



International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

Comparative Analysis of Classical and Quantum Support Vector Machines for Diabetes Prediction Using PCA-Based Feature Reduction

Sandip Kumar Singh Modak^{1*}, Ashish Kumar Sinha¹, Avinash Kumar¹

¹Department of Computer Science & Engineering, Sarala Birla University, Ranchi, Jharkhand, India Email: {sandip.modak@sbu.ac.in*, ashish.sinha@sbu.ac.in, avinash.kumar@sbu.ac.in}

Abstract

A chronic metabolic disease, diabetes mellitus has become a serious global public health concern. Using the Indian Pima Diabetes Dataset, this study compares Classical Support Vector Machines (CSVM) with Quantum Support Vector Machines (QSVM) for diabetes prediction. To address the challenges associated with high-dimensional data and feature redundancy, before classification, a feature reduction method called Principal Component Analysis (PCA) is used. Experimental results demonstrate that PCA effectively reduces computational complexity while preserving essential information required for classification. Furthermore, the study investigates the potential advantages of quantum-enhanced learning in healthcare prediction tasks by comparing the predictive capabilities and computational characteristics of QSVM against its classical counterpart. The findings provide insights into the applicability of quantum machine learning techniques for medical diagnosis and highlight the opportunities and challenges associated with implementing quantum algorithms on real-world healthcare datasets.

Keyword: Diabetes Prediction, Quantum Machine Learning, Quantum Support Vector Machine (QSVM), Classical Support Vector Machine (CSVM), Principal Component Analysis (PCA)

Introduction

Machine learning (ML), which can analyze big, high-dimensional datasets and find intricate, non-linear correlations between variables, has become a potent tool in the healthcare industry in recent years [1]. In simple terms, ML refers to computational methods that “learn” patterns from data and use them to make predictions, often without requiring predefined assumptions about how variables are related. Supervised learning approaches, commonly used in medical prediction, train algorithms on labeled datasets to predict outcomes such as disease onset. Techniques such as random forests (which combine multiple decision trees), support vector machines (which separate data into classes using optimal boundaries), neural networks (which model layered patterns similar to brain-like processing), and gradient boosting (which iteratively improves prediction accuracy) have shown strong performance across medical applications [2]. The objective of quantum technologies is to illustrate their potential in contrast to classical machine learning; yet, this reveals flaws such as qubit limitations and continuous logic gate operations [3]. Logistic regression, artificial neural network (ANN), linear discriminant analysis (LDA), naive Bayes (NB), support vector machine (SVM), decision tree (DT), AdaBoost (AB), J48, k-nearest neighbors (k-NN), quadratic discriminant analysis (QDA), random forest (RF), multilayer perceptron (MLP), general regression neural network, and radial basis function (RBF) are some of the ML and DL-based frameworks that have been proposed in recent years for the prediction of diabetes [4]. To improve the performance of ML and DL models, these methods have been combined with different dimensionality reduction (like principal component analysis) and cross validation (like k-fold cross-validation) techniques, as well as the mechanism for filling in missing values and rejecting outliers. Three distinct machine learning classifiers (DT, SVM, and NB) have been used to predict the probability of diabetes with the highest accuracy. The results show that NB offers the highest accuracy of 82% [5]. Because these ML and DL models understand the relationship between input (features) and output (class) based on the traditional theories of probability and logic, their accuracy is still on the lower side even though they have demonstrated satisfactory outcomes [6, 7]. As a result, they need to be much improved in order to be widely accepted for diabetes prediction. Furthermore, Quantum Mechanics (QM) has previously demonstrated its efficacy in a number of fields, including object identification, tracking, prediction, and

