



COMPARATIVE ANALYSIS OF CLASSICAL AND QUANTUM SVM MODELS ON MEDICAL DIAGNOSIS DATASETS

*Gamzepelin Aksoy¹, Zeynep Özpolat²

¹Isparta University of Applied Sciences, Faculty of Technology, Department of Computer Engineering, Isparta

²Mus Alparslan University, Faculty of Engineering-Architecture, Department of Software Engineering, Mus

(Geliş/Received: 08.06.2025, Kabul/Accepted: 19.06.2025, Yayınlanma/Published: 30.06.2025)

ABSTRACT

Quantum-assisted machine learning approaches have become a significant area of research in the healthcare domain by offering alternative solutions to classical methods, particularly when dealing with high-dimensional and complex datasets. This study presents a comparative evaluation of the classification performance of classical Support Vector Machines (SVM) and quantum-based algorithms Quantum Support Vector Machine (QSVM) and Pegasos-QSVM on healthcare data.

Experimental analyses were conducted using three distinct medical datasets related to liver disease, breast cancer, and heart failure. The results demonstrate that the QSVM model consistently achieved the highest and most stable classification accuracy. Although the Pegasos-QSVM model achieved comparable accuracy rates in certain configurations, its performance was generally more variable. Nevertheless, thanks to its lower computational cost and faster processing time, Pegasos-QSVM emerges as a promising alternative, particularly in resource-constrained environments. The findings suggest that quantum-assisted models can deliver performance levels competitive with classical approaches, particularly highlighting the effectiveness of QSVM on small- to medium-sized datasets.

Keywords: Pegasos-QSVM, QSVM, Medical Datasets, Quantum Machine Learning

TIBBİ TANI VERİ SETLERİ ÜZERİNDE KLASİK VE KUANTUM SVM MODELLERİNİN KARŞILAŞTIRMALI ANALİZİ

ÖZ

Kuantum destekli makine öğrenimi yaklaşımları, özellikle yüksek boyutlu ve karmaşık veri kümeleriyle çalışırken klasik yöntemlere alternatif çözümler sunarak sağlık alanında önemli bir araştırma konusu hâline gelmiştir. Bu çalışma, sağlık verileri üzerinde klasik Destek Vektör Makineleri (SVM) ile kuantum tabanlı algoritmalar olan Kuantum Destek Vektör Makinesi (QSVM) ve Pegasos-QSVM'nin sınıflandırma performanslarının karşılaştırmalı bir değerlendirmesini sunmaktadır.

Karaciğer hastalığı, meme kanseri ve kalp yetmezliğiyle ilgili üç farklı tıbbi veri seti kullanılarak deneysel analizler gerçekleştirilmiştir. Elde edilen sonuçlar, QSVM modelinin tutarlı bir şekilde en yüksek ve en istikrarlı sınıflandırma doğruluğunu sağladığını göstermektedir. Pegasos-QSVM modeli belirli konfigürasyonlarda benzer doğruluk oranlarına ulaşsa da, genel olarak daha değişken bir performans sergilemiştir. Bununla birlikte, daha

düşük hesaplama maliyeti ve daha hızlı işlem süresi sayesinde Pegasos-QSVM, özellikle kaynakların kısıtlı olduğu ortamlarda umut vadeden bir alternatif olarak öne çıkmaktadır. Bulgular, kuantum destekli modellerin klasik yaklaşımlarla rekabet edebilecek düzeyde performans gösterebildiğini ve özellikle QSVM'nin küçük ve orta ölçekli veri kümelerinde etkili olduğunu ortaya koymaktadır.

Anahtar kelimeler: Pegasos-QSVM, QSVM, Tıbbi veri setleri, Kuantum makine öğrenme

1. Introduction

Advances in technology have significantly improved the early detection and accurate diagnosis of diseases through the use of high-performance analytical methods. Machine Learning (ML) and Deep Learning (DL) algorithms increasingly serve as digital tools that support clinical decision-making processes [1,2]. As a result, modern technologies have become essential for the efficient collection, storage, and analysis of medical data. However, traditional ML approaches often fall short in handling high-dimensional and complex healthcare datasets, particularly in terms of classification accuracy, computational efficiency, and generalizability.

