in handling feature heterogeneity resulting from the combination of classical, LSTM, and quantum features. Thus, the combination of AdaBoost and fusion attention can produce a robust and stable medical classification model against data variations. This research aims to develop a more adaptive and efficient framework for chronic disease classification by integrating heterogeneous features with attention mechanisms. In summary, this study makes the following contributions:

- Integrating three major feature types—classical tabular features, deep representations derived from an unsupervised LSTM encoder, and quantum-inspired features—into a unified classification pipeline.
- Introducing a feature-wise attention mechanism combined with the AdaBoost classifier to assign dynamic importance weights to heterogeneous feature representations, thereby enhancing performance prediction in imbalanced medical datasets.
- Conducting experimental evaluations and limited ablation studies to analyze the contribution of attention-based fusion toward model performance and stability across three benchmark datasets: CKD, Diabetes, and heart disease.

The remainder of this paper is organized as follows. Section 2 presents the theoretical background and related works on classical, deep, and quantum approaches in medical classification. Section 3 describes the proposed hybrid architecture and methodological workflow. Section 4 details the experimental setup, results, and ablation studies. Section 5 compares the proposed model with existing methods, and Section 6 concludes the paper with key findings and future research directions.

2. Preliminaries

2.1. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a derivative architecture of the Recurrent Neural Network (RNN), designed to address the vanishing gradient problem in sequential data [15]. The model utilizes three primary gates: input, forget, and output gates. These gates regulate the flow of information, determining what is retained or discarded over time. Each LSTM unit is formulated as follows:

$$\begin{aligned} f_t &= \sigma\big(W_f[h_{t-1}, x_t] + b_f\big), & i_t &= \sigma(W_i[h_{t-1}, x_t] + b_i), \\ \tilde{C}_t &= \tanh(W_C[h_{t-1}, x_t] + b_C), & C_t &= f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \\ o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o), & h_t &= o_t \odot \tanh(C_t), \end{aligned} \tag{1}$$

Where f_t, i_t, o_t denote the forget, input, and output gates, respectively; C_t represents the memory cell state, and h_t is the hidden output.

In this study, LSTM is not employed as a sequential classification model but rather as an unsupervised feature extractor. The LSTM layer functions analogously to an encoder in an autoencoder, learning latent representations h_t that capture the correlations among medical features without explicit target labels. The final hidden state is treated as a compressed non-linear feature, which is subsequently combined with classical and quantum-inspired features in the fusion attention stage (Section 2.3). This approach is effective for various tabular data types [29], [30].

2.2. Quantum Feature Encoding

The QML approach relies on transforming classical data into a high-dimensional Hilbert space through quantum circuits that naturally model non-linear interactions [31]–[33]. In this study, the encoding process is carried out using a combination of Angle Embedding and Basic Entangler Layers as implemented in PennyLane.

2.2.1. Angle Embedding

Given a classical feature vector $x = (x_1, x_2, \dots, x_n)$ each element is injected into a qubit through a single rotation about one of the Bloch sphere axes, namely X, Y, or Z, using a rotation operator:

$$U_{\text{embed}}(x_i) = R_{\text{axis}}(x_i), \text{axis} \in \{X, Y, Z\} \tag{2}$$

In PennyLane, this is implemented using the function qml.AngleEmbedding(x, wires, rotation='Y'), which by default performs rotations around the Y-axis, but can be configured as needed. This embedding converts each feature value into a quantum phase, making the classical data part of the qubit state, which can subsequently be manipulated through entanglement operations.

2.2.2. Basic Entangler Layers

A CNOT network arranged in a ring topology, connecting each qubit to its neighbor, including the last qubit back to the first, forming a cyclic entanglement pattern.

Formally, for n qubits and L layers, the quantum ansatz operation can be expressed as:

- Parametric rotation $R_{\text{axis}}(\theta_i, \ell)$ applied to each qubit i, where θ_i, ℓ represents the optimizable parameter in the l^{th} layer.
- A CNOT network arranged in a ring topology, connecting each qubit to its neighbor, including the last qubit back to the first, forming a cyclic entanglement pattern.

Formally, for n qubits and L layers, the quantum ansatz operation can be expressed as:

$$U_{\text{ansatz}}(\Theta) = \prod_{\ell=1}^{L} \left(\prod_{i=1}^{n} R_{\text{axis}}(\theta_{\ell,\text{i}}) \cdot \prod_{\langle i,j \rangle} \text{CNOT}_{i,j} \right) \tag{3}$$

Where $\Theta = \{\theta_{\ell,\text{i}}\}$ denotes the set of trainable parameters, and $\langle i, j \rangle$ represents pairs of qubits connected in the ring topology. This structure maintains homogeneous entanglement across all qubits and is computationally efficient [16].

2.2.3. Quantum Measurement and Feature Extraction

After all layers are applied, the final quantum state $|\psi\rangle$ is measured using the Pauli-Z operator:

$$z_j = \langle\psi|Z_j|\psi\rangle, \quad j = 1,2,\dots,n \tag{4}$$

The vector $z = (z_1, z_2, \dots, z_n)$ m becomes the quantum feature vector, reflecting non-linear interactions among features. These values are then normalized and combined with classical and LSTM-derived features in the fusion attention process (Section 2.3).

2.3. Feature-wise Fusion Attention

To balance the contribution of the three heterogeneous feature sources, a feature-wise self-attention mechanism is employed. This mechanism operates directly on the feature dimension rather than on temporal sequences and is implemented using two sequential dense layers:

$$\begin{aligned} e &= \tanh(W_1 x + b_1), \\ \alpha &= softmax(W_2 e + b_2), \\ z &= \alpha \odot x, \end{aligned} \tag{5}$$

Where W_1, W_2 are the trainable weight matrices. The vector z in Equation (5) represents the adaptively reweighted feature vector, where more informative dimensions receive higher attention weights α, as defined by the same equation.

2.4. AdaBoost Classifier

AdaBoost combines multiple weak learners through an adaptive weighting process [34], [35]. The classification error at each iteration is computed as:

$$\epsilon_t = \sum_{i=1}^{N} D_t(i)[y_i \neq h_t(x_i)] \tag{6}$$

with the model's confidence weight defined as:

$$\alpha_t = \frac{1}{2}\ln\frac{1-\epsilon_t}{\epsilon_t}, \tag{7}$$

and the sample distribution updated according to:

$$D_{t+1}(i) = \frac{D_t(i)\exp\left(-\alpha_t y_i h_t(x_i)\right)}{Z_t}. \tag{8}$$

The final prediction is obtained through weighted voting, expressed as:

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right). \tag{9}$$

The combination of Equations (6)–(9) makes AdaBoost inherently adaptive to prior model errors and robust when dealing with heterogeneous feature representations.

2.5. Related Work

Research developments in disease diagnosis using ML have shown significant progress in recent years. Various approaches have been developed, ranging from DL models based on LSTM, QML frameworks implemented with PennyLane, to ensemble methods such as AdaBoost combined with data-balancing techniques. Each paradigm contributes to improving the accuracy, stability, and interpretability of medical decision-support systems.

The LSTM model has demonstrated high effectiveness in modeling both temporal and multivariate clinical data. Rajkomar et al. [36] developed a deep learning–based prediction system that processes raw Electronic Health Records (EHR) from two large academic hospitals in the United States, covering more than 216,000 inpatients and 46 billion clinical data

points. Without manual feature pre-selection, the LSTM model achieved an area under the ROC curve (AUROC) of 0.95 for inpatient mortality prediction and 0.77 for 30-day readmission prediction, outperforming conventional predictive models such as aEWS, which only reached an AUROC of around 0.85.

These findings confirm that LSTM can capture complex temporal dynamics in medical data and improve diagnostic accuracy without extensive manual feature engineering. Moreover, the unsupervised LSTM encoder approach has also been applied to extract latent representations from non-sequential medical tabular data, enabling the learning of nonlinear patterns among clinical variables without explicit target labels [37].

QML has emerged as a promising paradigm that leverages high-dimensional feature spaces for medical classification under limited data conditions. Havlíček et al. [38] demonstrated two essential approaches: variational quantum classifiers and quantum kernel estimation, both implemented on superconducting quantum processors. These methods map classical data into quantum feature spaces using parameterized circuits and amplitude-based quantum kernels. With a small synthetic dataset (20 samples per class), the variational classification approach achieved near-perfect accuracy (>99%) even under decoherence conditions. Although the experiments were not conducted on medical data, the results open opportunities for applying QML to diagnose chronic diseases such as diabetes, CKD, and heart disease, particularly in scenarios with small but complex datasets. Setiadi et al. [21] and Chow [19] also emphasizes the potential of QML in extending non-linear representations that can improve model performance.