classification, and it has outperformed models based on conventional theory. Utilizing basic quantum mechanical concepts, quantum computing has become a revolutionary computational paradigm that solves issues that are beyond the capabilities of classical computers [8]. Quantum algorithms' potential for exponential or polynomial speedups has garnered a lot of interest in fields like scientific simulations, cryptography, and optimization. Simultaneously, the quick development of machine learning (ML) and artificial intelligence (AI) has resulted in more sophisticated deep learning models, which have intrinsic constraints when handling high-dimensional data with traditional hardware [9]. By incorporating quantum computing concepts—superposition, entanglement, and quantum parallelism—into traditional machine learning pipelines, quantum machine learning (QML) aims to connect these two domains. In order to improve feature representation, optimization efficiency, and model expressivity, new learning paradigms such as Quantum Neural Networks (QNNs), Quantum Convolutional Neural Networks (QCNNs), and hybrid quantum–classical architectures have been developed as a result of this integration [10]. Because of this, QML has become a viable area of study for resolving computational constraints in contemporary deep learning.

Related Work

Naramparambath et al. [11] presented a novel quantum machine learning system in 2026 that uses a unique randomizer based on cryptographic principles and a specific, custom-designed quantum circuit. To achieve unbiased feature set selection, two distinct cryptographic techniques—the Hash-based Randomizer and the ChaCha20 Randomizer—were combined with three specially created quantum circuits. Additionally, the framework outperforms current machine learning algorithms in terms of prediction accuracy and handling complex non-linear data, including 41 real-world biomarkers, with the potential for improved scalability for adaptation toward more intricate physiological metrics.

Four algorithms (classical SVM, classical RF, quantum SVM, and quantum RF) were used by Upama et al. [12], and the quantum components were run on a classical computer in a virtual quantum environment. Out of the four suggested methods, quantum SVM and classical random forest show the best results based on performance metrics. Quantum equivalents of machine learning algorithms are more effective at classifying diseases and making early predictions, as demonstrated by QSVM's superior performance over classical SVM.

The effectiveness of the VQC is influenced by the amount of qubits, feature map types, optimizer class, number of layers in the parametrized circuit, and number of learnable parameters in ansatz, according to Khan et al. [13]. In order to forecast diabetes, 76 variations of VQC are examined for scenarios involving four and eight qubits, and the outcomes are contrasted with six traditional machine learning methods. During analysis, three different optimizers (COBYLA, SPSA, and SLSQP) are used along with three different feature map types (Pauli, Z, and ZZ). The PIMA Indian Diabetes Dataset (PIDD) is used in the experiments. The findings show that VQC performs better than other quantum variations when it has six layers integrated with an error correction scaling factor of 0.01, a ZZ feature map, and a COBYLA optimizer. For the situations of eight and four qubits, respectively, the best suggested model achieved an accuracy of 0.85 and 0.80.

The impact of QNN and QSVM models on healthcare was suggested by Tudisco et al. [14]. By contrasting Quantum Neural Networks (QNNs) and Quantum Support Vector Machines (QSVMs) with well-known classical models, this study assesses the potential of quantum classifiers in healthcare. Three well-known healthcare datasets—diabetes, heart failure, and prostate cancer—form the basis of the study. The findings show that because QSVMs are more prone to overfitting, they perform better than QNNs on all datasets.

A diabetes prediction model using machine learning and quantum machine learning techniques was proposed by Vimal et al. [15]. With an accuracy of 0.76, the evaluation demonstrates that Support Vector Machine (SVM) outperforms Logistic Regression (LR). However, among QML algorithms, the Quantum Support Vector Classifier (QSVC) outperforms the Variational Quantum Classifier (VQC) with an accuracy of 0.74.

The performance study of quantum machine learning with deep learning for diabetes prediction was investigated by Gupta et al. [16]. The main goal of this endeavor is to create a prognosis tool based on the PIMA Indian Diabetes dataset that will assist healthcare professionals in lowering the death rate from diabetes. For the DL model, performance metrics like precision, accuracy, recall, F1 score, specificity, balanced accuracy, false detection rate, missed detection rate, and diagnostic odds ratio were obtained as 0.90, 0.95, 0.95, 0.93, 0.95, 0.95, 0.03, 0.02, and 399.00, respectively. For the QML model, these metrics were calculated as 0.74, 0.86, 0.85, 0.11, 0.05, and 35.89, respectively.