The exponential growth of the global population has led to a substantial increase in healthcare-related data, rendering conventional computational methods inadequate for managing and analyzing large-scale datasets. Although supercomputers were initially employed to overcome these challenges, the increasing complexity of data has begun to exceed their processing capacity, thereby exposing the limitations of classical computing. Consequently, quantum computing based on fundamentally different algorithmic principles has emerged as a promising paradigm for data intensive applications in healthcare analytics [3,4].

Unlike classical computers, which rely on binary logic, quantum computers are grounded in the principles of quantum physics and operate using phenomena such as superposition and entanglement concepts that transcend classical mechanics [5]. These quantum properties enable systems to process multiple computational paths simultaneously, allowing them to encode and manipulate complex probability distributions. This parallelism proves especially advantageous in analyzing high-dimensional and large-scale healthcare datasets, as it facilitates the exploration of numerous possible outcomes in a single computational cycle [6,7].

Within this framework, Quantum Machine Learning (QML), which represents the integration of quantum computing with machine learning techniques, has emerged as a promising paradigm for enhancing data processing efficiency, accelerating computation, and enabling more robust predictive modeling [8]. QML algorithms are increasingly utilized in the literature to tackle the limitations of classical methods, particularly in managing complex and high-dimensional healthcare datasets [9]. A noteworthy example is the Quantum Support Vector Machine (QSVM), an extension of classical SVMs, which has demonstrated potential in reducing computational complexity while improving classification performance in healthcare applications [10].

In recent years, quantum-based algorithms that offer promising alternatives to classical methods have gained traction in healthcare-related research. Consequently, recent studies that utilize classical SVM, QSVM, and Pegasos-QSVM techniques for medical classification tasks have been examined and reviewed.

Guido et al. [11], in their review on the application of SVM in healthcare, highlighted its strong performance on high-dimensional and imbalanced datasets. Nevertheless, they pointed out persistent limitations, particularly in interpretability and handling of missing values. Kodipalli and Devi [12], compared the Fuzzy TOPSIS and SVM methods for the joint prediction of PCOS and mental health issues, concluding that Fuzzy TOPSIS yielded a higher accuracy of 98.20%. They also noted the superiority of fuzzy approaches in processing linguistic data. Kareem et al. [13], employed the IQ-OTH/NCCD CT lung cancer dataset and achieved 89.88% accuracy through an image processing and SVM-based approach. The best results were obtained by combining a polynomial kernel SVM with Gabor filters and Gray-Level Co-occurrence Matrix (GLCM) features. Öztekin et al. [14], compared six machine learning algorithms, including SVM, for breast cancer diagnosis using the Breast-XD dataset. The highest accuracy (94.34%) was obtained using the explainable XGBoost model (X²GAI). Finally, Ramu et al. [15], proposed a hybrid model integrating SVM and convolutional neural networks (CNN)

for early prediction of chronic kidney disease. This model surpassed the performance of classical SVM (94.8%) and Random Forest, reaching an accuracy of 96.8%.

Tudisco et al. [16], compared quantum models, including QSVM and QNN, with classical algorithms such as SVM on datasets related to heart failure, diabetes, and prostate cancer. Their results indicated that QSVM outperformed classical models, especially in imbalanced datasets, by achieving higher recall rates. Similarly, Khushal and Fatima [17], integrated quantum models such as QSVM and Q-KNN with fuzzy logic (FL) to improve both computational efficiency and classification accuracy. Their proposed Fuzzy Quantum Machine Learning (FQML) approach achieved superior results compared to standard QML techniques on chronic disease datasets. Maheshwari et al. [18], employed optimized QSVM (OQSVM) and hybrid quantum multilayer perceptron (HQMLP) models for the classification of cardiovascular diseases. They highlighted that the use of PauliFeatureMap and fine-tuned hyperparameters led to enhanced accuracy in their models. In the field of breast cancer diagnosis, Jose P. et al. [19], demonstrated that the QSVM model optimized using the Elitist Non-Dominated Sorting Genetic Algorithm (ENSGA) significantly outperformed traditional SVM variants, underscoring the algorithm's potential in medical classification tasks.