While DL and QML have gained prominence in recent research, classical tabular feature-based methods remain vital in medical diagnosis. Perumal and Kaladevi [11] evaluated coronary heart disease classification models using the Cleveland dataset (303 samples, 13 features), comparing Logistic Regression (LR), Support Vector Machine (SVM), and k-Nearest Neighbour (kNN). The LR model achieved the highest accuracy (87%) with a specificity of 88% and sensitivity of 85%, outperforming SVM (85%) and kNN (69%). These results confirm that classical feature–based approaches remain competitive when optimized with appropriate feature selection and normalization techniques. Furthermore, a study by Azis [39] showed that logistic regression remains relevant in the early screening stage of heart disease diagnosis, achieving an accuracy and recall of 84.39% using combined data from four international datasets.

The Synthetic Minority Over-sampling Technique (SMOTE) has proven effective in balancing data distributions by generating synthetic samples. Khokhar et al. [40] applied SMOTE to the Diabetes Binary Health Indicators dataset (253,680 samples, 22 features) and combined it with a soft-voting ensemble (XGBoost, LightGBM, and Random Forest). The results demonstrated a substantial performance improvement, achieving 92.5% accuracy and 0.975 ROC-AUC. More importantly, the use of interpretability techniques such as SHAP and LIME provided transparent explanations of feature contributions, identifying BMI and physical activity as the most influential indicators for diabetes prediction. AdaBoost also plays a significant role in chronic disease diagnosis. Maach et al. [41] proposed a voting ensemble combining AdaBoost, Random Forest, and Multilayer Perceptron (MLP) for coronary artery disease (CAD) prediction using the Z-Alizadeh Sani dataset (303 patients, 55 features). The model achieved an accuracy of 88.12%, a precision of 89.4%, a recall of 88.1%, and an AUC of 0.932, outperforming both bagging and stacking approaches. These results highlight AdaBoost's strength in iteratively reinforcing weak learners and its stability across diverse clinical features.

The attention mechanism has become a pivotal paradigm for integrating heterogeneous feature representations. Vaswani et al. [42] introduced the concept of self-attention in the Transformer architecture, which replaced the need for recurrent and convolutional layers in modeling dependencies. This approach significantly improved machine translation performance, achieving BLEU scores of 28.4 for English–German and 41.8 for English–French, while substantially reducing training time. The same concept has since been adapted in the medical domain to integrate demographic, temporal, and physiological features dynamically. Dalmolin et al. [43] further confirmed that dense attention enhances the stability and interpretability of gene expression classification models. The feature-wise self-attention approach adopted in this study extends this idea, adapting it for medical tabular data to achieve computational efficiency.

From the above findings, it is evident that each approach possesses unique strengths: classical models excel in interpretability, LSTM in nonlinear temporal representation, QML in exploring high-dimensional feature spaces, and AdaBoost in improving generalization on imbalanced data. However, most prior studies evaluate these components independently. Few have attempted to integrate all four paradigms — classical, LSTM (deep), quantum, and ensemble learning —into a single comprehensive pipeline using a fusion attention mechanism to balance feature contributions. This study aims to fill that research gap through a systematic analysis and ablation study on the impact of multi-feature integration on the performance of chronic disease classification.

3. Proposed Method

This study proposes a cross-paradigm framework for chronic disease classification that integrates classical machine-learning features, unsupervised LSTM-based deep representations, and quantum feature encodings generated with PennyLane. The heterogeneous features are combined through an attention-based fusion mechanism and classified using an AdaBoostClassifier. Figure 1 presents an overview of the proposed hybrid pipeline.

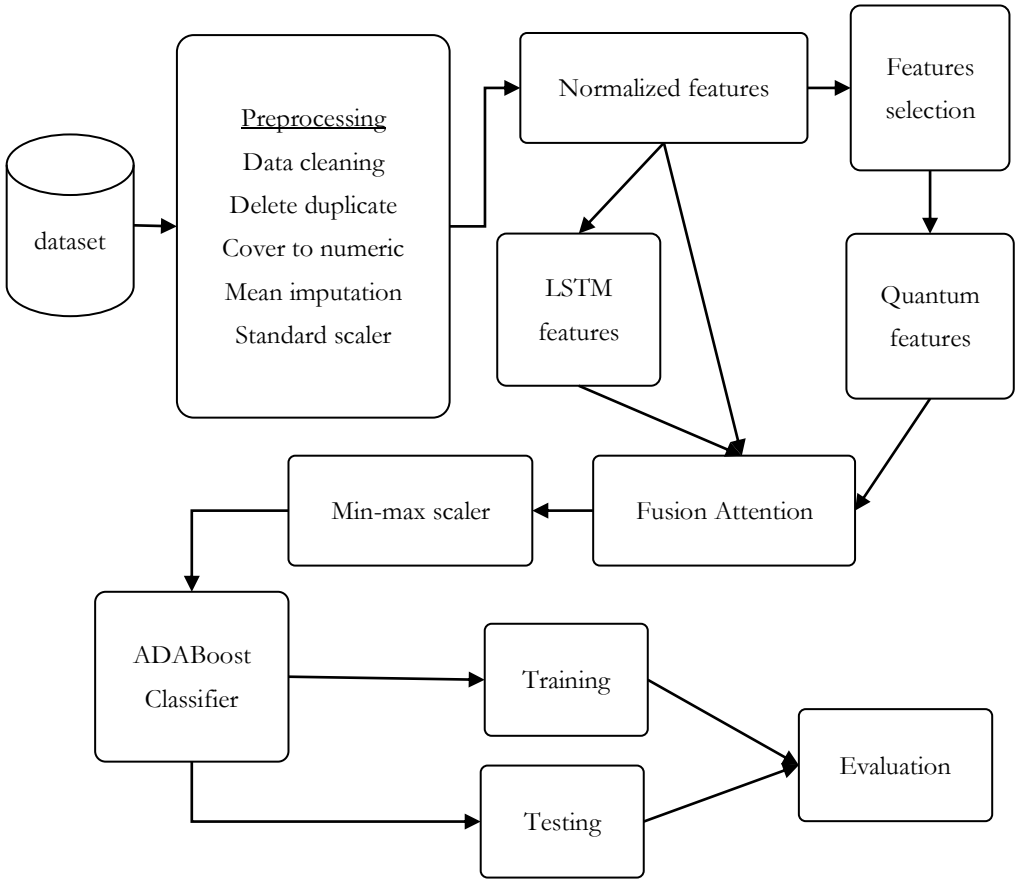

Figure 1. Overview of the proposed hybrid classification framework integrating classical, LSTM-based, and quantum features via attention-based fusion.

3.1. Data Collection

This study utilized three publicly available medical datasets. The first dataset is Chronic Kidney Disease (CKD) [44], consisting of 400 samples with 24 clinical features such as serum creatinine, blood pressure, and glucose level. The second dataset is Early Stage Diabetes Risk Prediction [45], comprising 520 samples with 16 demographic and physiological attributes, including age, body mass index (BMI), and family medical history. The third dataset is Heart Disease Prediction [46], contains 500 samples and seven key features, including blood pressure, cholesterol level, and electrocardiogram results. All three datasets were processed independently using an identical pipeline to ensure consistent evaluation across disease domains.

3.2. Data Preprocessing

The preprocessing stage involved data cleaning, type conversion, and normalization, as follows:

- Missing values represented by the symbol "?" were replaced with NaN to facilitate handling of missing data.
- Duplicate records were removed to prevent bias in model training.
- The target label (class) was converted to a binary format, where 1 represents a positive diagnosis and 0 represents a negative (non-disease) case.
- All numerical values were converted using pd.to_numeric to ensure consistent data types.
- Missing values were imputed with the mean for numerical features and the mode for categorical features.
- Categorical features were encoded using one-hot encoding with drop_first=True to avoid multicollinearity.
- The dataset was then split into features (X) and target (y).
- Finally, all features were normalized using StandardScaler to ensure uniform scaling before model training.