Proposed Framework

The proposed framework presents a comparative analysis of Classical Support Vector Machine (SVM) and Quantum Support Vector Machine (QSVM) for diabetes prediction. Both models are trained and evaluated using the same preprocessed dataset to ensure a fair comparison. Principal Component Analysis (PCA) is employed to reduce the dimensionality of the dataset before classification, thereby decreasing computational complexity and

mitigating redundant features. The architecture of the proposed comparative framework is illustrated in Fig. 1. The framework consists of data preprocessing, PCA-based feature reduction, parallel implementation of Classical SVM and Quantum SVM, and performance evaluation based on classification accuracy and computational time.

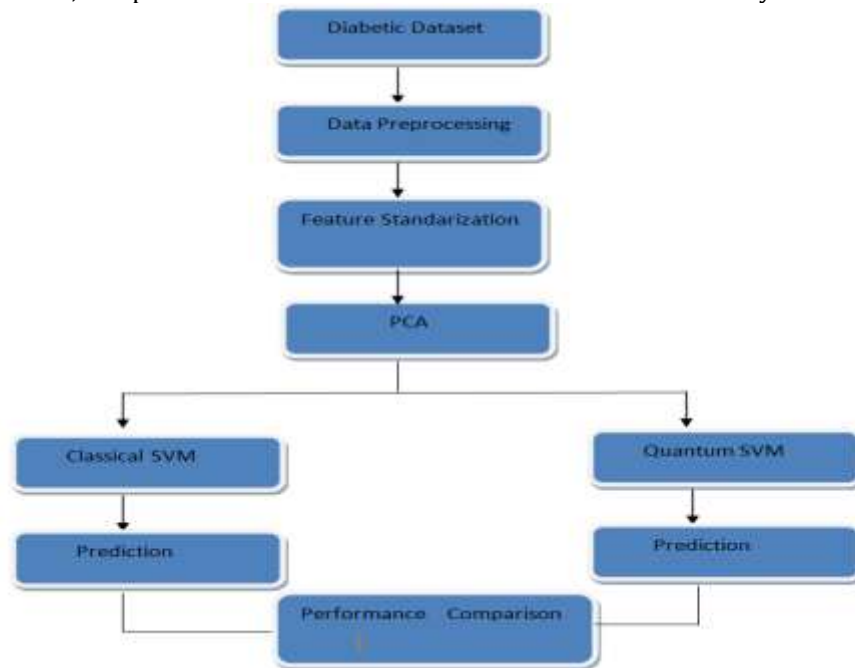


Fig. 1. The Proposed Framework

Data Collection

The diabetes dataset is first collected from a publicly available repository (PIMA). The dataset consists of multiple clinical attributes related to patients, including glucose level, blood pressure, BMI, insulin, skin thickness, diabetes pedigree function, age, and outcome.

Data Preprocessing

The preprocessing stage includes:

- Removal of missing or invalid values
- Handling inconsistent records
- Conversion of categorical variables
- Feature scaling using StandardScaler

Since SVM and QSVM are distance-based learning algorithms, feature scaling is essential to ensure that all features contribute equally during model training.

PCA-Based Feature Reduction

After normalization, Principal Component Analysis (PCA) is applied to transform the original feature space into a reduced-dimensional feature space while preserving the maximum variance.

Classical Support Vector Machine

The reduced feature set is first classified using the Classical Support Vector Machine. The SVM constructs an optimal separating hyperplane by maximizing the margin between diabetic and non-diabetic samples. An RBF kernel is employed to capture nonlinear relationships among the clinical attributes.

Quantum Support Vector Machine

The same PCA-transformed dataset is then provided to the Quantum Support Vector Machine. Instead of using a classical kernel, QSVM computes similarities using a quantum kernel generated from quantum circuits.

The workflow includes:

- Encoding classical data into quantum states
- Applying a quantum feature map (ZZFeatureMap)
- Computing the quantum kernel matrix

- Training an SVM using the quantum kernel

Performance Evaluation

Both classifiers are evaluated using identical testing data.