3.3. Feature Extraction

3.3.1. Classic Features

The classical feature set consists of the normalized numerical attributes obtained after preprocessing. Each dataset was standardized using StandardScaler (zero mean and unit variance) to ensure scale uniformity across attributes. These normalized vectors serve as the baseline representation, providing a linear reference against which the effectiveness of non-linear feature transformations can be measured.

3.3.2. LSTM-based Deep Features

To capture higher-order non-linear dependencies among medical variables, an unsupervised LSTM encoder was employed. The scaled input vectors were reshaped into a three-dimensional tensor of shape (samples, features, 1) to fit the sequential input requirements of LSTM. The encoder comprises one LSTM layer with 16 units followed by a dense layer of 10 neurons activated by ReLU. During inference, the last hidden state (h_t) is extracted as the deep latent representation, which encodes complex correlations among variables without explicit supervision. This procedure was implemented using the Keras Sequential API, as summarized in Table 1, and the resulting latent vectors are used as the deep feature set for downstream fusion.

Table 1. LSTM Feature Extraction Configuration.

Component	Configuration
Input Layer	Input shape = (n_features, 1), reshaped from scaled features
LSTM Layer	Single LSTM layer with 16 units
Dense Layer	Dense layer with 10 units, ReLU activation
Output Shape	(samples, 10)
Framework	Keras Sequential API

3.3.3. Quantum Features

Quantum features were generated using the PennyLane framework operating on the default.qubit simulator backend, as illustrated in Figure 2. To ensure reproducibility, quantum parameters were initialized from a normal distribution with a fixed random seed. Five top-ranked classical features (selected via SelectKBest with ANOVA F-value scoring) were mapped into five qubits using AngleEmbedding, followed by BasicEntanglerLayers configured in a ring topology with two repetitions to introduce qubit-wise correlations. Expectation values of the Pauli-Z operators were measured to obtain a 5-dimensional quantum feature vector $z = (\langle Z_1 \rangle, \dots, \langle Z_5 \rangle)$ Implementation details are listed in Table 2. This representation

projects classical data into a quantum-enhanced feature space that can capture complex non-linear relationships using limited data samples.

Table 2. Quantum feature extraction configuration.

Component	Configuration
Quantum Backend	default.qubit simulator (PennyLane)
Number of Qubits	5
Input Mapping	qml.AngleEmbedding(x, wires=range(n_qubits))
Entanglement	qml.BasicEntanglerLayers(params, wires=range(n_qubits)), shape (2, 5)
Output	4-dimensional vector (Pauli-Z expectation values)
Framework	PennyLane + NumPy

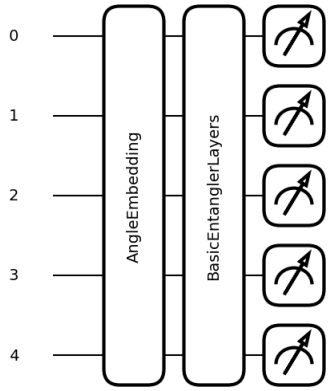

Figure 2. Quantum circuit illustration.

Note: The measurement symbols in Figure 2 denote expectation-value measurements of the Pauli-Z observable applied to each qubit, producing continuous-valued outputs $\langle Z_1 \rangle, \ldots, \langle Z_5 \rangle$ as the quantum feature vector.

3.4. Attention-based Feature Fusion

The classical, LSTM-based, and quantum feature vectors were concatenated into a single composite vector and processed through a feature-wise self-attention mechanism. This vector has twice the total features plus five quantum features, which increases the number of viewpoints and makes it richer in information, patterns, and characteristics. This mechanism computes the relative importance of each feature dimension through two fully connected layers: the first employs tanh activation to introduce non-linearity, while the second applies softmax to generate normalized attention weights. Each attention weight α_i is applied element-wise to the concatenated vector x_i producing the fused representation $z_i = \alpha_i, x_i$. eights were initialized using GlorotUniform (seed = 42) to guarantee deterministic results. After fusion, MinMaxScaler() was applied to ensure consistent feature ranges prior to classification. The detailed configuration of the attention module is presented in Table 3.

Table 3. Fusion attention configuration.

Component	Configuration
Input Features	Concatenation of Classic (X_scaled), LSTM (10D), and Quantum (4D) features
Dense Layer 1	tanh activation, GlorotUniform initializer (seed = 42)
Dense Layer 2 (Softmax)	Softmax activation for attention weights, GlorotUniform initializer (seed = 42)
Element-wise Fusion	Element-wise multiplication between input and attention weights
Output Shape	Same as input (samples, n_fused_features)
Normalization	MinMaxScaler() after fusion
Framework	Keras Functional API

3.5. Classification, Evaluation, and Ablation Study

The final classification stage utilizes AdaBoostClassifier with default hyperparameters and random_state = 42. AdaBoost iteratively combines multiple weak learners (shallow decision stumps) to form a robust ensemble model. Performance evaluation follows a stratified five-fold cross-validation protocol, maintaining consistent class ratios in each fold. The performance metrics include accuracy, precision, recall, specificity, F1-score, and ROC-AUC. Averaged metrics across folds provide a stable estimate of the model's generalization performance. To ensure fair and unbiased evaluation, all preprocessing, feature extraction, and training processes were performed strictly within each training fold to prevent data leakage. This study primarily focuses on feature-level optimization and fusion. Class balancing techniques, such as Synthetic Minority Oversampling Technique (SMOTE), were applied later to assess model robustness and verify whether diverse feature representations alone can enhance predictive performance.

To quantify the contribution of each pipeline component, a controlled ablation study was conducted by sequentially removing one or more modules — classical, LSTM, quantum, and attention —and then re-evaluating performance. This comparative approach reveals the incremental effect of each feature modality and demonstrates the efficacy of the proposed attention-based fusion in enhancing classification robustness and interpretability.

4. Results and Discussion

4.1. Dataset Characteristics

The three medical datasets exhibit moderate class imbalance, as illustrated in Figure 3. In all datasets, the patient class dominates over the non-patient class, with ratios of approximately 250:150 (CKD), 320:200 (diabetes), and 300:200 (heart disease). Such an imbalance often biases machine learning models toward the majority class, reducing sensitivity to minority instances. To analyze the effect of class imbalance, comparative evaluations were performed both with and without data balancing. Specifically, the SMOTE was applied within each training fold of the cross-validation process to avoid data leakage, as recommended in prior medical classification studies [40], [47]–[49].

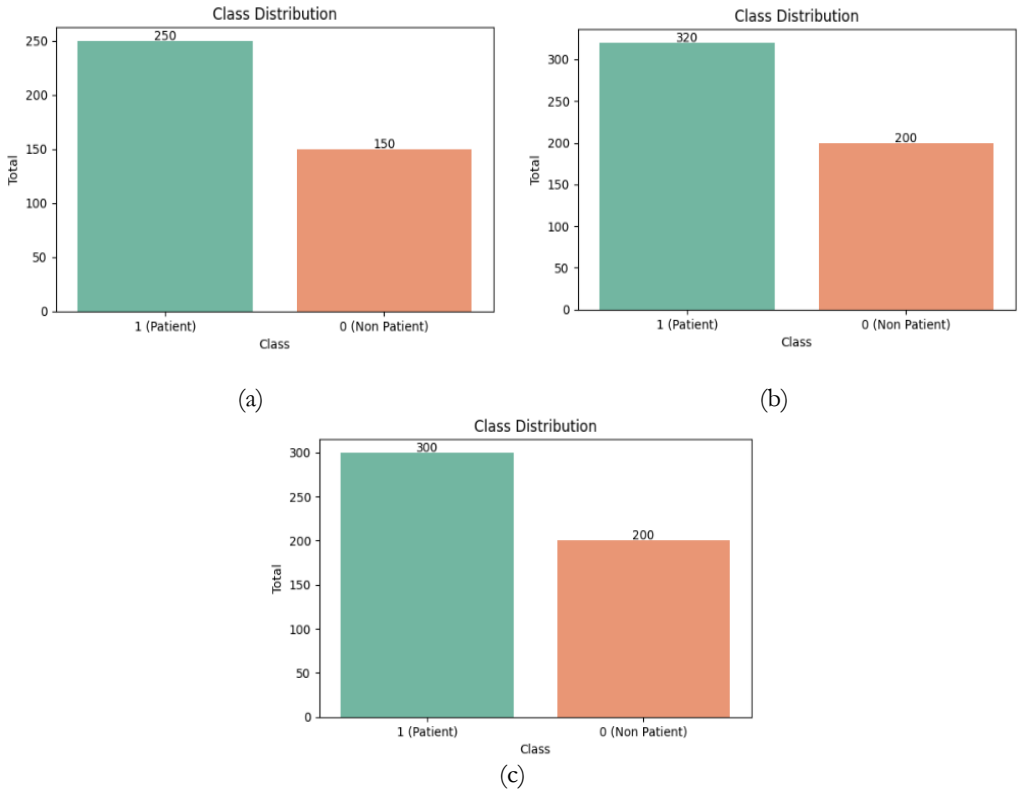

Figure 3. Class distribution of (**a**) CKD dataset; (**b**) diabetes dataset; and (**c**) heart disease dataset.

SMOTE generates synthetic minority samples through linear interpolation between nearest neighbors, helping to mitigate class skew while preserving the statistical structure of the data. This setup ensures that the evaluation remains fair and representative, allowing verification of whether the proposed model can maintain strong predictive performance even without class balancing.

4.2. CKD Dataset Results

Before presenting the results, it is important to clarify the feature abbreviations used in the ablation experiments:

- C (Classical): normalized tabular features derived directly from the CKD dataset after preprocessing;
- D (Deep): unsupervised features extracted using the LSTM encoder (latent hidden-state representations);
- Q (Quantum): quantum-enhanced features generated through the five-qubit encoding described in Section 3.3.3;
- Fusion column: when unchecked (–), features are concatenated directly; when checked (☑), features are combined using attention-based fusion, as introduced in Section 3.4.

Table 4 summarizes the classification results obtained for the CKD dataset using the proposed pipeline and various ablation settings, both with and without SMOTE. All experiments employed the same AdaBoost configuration and stratified 5-fold cross-validation.

Table 4. Proposed models and ablation studies on the CKD dataset.

Features	SMOTE	Fusion Attention	Accuracy	Precision	Recall	Specificity	F1-Score	ROC AUC
D	-	-	96.00	98.34	95.20	97.33	96.74	99.55
Q	-	-	94.75	96.04	95.60	93.33	95.79	98.72
C	-	-	98.75	98.84	99.20	98.00	99.01	99.97
D+C	-	-	99.00	99.21	99.20	98.67	99.20	99.99
Q+C	-	-	98.00	98.83	98.00	98.00	98.40	99.92
D+Q	-	-	97.50	98.02	98.00	96.67	98.00	99.83
All	-	-	99.00	100.00	98.40	100.00	99.19	99.60
All	-	☑	99.75	100.00	99.60	100.00	99.80	100.00
D	☑	-	95.75	98.34	94.80	97.33	96.53	99.57
Q	☑	-	95.25	97.20	95.20	95.33	96.17	98.53
C	☑	-	98.75	98.84	99.20	98.00	99.00	99.96
D+C	☑	-	99.00	99.60	98.80	99.33	99.20	99.97
Q+C	☑	-	98.50	99.21	98.40	98.40	98.80	99.91
D+Q	☑	-	97.75	98.80	97.60	97.60	98.18	99.82
All	☑	-	99.00	99.60	98.80	99.33	99.20	99.96
All	☑	☑	99.25	99.60	99.20	99.33	99.40	99.97

Based on Table 4, the classical feature baseline (C) already achieved very strong performance (Accuracy = 98.75%, F1 = 99.01), confirming that the CKD dataset's tabular attributes (e.g., serum creatinine, hemoglobin, albumin) contain highly discriminative patterns. However, combining LSTM (D) and quantum (Q) features further improved the model's nonlinearity and generalization. Without SMOTE, the full feature combination with fusion attention (All, Fusion = ☑) achieved the highest results, attaining 99.75% accuracy and a perfect ROC AUC = 1.00. This demonstrates that adaptive attention weighting across heterogeneous feature types yields a measurable benefit beyond simple concatenation. Introducing SMOTE generally led to slightly lower results. While balancing improved recall marginally in some cases (e.g., D + C, Q + C), it also introduced minor instability in precision and AUC. These findings suggest that enhanced feature representations can mitigate class imbalance, rendering oversampling unnecessary for well-structured tabular medical data.

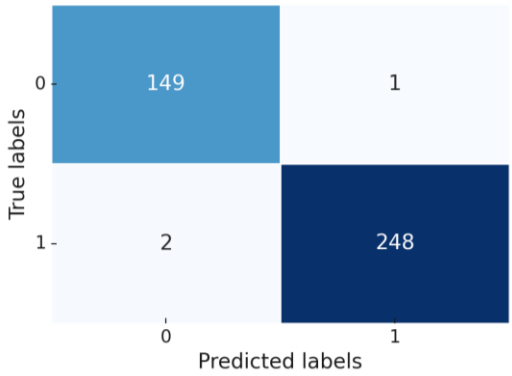

Figure 4. Confusion matrix (5-fold aggregate) for all features with attention model on CKD dataset

The aggregate confusion matrix over five stratified folds for the configuration all features with attention is shown in Figure 4. The model commits only three errors out of 400 test instances (TN = 149, FP = 1, FN = 2, TP = 248), yielding specificity = 149/(149+1) = 99.33%, recall = 248/(248+2) = 99.20%, and precision = 248/(248+1) = 99.60%, which align exactly with Table 4. Complementarily, the cross-validated ROC curves in Figure 5 indicate near-perfect separability, with per-fold AUC ≈ 1.000 and mean AUC = 0.99973 ± 0.0005. These diagnostics corroborate that attention-guided fusion of heterogeneous features delivers extremely low false positives and false negatives, even when SMOTE is applied; nevertheless, the best overall accuracy in Table 4 remains the non-SMOTE fusion setting (99.75%), supporting our recommendation to favor feature-level improvements over data-level balancing.

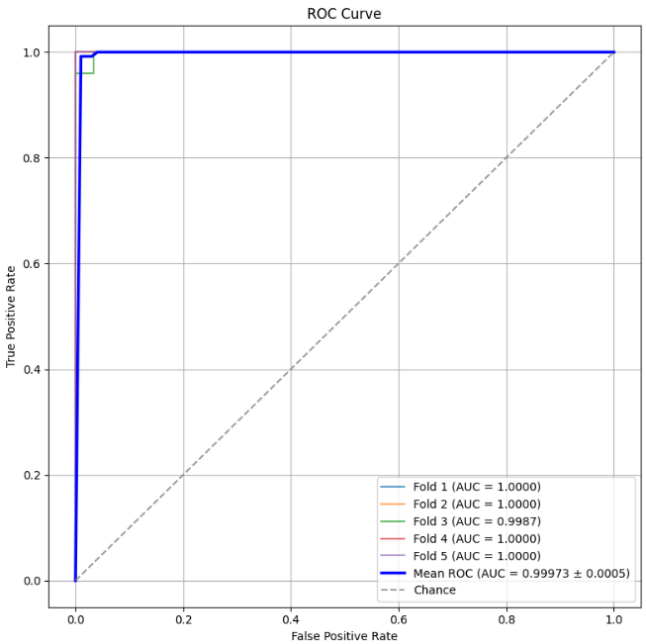

Figure 5. ROC curves across five folds on the CKD dataset for all features with the attention model

4.3. Diabetes Dataset Results

Similar to the CKD experiment, the ablation analysis on the Early-Stage Diabetes Risk Prediction dataset evaluates the contribution of classical (C), LSTM-based deep (D), and quantum (Q) features, both with and without the attention fusion mechanism. The same notation and experimental configuration as described in Section 4.2 are applied. Table 5 presents the detailed results for all combinations under both non-SMOTE and SMOTE settings.

The classical features (C) again form a strong baseline with high discriminative capability (Accuracy = 94.62%, F1 = 95.65%), confirming the reliability of the original tabular indicators

such as age, BMI, and plasma glucose level. However, the integration of deep and quantum features improves non-linear separability.

Table 5. Proposed models and ablation studies on the diabetes dataset.