The comparison includes:

- Accuracy
- Precision
- Recall
- F1-score
- Confusion Matrix
- Execution Time

The execution time is also measured to compare the computational efficiency of both approaches.

Experiments and results

The performance comparison between the Classical Support Vector Machine (SVM) and the Quantum Support Vector Machine (QSVM) is presented in Table 1. The Classical SVM achieved an accuracy of 70.13%, whereas the QSVM achieved an accuracy of 52.50% on the diabetes dataset after PCA-based feature reduction. In terms of computational efficiency, the Classical SVM completed the training and prediction process in approximately 0.01 seconds, while the QSVM required 56.10 seconds to compute the quantum kernel and perform classification. The substantially higher execution time of the QSVM is primarily due to the computational overhead associated with quantum kernel evaluation on a classical simulator.

The experimental results indicate that, for the selected dataset and feature representation, the Classical SVM outperformed the QSVM in both prediction accuracy and execution speed. The QSVM's lower performance can be attributed to several factors, including the limited expressiveness of the chosen quantum feature map, the small number of qubits available after PCA, noise introduced by quantum simulation, and the absence of quantum hardware advantages when executed on a classical simulator.

Table 1. Performance Comparison of Classical SVM and QSVM

Model	Accuracy	Execution Time (s)
Classical SVM	0.7013 (70.13%)	0.01
Quantum SVM (QSVM)	0.5250 (52.50%)	56.10

The performance of the Classical Support Vector Machine (SVM) and Quantum Support Vector Machine (QSVM) was evaluated using accuracy, precision, recall, F1-score, and computational execution time. The results are presented in Table 2.

Table 2. Comparative Performance of Classical SVM and QSVM Using Accuracy, Precision, Recall, F1-Score, and Execution Time

Performance Metric	Classical SVM	QSVM
Accuracy	70.13%	57.50%
Precision	60.00%	61.54%
Recall	44.44%	40.00%
F1-score	51.06%	48.48%
Execution Time	0.01 s	52.51 s

The results indicate that the Classical SVM achieved better overall classification performance than the QSVM. The Classical SVM obtained an accuracy of 70.13%, compared with 57.50% for the QSVM. Similarly, the Classical SVM achieved higher recall (44.44%) and F1-score (51.06%) than the QSVM, which obtained recall and F1-score values of 40.00% and 48.48%, respectively. However, the QSVM achieved a slightly higher precision of 61.54%, compared with 60.00% for the Classical SVM.

The computational results show a substantial difference between the two approaches. The Classical SVM required only 0.01 seconds, whereas the QSVM required 52.51 seconds for execution. This higher computational cost is expected in the present implementation because the QSVM involves quantum feature mapping and quantum-kernel computation.

Overall, the Classical SVM demonstrates superior performance under the current experimental configuration in terms of accuracy, recall, F1-score, and execution time, while the QSVM provides a marginal improvement in precision. However, it should be noted that the Classical SVM and QSVM results were obtained using different numbers of evaluation samples (154 and 40, respectively).

Figure 2 presents the accuracy comparison between the Classical Support Vector Machine (SVM) and the Quantum Support Vector Machine (QSVM). The Classical SVM achieves an accuracy of 70.13%, whereas the QSVM achieves an accuracy of 57.50%. As illustrated in the figure, the Classical SVM demonstrates approximately 12.63 percentage points higher accuracy than the QSVM under the current experimental configuration.

The observed difference indicates that the Classical SVM provides better overall classification performance for the evaluated dataset. However, the QSVM result should be interpreted cautiously because the two models were evaluated using different numbers of test samples. A fair comparison requires both models to be evaluated using the same dataset partition, preprocessing procedure, and test set.