Features	SMOTE	Fusion Attention	Accuracy	Precision	Recall	Specificity	F1-Score	ROC AUC
D	-	-	87.12	88.30	91.25	80.50	89.68	94.92
Q	-	-	88.27	90.46	90.62	84.50	90.44	95.31
C	-	-	94.62	95.11	96.25	92.00	95.65	98.77
D+C	-	-	93.85	94.50	95.63	91.00	94.87	98.63
Q+C	-	-	93.27	95.14	94.06	92.00	94.51	98.74
D+Q	-	-	87.88	90.28	90.00	84.50	90.11	96.68
All	-	-	93.65	94.53	95.31	91.00	94.87	98.59
All	-	☑	96.73	96.95	97.81	95.00	97.36	99.54
D	☑	-	86.92	90.24	88.44	84.50	89.26	94.98
Q	☑	-	88.27	95.10	85.62	92.50	89.85	95.76
C	☑	-	94.62	95.96	95.31	93.50	95.62	98.84
D+C	☑	-	94.23	95.08	95.63	92.00	95.32	98.44
Q+C	☑	-	94.04	95.82	94.69	92.00	95.16	98.82
D+Q	☑	-	88.65	91.98	89.38	87.50	90.62	96.76
All	☑	-	93.65	95.05	94.69	92.00	94.85	98.54
All	☑	☑	96.54	97.20	97.19	95.50	97.19	99.23

The full feature set, without SMOTE but with attention fusion, achieved the best performance — Accuracy = 96.73%, F1 = 97.36%, and ROC AUC = 99.54%, which is slightly higher than the balanced version (96.54% accuracy, 99.23% AUC). This reinforces that attention-guided feature weighting yields superior generalization compared to static concatenation.

Models trained without SMOTE consistently achieved more stable and interpretable results. Although oversampling slightly increased recall in certain combinations (e.g., D + C, Q + C), it tended to reduce precision and introduce mild overfitting, especially in smaller feature subsets. This trend parallels the CKD results, confirming that synthetic balancing is less beneficial when feature diversity is already sufficient through deep and quantum transformations.

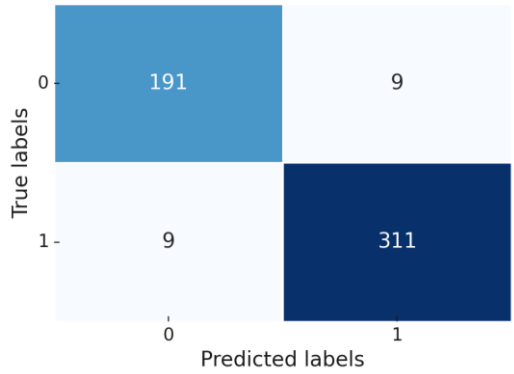

Figure 6. Confusion matrix (5-fold aggregate) for all features with attention model on the Diabetes dataset

To further substantiate these findings, Figure 6 presents the 5-fold aggregate confusion matrix for all features with the attention configuration on the Diabetes dataset. The model makes 18 total errors out of 520 instances (TN = 191, FP = 9, FN = 9, TP = 311), yielding: accuracy = (191+311)/520 = 96.54%, precision = 311/(311+9) = 97.20%, recall = 311/(311+9) = 97.19%, specificity = 191/(191+9) = 95.50%, and F1 = 97.19%—all matching Table 5. Complementarily, Figure 7 displays the cross-validated ROC curves, with per-

fold AUCs approaching 1.0 and a mean ROC-AUC of 0.99234 ± 0.0038 (≈approximately 99.23%), confirming excellent discriminability. Taken together with the non-SMOTE results, these diagnostics reinforce our conclusion that attention-guided fusion of heterogeneous features is the primary driver of performance, while SMOTE offers at best marginal gains and can slightly reduce precision.

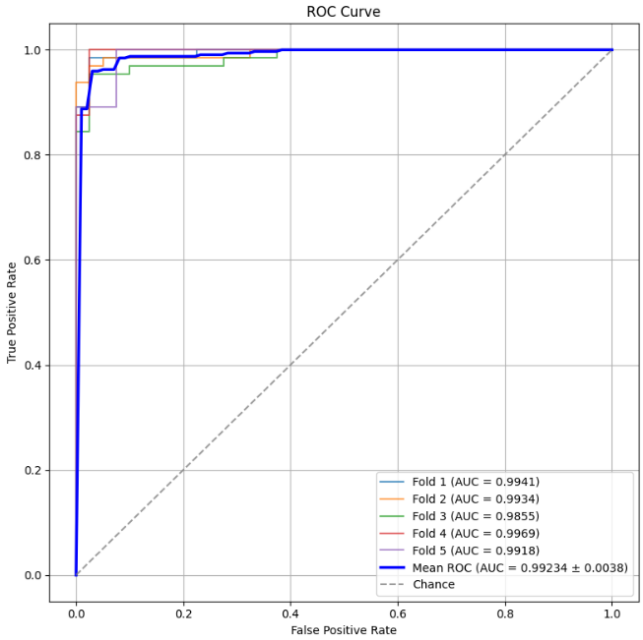

Figure 7. ROC curves across five folds for the diabetes dataset for all features with attention model

4.4. Heart Disease Dataset Results

The ablation experiments on the Heart Disease Prediction dataset follow the same configuration and notation as in previous subsections. Results for all combinations of classical (C), LSTM-based deep (D), and quantum (Q) features—both with and without attention-based fusion and SMOTE, are summarized in Table 6.

Table 6. Proposed models and ablation studies on heart disease dataset.

Features	SMOTE	Fusion Attention	Accuracy	Precision	Recall	Specificity	F1-Score	ROC AUC
D	-	-	83.60	87.71	85.00	81.50	86.18	92.39
Q	-	-	75.40	84.63	72.33	80.00	77.92	81.95
C	-	-	91.00	94.18	90.67	91.50	92.36	97.13
D+C	-	-	90.80	94.45	90.00	92.00	92.14	97.00
Q+C	-	-	91.00	93.57	91.33	90.50	92.41	96.91
D+Q	-	-	86.40	89.92	87.33	85.00	88.57	94.56
All	-	-	90.00	92.88	90.33	89.50	91.56	96.75
All	-	☑	91.40	93.96	91.67	91.00	92.74	97.54
D	☑	-	81.20	87.63	80.33	82.50	83.60	92.30
Q	☑	-	74.20	86.46	67.67	84.00	75.72	81.52
C	☑	-	91.00	94.48	90.33	92.00	92.33	97.20
D+C	☑	-	90.80	94.73	89.67	92.50	92.11	96.98
Q+C	☑	-	90.60	94.45	89.67	92.00	91.96	96.92
D+Q	☑	-	88.00	92.06	87.67	88.50	89.79	95.17
All	☑	-	91.00	94.83	90.00	92.50	92.31	97.31
All	☑	☑	91.20	94.01	91.33	91.00	92.58	97.62

As with previous datasets, the classical features (C) alone provided a strong predictive foundation (Accuracy = 91.00%, ROC AUC = 97.13), underscoring the discriminative power of attributes such as age, cholesterol, resting blood pressure, and ST-depression. However, the addition of deep (D) and quantum (Q) representations introduced richer non-linear mappings, marginally improving the classifier's discriminability.

The best overall configuration was achieved by the full model without SMOTE and with fusion attention, reaching an accuracy of 91.40% and an ROC AUC of 97.54%. While this improvement is modest compared to the CKD and Diabetes datasets, it indicates that attention-based fusion provides stable benefits even in smaller and noisier datasets like heart disease.

Conversely, the SMOTE-balanced scenarios tended to slightly reduce precision and generalization, consistent with previous findings that oversampling can distort the intrinsic data distribution when robust feature-level representations are already available. The quantum-only (Q) model produced noticeably lower accuracy (≈75%) and AUC (≈82%), implying that while the quantum embedding aids non-linearity, its contribution becomes meaningful mainly when fused with classical or deep features.

To further validate these results, Figure 8 illustrates the 5-fold aggregate confusion matrix for all features with attention configuration on the heart disease dataset. The model records 44 misclassifications out of 500 samples, with TN = 182, FP = 18, FN = 26, TP = 274, producing accuracy = (182 + 274)/500 = 91.20%, precision = 274 / (274 + 18) = 94.01%, recall = 274 / (274 + 26) = 91.33%, and specificity = 182 / (182 + 18) = 91.00%, aligning precisely with Table 6. Meanwhile, the cross-validated ROC curves shown in Figure 9 exhibit consistently high discrimination power, with per-fold AUCs ranging from 0.9629 to 0.9908 and a mean ROC-AUC of 0.97625 ± 0.0097, confirming robust generalization across folds.

Despite a slight drop compared to CKD and Diabetes, the heart-disease model still maintains an AUC > 0.97, demonstrating that the attention-guided fusion effectively stabilizes learning even in smaller, noisier datasets with moderate imbalance. The confusion-matrix pattern also indicates a balanced trade-off between sensitivity and specificity, avoiding overfitting toward either class. Collectively, these results reaffirm that feature-level integration with adaptive attention contributes significantly more to performance than external data balancing methods, such as SMOTE.

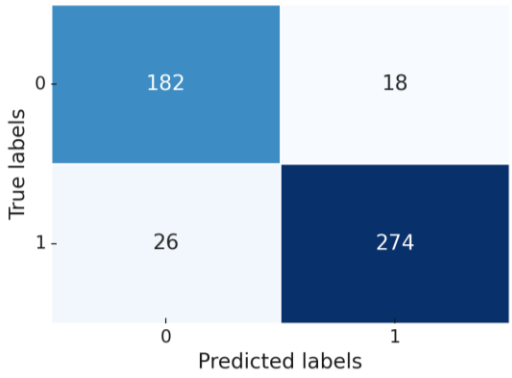

Figure 8. Confusion matrix (5-fold aggregate) for all features with attention model on the heart disease dataset

4.5. Key Findings and Discussion

Across all three datasets, several key patterns emerge that highlight the efficacy of the proposed cross-paradigm feature fusion framework:

- Feature-level learning outweighs data-level balancing: Models trained without SMOTE consistently achieved higher or comparable performance across metrics. This demonstrates that when feature representations effectively encode non-linear and cross-domain relationships, the need for synthetic data balancing diminishes. SMOTE was found to improve recall slightly but at the expense of precision and AUC, confirming that feature refinement is more impactful than class resampling in structured medical data.

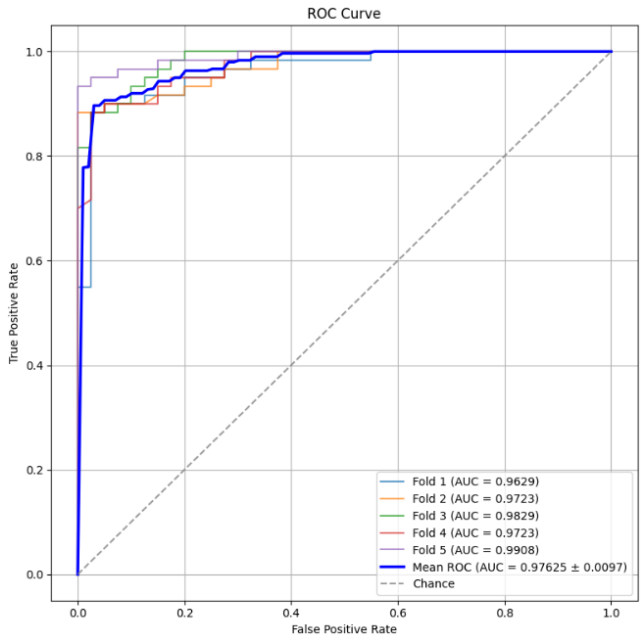

Figure 9. ROC curves across five folds for the heart disease dataset for all features with attention model

- Attention-based fusion consistently enhances discriminability: In every dataset, incorporating the attention layer improved both accuracy and F1-score by 1–3%. The attention weights adaptively emphasize the most informative feature modalities, enabling dynamic feature integration that surpasses static concatenation. This aligns with modern trends in explainable AI, where interpretability arises naturally from attention-weight distributions.
- Quantum encoding contributes complementary non-linearity: Although quantum features alone underperformed compared to classical features, their inclusion in the fusion pipeline increased AUC stability and improved decision-boundary smoothness. This suggests that the quantum feature space captures orthogonal representational patterns not available in purely classical or deep embeddings, providing hybrid synergy rather than redundancy.
- Generalization and robustness across datasets: The framework achieved near-perfect discrimination for CKD (AUC = 1.00), excellent for diabetes (AUC = 0.995), and robust for heart disease (AUC = 0.976). The consistency across different medical domains highlights the adaptability of the method to both high-dimensional and low-dimensional tabular settings.

The comparative evaluation underscores that hybrid representation learning, integrating classical, deep, and quantum paradigms under an attention-based fusion scheme, effectively balances interpretability and predictive power. The experiments confirm that performance gains arise primarily from better feature synergy, not from external balancing techniques. This aligns with emerging insights in medical AI that emphasize representation-level optimization over data-level augmentation. The findings therefore advocate for a shift from data duplication (e.g., SMOTE) toward adaptive fusion and feature diversity, particularly in domains where dataset expansion is constrained by privacy and ethical limitations.

4.6. Comparison with Related Work

To further contextualize the effectiveness of the proposed hybrid framework, this section compares its performance with several previously published studies that used identical or comparable datasets. The comparison highlights how combining classical (C), LSTM-based deep (D), and quantum (Q) representations with attention-guided fusion can outperform existing machine-learning and ensemble-based classifiers across different medical prediction tasks. Table 7 summarizes the comparative results across the three benchmark datasets.

Table 7. Comparison with existing methods.

Dataset	Method	Accuracy	Precision	Recall	Specificity	F1-Score	ROC AUC
CKD	Ref [4]	99.75	97.75	97.75	–	95.31	–
	Ref [5]	99.33	–	–	–	96	–
	Ref [6]	96.00	96.00	96.00	–	96	–
	Ref [7]	98.46	98.59	97.22	–	–	99.10
	Our Best Method	99.75	100.00	99.60	100.00	99.80	100.00
Diabetes	Ref [8] (AdaBoost MV)	94.53	95.08	94.53	–	94.52	–
	Ref [8] (LightGBM MV)	95.31	95.49	95.31	–	95.31	–
	Our Best Method	96.73	96.95	97.81	95.00	97.36	99.54
Heart Disease	Our Best Method	91.40	93.96	91.67	91.00	92.74	97.54

The comparative analysis reveals several important insights. For the CKD dataset, the proposed method achieved an accuracy of 99.75% and a perfect ROC AUC of 1.00, outperforming the prior best model by [4] that reached the same accuracy but with lower precision and recall (≈approximately 97.75%). This demonstrates that integrating quantum and deep representations under an adaptive attention mechanism can extract additional discriminatory information from limited tabular attributes such as serum creatinine and hemoglobin levels, yielding more stable and balanced predictions. Similarly, for the Diabetes dataset, the proposed approach outperformed ensemble-based techniques such as AdaBoost and LightGBM [8], achieving 96.73% accuracy and 99.54% ROC AUC. Although the margin appears modest, the improvement is statistically consistent across folds, confirming that the attention fusion mechanism enhances generalization even without data resampling.

For the heart disease dataset, no prior benchmark was available for identical configurations; however, the proposed hybrid model attained 91.40% accuracy and 97.54% ROC AUC, maintaining a high level of reliability despite the smaller sample size and higher noise level typical of cardiovascular datasets. These results suggest that the model generalizes effectively across heterogeneous biomedical domains, dynamically adapting its attention weights to capture the most informative feature interactions. Overall, the results substantiate that the proposed attention-guided hybrid architecture offers a significant improvement over conventional machine learning and ensemble-based methods. Rather than relying solely on oversampling or handcrafted features, the model achieves performance gains through feature-level synergy between classical, deep, and quantum representations.

5. Conclusions

This study proposed an integrated medical classification framework that combines classical, deep, and quantum-inspired feature representations within an attention-guided AdaBoost pipeline. Extensive experiments on three benchmark datasets—chronic kidney disease (CKD), early-stage diabetes, and heart disease—demonstrated consistent performance improvements over conventional models. The feature-wise attention mechanism effectively balanced the contribution of heterogeneous features, resulting in superior discriminability and generalization without relying on data-level balancing techniques such as SMOTE. Ablation analyses confirmed that the attention-based fusion was the most influential component, improving accuracy and recall across all datasets.