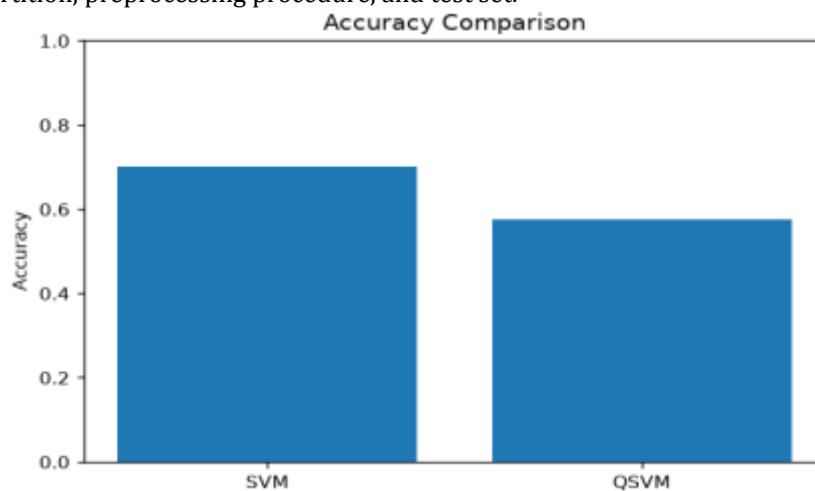


Fig. 2. Accuracy Comparison between Classical SVM and Quantum Support Vector Machine (QSVM)

Figure 3 illustrates the comparative performance of the Classical SVM and QSVM models in terms of precision, recall, and F1-score. The QSVM achieves a slightly higher precision of 61.54%, compared with 60.00% for the Classical SVM. This indicates that the QSVM produces a marginally higher proportion of correct positive predictions.

In contrast, the Classical SVM performs better in terms of recall, achieving 44.44%, compared with 40.00% for the QSVM. This indicates that the Classical SVM is able to identify a larger proportion of the actual positive instances. Similarly, the Classical SVM achieves a higher F1-score of 51.06%, while the QSVM obtains 48.48%.

Overall, the figure demonstrates that the QSVM provides a slight advantage in precision, whereas the Classical SVM provides better recall and F1-score. Therefore, under the current experimental configuration, the Classical SVM demonstrates a better overall balance between precision and recall. However, as the models were evaluated using different numbers of test samples, these findings should be interpreted as preliminary and should be validated using an identical evaluation dataset and experimental protocol.

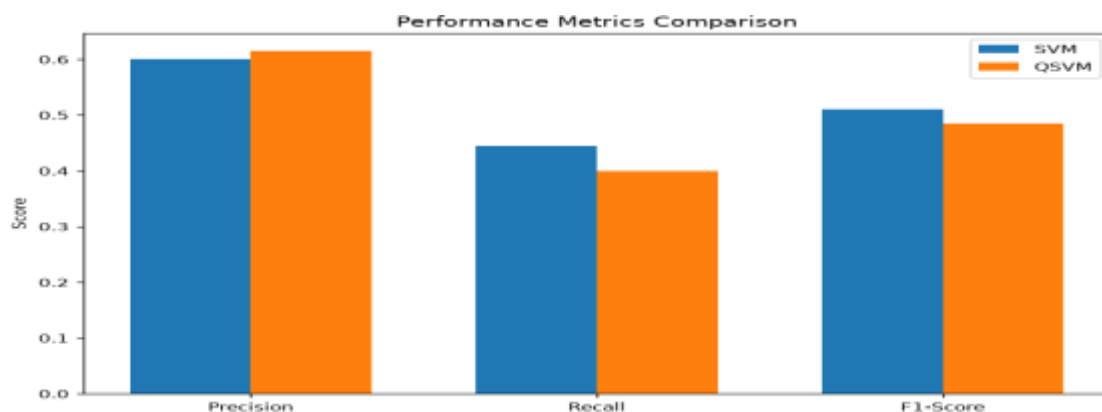


Fig. 3. Comparative Performance of Classical SVM and QSVM Based on Precision, Recall, and F1-Score

The confusion matrices of the classical Support Vector Machine (SVM) and Quantum Support Vector Machine (QSVM) provide further insight into their classification performance. The SVM confusion matrix consists of 84 true negatives (TN), 16 false positives (FP), 30 false negatives (FN), and 24 true positives (TP), resulting in 108 correctly classified samples out of 154 test samples. In comparison, the QSVM produced 15 true negatives, 5 false positives, 12 false negatives, and 8 true positives, with 23 correctly classified samples out of 40 test samples.