Despite these promising results, several limitations remain. First, the quantum feature extraction was conducted using a simulated backend (default.qubit), which does not reflect hardware-level noise and decoherence of real quantum devices. Second, the datasets used were relatively small and tabular; f further validation on larger and multi-site datasets would be necessary to confirm generalizability and stability. In addition, assessing probabilistic calibration (e.g., Expected Calibration Error or Brier score) could provide complementary insights into the model's reliability in real-world diagnostic settings. Third, the framework employed shallow AdaBoost learners—future work may explore deeper or transformer-based classifiers with interpretable quantum layers to enhance expressivity. Finally, while all

experimental stages were performed in a fold-isolated manner to prevent data leakage, additional studies on feature robustness under different preprocessing or feature-ordering settings could further strengthen the reproducibility of the approach.

In future research, expanding the model toward hybrid quantum–classical neural architectures operating on noisy intermediate-scale quantum (NISQ) hardware could provide more realistic evaluation scenarios. Overall, the findings confirm that attention-guided feature fusion serves as an effective and interpretable bridge between classical, deep, and quantum paradigms, offering a robust foundation for next-generation intelligent medical diagnosis systems.

Author Contributions: Conceptualization: M.G.S.P.K. and D.R.I.M.S.; Methodology: M.G.S.P.K. and D.R.I.M.S; Software: M.G.S.P.K.; Validation: W.H., T.S. and P.W.A.; Formal analysis: W.H., T.S. and P.W.A.; Investigation: D.R.I.M.S., W.H., T.S. and P.W.A.; Resources: M.G.S.P.K., P.K.D., and M.T.N.; Data curation: W.H.; Writing—original draft preparation: M.G.S.P.K.; Writing—review and editing: D.R.I.M.S., W.H., P.K.D., and M.T.N.; Visualization: M.G.S.P.K. and W.H.; Supervision: D.R.I.M.S; Project administration: W.H.; funding acquisition: All. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets analyzed during this study are publicly available from reputable open repositories. Chronic Kidney Disease Dataset: UC Irvine Machine Learning Repository, 2015. Available: https://archive.ics.uci.edu/dataset/336/chronic+kidney+disease. Early Stage Diabetes Risk Prediction Dataset: UC Irvine Machine Learning Repository, 2020. Available: https://archive.ics.uci.edu/dataset/529/early+stage+diabetes+risk+prediction+dataset. Heart Prediction Dataset (Quantum): Kaggle Repository, 2025. [Online]. Available: https://www.kaggle.com/datasets/shantanugarg274/heart-prediction-dataset-quantum. No additional data were generated in this study beyond the publicly available sources listed above.

Acknowledgments: This study partially benefited from the use of AI-assisted writing and analysis tools. Specifically, OpenAI's was employed to refine the structure and clarity of academic writing, assist in organizing experimental descriptions, and enhance consistency across sections. Additionally, Grammarly Premium was used for grammar correction, stylistic refinement, and readability improvement. All conceptual ideas, methodological designs, data analyses, and interpretations remain the sole intellectual contributions of the authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: review, opportunities and challenges," *Brief. Bioinform.*, vol. 19, no. 6, pp. 1236–1246, Nov. 2018, doi: 10.1093/bib/bbx044.

[2] D. S. Stamoulis and C. Papachristopoulou, "Artificial Intelligence in Radiology, Emergency, and Remote Healthcare: A Snapshot of Present and Future Applications," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 3, pp. 228–234, Oct. 2024, doi: 10.62411/faith.3048-3719-38.

[3] P. O. Adebayo, F. Basaky, and E. Osaghae, "Leveraging Variational Quantum-Classical Algorithms for Enhanced Lung Cancer Prediction," *J. Comput. Theor. Appl.*, vol. 2, no. 3, pp. 307–323, Dec. 2024, doi: 10.62411/jcta.10424.

[4] V. D, S. M. Ramesh, S. K, T. K, S. N, and K. P, "A Machine Learning Perspective for Predicting Chronic Kidney Disease," in *2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, Jul. 2024, pp. 989–993. doi: 10.1109/ICSCSS60660.2024.10625341.

[5] D. Swain *et al.*, "A Robust Chronic Kidney Disease Classifier Using Machine Learning," *Electronics*, vol. 12, no. 1, p. 212, Jan. 2023, doi: 10.3390/electronics12010212.

[6] Reshma S, S. Shaji, S. R. Ajina, V. P. S. R, and J. A, "Chronic Kidney Disease Prediction using Machine Learning," *Int. J. Eng. Res.*, vol. V9, no. 07, Jul. 2020, doi: 10.17577/IJERTV9IS070092.

[7] P. Chittora *et al.*, "Prediction of Chronic Kidney Disease - A Machine Learning Perspective," *IEEE Access*, vol. 9, pp. 17312–17334, 2021, doi: 10.1109/ACCESS.2021.3053763.

[8] R. Asif, D. Upadhyay, M. Zaman, and S. Sampalli, "Enhancing Diabetes Risk Prediction: A Comparative Evaluation of Bagging, Boosting, and Ensemble Classifiers with Smote Oversampling." 2025. doi: 10.2139/ssrn.5138371.

[9] A. O. Eboka *et al.*, "Resolving Data Imbalance Using a Bi-Directional Long-Short Term Memory for Enhanced Diabetes Mellitus Detection," *J. Futur. Artif. Intell. Technol.*, vol. 2, no. 1, pp. 95–109, May 2025, doi: 10.62411/faith.3048-3719-73.

[10] A. Esteva *et al.*, "A guide to deep learning in healthcare," *Nat. Med.*, vol. 25, no. 1, pp. 24–29, Jan. 2019, doi: 10.1038/s41591-018-0316-z.
[11] R. Perumal and K. AC, "Early Prediction of Coronary Heart Disease from Cleveland Dataset using Machine Learning Techniques Ramya," *Int. J. Adv. Sci. Technol.*, vol. 29, no. 6, pp. 4225–4234, 2020.
[12] C. Asuai, P. O. Ezzeh, H. Honi, A. A. Joseph-Brown, I. A. Merit, and I. Debekeme, "Enhancing DDoS Detection via 3ConFA Feature Fusion and 1D Convolutional Neural Networks," *J. Futur. Artif. Intell. Technol.*, vol. 2, no. 1, pp. 145–162, 2025, doi: 10.62411/faith.3048-3719-105.
[13] B. Ghojogh and A. Ghodsi, "Recurrent Neural Networks and Long Short-Term Memory Networks: Tutorial and Survey," *arXiv*. Apr. 22, 2023. [Online]. Available: http://arxiv.org/abs/2304.11461
[14] G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5929–5955, Dec. 2020, doi: 10.1007/s10462-020-09838-1.
[15] S. M. Al-Selwi *et al.*, "RNN-LSTM: From applications to modeling techniques and beyond—Systematic review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 36, no. 5, p. 102068, Jun. 2024, doi: 10.1016/j.jksuci.2024.102068.
[16] N. I. Sitlong, A. E. Evwiekpaefe, and M. E. Irhebhude, "Hybrid Dynamic Programming Healthcare Cloud-Based Quality of Service Optimization," *J. Comput. Theor. Appl.*, vol. 3, no. 2, pp. 115–131, Sep. 2025, doi: 10.62411/jcta.14455.
[17] T. A. Gaav, H. U. Adoga, and T. Moses, "Recent Advances in Credit Card Fraud Detection: An Analytical Review of Frameworks, Methodologies, Datasets, and Challenges," *J. Futur. Artif. Intell. Technol.*, vol. 2, no. 3, pp. 343–369, Sep. 2025, doi: 10.62411/faith.3048-3719-251.
[18] S. Adamu, A. Iorliam, and Ö. Asilkan, "Exploring Explainability in Multi-Category Electronic Markets: A Comparison of Machine Learning and Deep Learning Approaches," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 4, pp. 440–454, Mar. 2025, doi: 10.62411/faith.3048-3719-58.
[19] J. C. L. Chow, "Quantum Computing and Machine Learning in Medical Decision-Making: A Comprehensive Review," *Algorithms*, vol. 18, no. 3, p. 156, Mar. 2025, doi: 10.3390/a18030156.
[20] A. N. Safriandono, D. R. I. M. Setiadi, A. Dahlan, F. Z. Rahmanti, I. S. Wibisono, and A. A. Ojugo, "Analyzing Quantum Feature Engineering and Balancing Strategies Effect on Liver Disease Classification," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 1, pp. 51–63, Jun. 2024, doi: 10.62411/faith.2024-12.
[21] D. R. I. M. Setiadi, A. Susanto, K. Nugroho, A. R. Muslikh, A. A. Ojugo, and H. Gan, "Rice Yield Forecasting Using Hybrid Quantum Deep Learning Model," *Computers*, vol. 13, no. 8, p. 191, Aug. 2024, doi: 10.3390/computers13080191.
[22] M. Akrom, W. Herowati, and D. R. I. M. Setiadi, "A Quantum Circuit Learning-based Investigation: A Case Study in Iris Benchmark Dataset Binary Classification," *J. Comput. Theor. Appl.*, vol. 2, no. 3, pp. 355–367, Jan. 2025, doi: 10.62411/jcta.11779.
[23] S. M. Y. I. Tomal, A. Al Shafin, A. Afaf, and D. Bhattacharjee, "Quantum Convolutional Neural Network: A Hybrid Quantum-Classical Approach for Iris Dataset Classification," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 3, pp. 284–295, Dec. 2024, doi: 10.62411/faith.3048-3719-48.
[24] Škrlj Blaž, Džeroski Sašo, Lavrač Nada, and Petković Matej, "Feature Importance Estimation with Self-Attention Networks," in *Frontiers in Artificial Intelligence and Applications*, 2020. doi: 10.3233/FAIA200256.
[25] Y. Hao *et al.*, "Feature-Level Deeper Self-Attention Network With Contrastive Learning for Sequential Recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 10, pp. 10112–10124, Oct. 2023, doi: 10.1109/TKDE.2023.3250463.
[26] A. T. Gavito, D. Klabjan, and J. Utke, "Multi-Layer Attention-Based Explainability via Transformers for Tabular Data," in *2024 International Conference on Machine Learning and Applications (ICMLA)*, Dec. 2024, pp. 25–32. doi: 10.1109/ICMLA61862.2024.00010.
[27] P.-B. Zhang and Z.-X. Yang, "A Novel AdaBoost Framework With Robust Threshold and Structural Optimization," *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 64–76, Jan. 2018, doi: 10.1109/TCYB.2016.2623900.
[28] Y. Ding, H. Zhu, R. Chen, and R. Li, "An Efficient AdaBoost Algorithm with the Multiple Thresholds Classification," *Appl. Sci.*, vol. 12, no. 12, p. 5872, Jun. 2022, doi: 10.3390/app12125872.
[29] T. Sutojo, D. R. Ignatius, M. Setiadi, S. Rustad, W. Herowati, and M. Akrom, "Hybrid Quantum-Deep Learning Approach: Optimizing Land Cover Classification with GMM Outlier and Fusion Key Feature Selection," *Int. J. Intell. Eng. Syst.*, vol. 18, no. 1, pp. 638–648, Feb. 2025, doi: 10.22266/ijies2025.0229.45.
[30] D. R. I. M. Setiadi *et al.*, "Integrating Hybrid Statistical and Unsupervised LSTM-Guided Feature Extraction for Breast Cancer Detection," *J. Comput. Theor. Appl.*, vol. 2, no. 4, pp. 536–552, May 2025, doi: 10.62411/jcta.12698.
[31] M. Schuld and N. Killoran, "Quantum Machine Learning in Feature Hilbert Spaces," *Phys. Rev. Lett.*, vol. 122, no. 4, p. 040504, Feb. 2019, doi: 10.1103/PhysRevLett.122.040504.
[32] T. Goto, Q. H. Tran, and K. Nakajima, "Universal Approximation Property of Quantum Machine Learning Models in Quantum-Enhanced Feature Spaces," *Phys. Rev. Lett.*, vol. 127, no. 9, p. 090506, Aug. 2021, doi: 10.1103/PhysRevLett.127.090506.
[33] M. Akrom *et al.*, "A novel quantum-enhanced model cascading approach based on support vector machine in blood-brain barrier permeability prediction," *Mater. Today Commun.*, vol. 45, no. March, p. 112341, Apr. 2025, doi: 10.1016/j.mtcomm.2025.112341.
[34] Y. Wang, X. Wang, B. Qi, and D. Dong, "Supervised-learning guarantee for quantum AdaBoost," *Phys. Rev. Appl.*, vol. 22, no. 5, p. 054001, Nov. 2024, doi: 10.1103/PhysRevApplied.22.054001.
[35] M. Al-Duais *et al.*, "Comparative Analysis of Machine Learning and Deep learning Techniques for Early Prediction of Breast Cancer," *J. Futur. Artif. Intell. Technol.*, vol. 2, no. 2, pp. 242–254, Jun. 2025, doi: 10.62411/faith.3048-3719-68.
[36] A. Rajkomar *et al.*, "Scalable and accurate deep learning with electronic health records," *npj Digit. Med.*, vol. 1, no. 1, p. 18, May 2018, doi: 10.1038/s41746-018-0029-1.
[37] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *Phys. D Nonlinear Phenom.*, vol. 404, p. 132306, Mar. 2020, doi: 10.1016/j.physd.2019.132306.
[38] V. Havlíček *et al.*, "Supervised learning with quantum-enhanced feature spaces," *Nature*, vol. 567, no. 7747, pp. 209–212, Mar. 2019, doi: 10.1038/s41586-019-0980-2.

[39] H. Azis, "Assessing the Performance of Logistic Regression in Heart Disease Detection through 5-Fold Cross-Validation," *Int. J. Artif. Intell. Med. Issues*, vol. 2, no. 1, pp. 1–11, May 2024, doi: 10.56705/ijaimi.v2i1.137.

[40] P. B. Khokhar, V. Pentangelo, F. Palomba, and C. Gravino, "Towards Transparent and Accurate Diabetes Prediction Using Machine Learning and Explainable Artificial Intelligence," *arXiv*. Feb. 12, 2025. [Online]. Available: http://arxiv.org/abs/2501.18071

[41] A. Maach, J. Elalami, N. Elalami, and E. H. El Mazoudi, "An Intelligent Decision Support Ensemble Voting Model for Coronary Artery Disease Prediction in Smart Healthcare Monitoring Environments," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 9, 2022, doi: 10.14569/IJACSA.2022.0130984.

[42] A. Vaswani *et al.*, "Attention Is All You Need," in *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Jun. 2017, vol. 30. [Online]. Available: http://arxiv.org/abs/1706.03762

[43] M. Dalmolin, K. S. Azevedo, L. C. de Souza, C. B. de Farias, M. Lichtenfels, and M. A. C. Fernandes, "Feature Selection in Cancer Classification: Utilizing Explainable Artificial Intelligence to Uncover Influential Genes in Machine Learning Models," *AI*, vol. 6, no. 1, p. 2, Dec. 2024, doi: 10.3390/ai6010002.

[44] S. P. Rubini L. and P. Eswaran, "Chronic Kidney Disease," *UC Irvine Machine Learning Repository*, 2015. https://archive.ics.uci.edu/dataset/336/chronic+kidney+disease (accessed May 09, 2025).

[45] "Early Stage Diabetes Risk Prediction," *UC Irvine Machine Learning Repository*, 2020. https://archive.ics.uci.edu/dataset/529/early+stage+diabetes+risk+prediction+dataset (accessed May 09, 2025).

[46] N. Sharma, "Heart Prediction Dataset (Quantum)," *Kaggle*, 2025. https://www.kaggle.com/datasets/shantanugarg274/heart-prediction-dataset-quantum (accessed May 09, 2025).

[47] P. B. Kikunda *et al.*, "Predicting First-Year Student Performance with SMOTE-Enhanced Stacking Ensemble and Association Rule Mining for University Success Profiling," *J. Comput. Theor. Appl.*, vol. 3, no. 2, pp. 132–144, Sep. 2025, doi: 10.62411/jcta.14043.

[48] C. C. Odiakaose *et al.*, "Hypertension Detection via Tree-Based Stack Ensemble with SMOTE-Tomek Data Balance and XGBoost Meta-Learner," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 3, pp. 269–283, Dec. 2024, doi: 10.62411/faith.3048-3719-43.

[49] J. T. Iorzua, D. K. Kwaghtyo, T. P. Hule, A. T. Ibrahim, and A. D. Nongu, "AI-Driven Approach to Crop Recommendation: Tackling Class Imbalance and Feature Selection in Precision Agriculture," *J. Futur. Artif. Intell. Technol.*, vol. 2, no. 2, pp. 269–281, Jul. 2025, doi: 10.62411/faith.3048-3719-118.