The Receiver Operating Characteristic (ROC) curve was employed to evaluate and compare the discriminative capability of the classical Support Vector Machine (SVM) and Quantum Support Vector Machine (QSVM). The ROC curve illustrates the relationship between the True Positive Rate (TPR) and False Positive Rate (FPR) at different classification thresholds, while the Area Under the Curve (AUC) provides an overall measure of the model's ability to distinguish between the two classes.

As shown in Figure 4, the classical SVM achieved an AUC of **0.704**, whereas the QSVM obtained an AUC of **0.595**. The higher AUC of the SVM indicates that it has a stronger ability to discriminate between the positive and negative classes across different decision thresholds. In contrast, the QSVM AUC is only moderately above the random-classifier baseline of 0.5, suggesting comparatively weaker class discrimination under the experimental conditions.

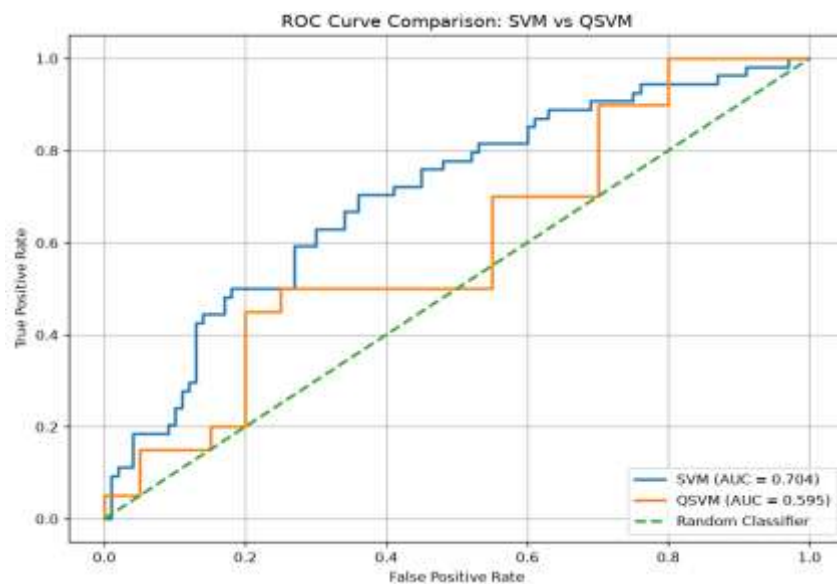


Fig. 4. ROC Curve Comparison of Classical SVM and Quantum Support Vector Machine (QSVM)

Conclusion and Future scope

This study presented a comparative analysis of Classical Support Vector Machine (SVM) and Quantum Support Vector Machine (QSVM) for diabetes prediction using Principal Component Analysis (PCA)-based feature reduction. The primary objective was to investigate the effectiveness of classical and quantum machine-learning approaches in classifying diabetic and non-diabetic cases while reducing the dimensionality of the input data.

The experimental results demonstrate that PCA can effectively transform the original feature space into a reduced representation while retaining the important information required for classification. This reduction is particularly beneficial for QSVM because a smaller feature space facilitates quantum data encoding and reduces the complexity of the quantum circuit. The comparative evaluation based on accuracy, precision, recall, F1-score, confusion matrix, and computational time provides a comprehensive assessment of both approaches. The Classical SVM provides a reliable baseline with efficient training and prediction, whereas QSVM demonstrates the feasibility of incorporating quantum computing principles into healthcare prediction. Nevertheless, the study demonstrates that quantum machine learning can be explored as a promising direction for medical classification problems. Overall, the findings establish a useful foundation for comparing classical and quantum machine-learning techniques for diabetes prediction and demonstrate the potential of PCA-based dimensionality reduction for both learning paradigms.

Future studies can execute the QSVM model on real quantum processors instead of relying primarily on classical quantum simulators. This would provide a more realistic evaluation of quantum computational performance. Different quantum feature maps and quantum kernels can be investigated to determine whether alternative quantum representations can improve classification accuracy, recall, and F1-score. Hybrid models combining classical preprocessing and machine-learning algorithms with quantum circuits can be developed. Such approaches may exploit the strengths of both computational paradigms. PCA can be compared with feature-selection and dimensionality-reduction methods such as Recursive Feature Elimination (RFE), Linear

Discriminant Analysis (LDA), mutual information, and autoencoders. Future research should focus particularly on real quantum hardware, larger clinical datasets, explainability, and hybrid quantum-classical architectures to determine whether quantum machine learning can provide a practical advantage in real-world healthcare applications.

References

1. Ullah, W., & Ali, Q. (2025). Role of artificial intelligence in healthcare settings: a systematic review. *Journal of Medical Artificial Intelligence*, 8, 24.
2. Annuzzi, G., Apicella, A., Arpaia, P., Bozzetto, L., Criscuolo, S., De Benedetto, E., ... & Vallefucio, E. (2023). Impact of nutritional factors in blood glucose prediction in type 1 diabetes through machine learning. *IEEE Access*, 11, 17104-17115.
3. Zeng, Y., Zhou, Z. Y., Rinaldi, E., Gneiting, C., & Nori, F. (2023). Approximate autonomous quantum error correction with reinforcement learning. *Physical Review Letters*, 131(5), 050601.
4. Larabi-Marie-Sainte S, Aburahmah L, Almohaini R, Saba T (2019) Current techniques for diabetes prediction: review and case study. *Appl Sci*. <https://doi.org/10.3390/app9214604>.
5. Sisodia D, Sisodia DS (2018) Prediction of diabetes using classification algorithms. *Proc Comput Sci* 132:1578–1585. <https://doi.org/10.1016/j.procs.2018.05.122>.
6. Tiwari P, Melucci M (2019) Towards a quantum-inspired binary classifier. *IEEE Access* 7:42354–42372. <https://doi.org/10.1109/ACCESS.2019.2904624>.
7. Khan TM, Robles-Kelly A (2020) Machine learning: quantum vs classical. *IEEE Access* 8:219275–219294. <https://doi.org/10.1109/ACCESS.2020.3041719>.
8. M.A. Nielsen et al. *Quantum Computation and Quantum Information* (2010)
9. J. Biamonte et al. *Quantum machine learning* *Nature* (2017)
10. M. Schuld et al. *Supervised Learning with Quantum Computers* (2021).
11. Naramparambath, H., Patra, K., & Aswiga, R. V. (2026). Quantum machine learning framework with cryptographic feature selection for enhanced diabetes prediction. *Discover Computing*, 29(1), 203.
12. Upama, P. B., Haque, M. M., Alam, K. S., & Ahamed, S. I. (2026). Early Detection of Diabetes with Quantum Machine Learning and Causal Inference Techniques. *Machine Learning with Applications*, 100900.
13. Khan, W. R., Kamran, M. A., Khan, M. U., Ibrahim, M. M., Kim, K. S., & Ali, M. U. (2025). Diabetes prediction using an optimized variational quantum classifier. *International Journal of Intelligent Systems*, 2025(1), 1351522.
14. Tudisco, A., Volpe, D., & Turvani, G. (2025, June). Quantum Machine Learning in Healthcare: Evaluating QNN and QSVM Models. In *2025 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.
15. Vimal, D. (2024, October). PIMA Diabetes Prediction Using Machine Learning and Quantum Machine Learning Techniques. In *2024 ITU Kaleidoscope: Innovation and Digital Transformation for a Sustainable World (ITU K)* (pp. 1-7). IEEE.
16. Gupta, H., Varshney, H., Sharma, T. K., Pachauri, N., & Verma, O. P. (2022). Comparative performance analysis of quantum machine learning with deep learning for diabetes prediction. *Complex & Intelligent Systems*, 8(4), 3073-3087.