Chatterjee ve Das [20], proposed a Staged Pegasos Quantum Support Vector Classifier for breast cancer diagnosis, which stratifies patients based on the severity of their condition. The model's performance was further enhanced through the incorporation of fuzzy logic and expert knowledge-based weighted algorithms. Nasir et al. [21], conducted a comparative study of classical and quantum machine learning algorithms for the classification of dermatological diseases. While classical models such as Bagging and Decision Trees achieved the highest accuracy of 98%, among quantum-based approaches, the Pegasos-QSVM model stood out with an accuracy of 84%, surpassing classical SVM in terms of precision. Singh and Pokhrel [22], assessed several quantum machine learning algorithms using various feature mapping techniques for the binary classification of genomic sequence data. Their analysis revealed that Pegasos-QSVM yielded particularly high recall rates. Munshi et al. [23], compared the performance of Pegasos-QSVM and Variational Quantum Classifier (VQC) on a lung cancer dataset and reported that Pegasos-QSVM achieved a superior accuracy of 85%, outperforming VQC.

As a result of the literature review, it is evident that studies based on Quantum Machine Learning (QML) are frequently employed in the healthcare domain, at least as commonly as classical algorithms. However, research specifically focused on the Pegasos-QSVM algorithm remains relatively scarce. This study aims to fill the existing gap in the literature by comprehensively evaluating the performance of both QSVM and Pegasos-QSVM models in real-world healthcare classification tasks, providing a comparative analysis against classical SVM models.

As evidenced by the literature review, Quantum Machine Learning (QML) techniques are increasingly employed in healthcare applications, offering performance levels that often rival those of classical machine learning algorithms. However, studies focusing specifically on the Pegasos-QSVM model remain limited. This study seeks to address this gap by comprehensively evaluating the performance of QSVM and Pegasos-QSVM models in real-world healthcare classification tasks and providing a comparative analysis against classical SVM methods. Through this investigation, the potential advantages of quantum-based approaches in processing complex medical data are highlighted.

The structure of the remainder of this paper is organized as follows: Section 2 introduces the datasets utilized in the study, details the preprocessing procedures, and outlines the classification algorithms applied. Section 3 presents the experimental results obtained from the implementation of the algorithms. Section 4 discusses the study's contributions to the existing literature, provides an evaluation of the findings, and outlines directions for future research.

2. Material and Method

In this study, three publicly available medical datasets were utilized to evaluate and compare the performances of classical and quantum-based classification algorithms. The datasets employed were the Breast Cancer Coimbra Dataset, Heart Failure Clinical Records Dataset, and the Indian Liver Patient Dataset. These datasets were selected due to their binary classification structure, which aligns with the requirements of the Pegasos-QSVM algorithm.

Three distinct classification methods were implemented: classical Support Vector Machines (SVM), Quantum Support Vector Machines (QSVM), and Pegasos-QSVM. Prior to classification, multiple preprocessing steps were conducted to ensure data consistency and enhance model performance. Initially, all features were standardized using the StandardScaler, transforming each feature to have zero mean and unit variance. Subsequently, Principal Component Analysis (PCA) was applied for dimensionality reduction, reducing the number of features to between 3 and 9 components to meet the input dimensionality constraints of quantum circuits. Finally, the PCA-transformed data were rescaled using MinMaxScaler to the range $[0, \pi]$, which is commonly preferred for quantum state encoding due to the periodic nature of quantum gates. This rescaling step enabled the conversion of classical data into a qubit-compatible format, facilitating its integration into quantum circuits.

For quantum classification, the PCA-reduced datasets were mapped from classical bit format to quantum qubit format using four distinct quantum feature maps: X, Y, Z, and ZZ. These feature maps function as encoding mechanisms that embed classical data into high-dimensional Hilbert spaces via parameterized quantum circuits. To evaluate the effect of circuit depth on model performance, each feature map was tested under two repetition settings: $\text{reps} = 1$ and $\text{reps} = 2$. This allowed for a comparative analysis of how entanglement complexity influences the expressivity of the quantum kernels.

The quantum models were trained using the QSVM and Pegasos-QSVM algorithms within a simulated quantum environment. All simulations were carried out using the Qiskit framework with the statevector_simulator backend. In the QSVM implementation, the FidelityQuantumKernel class was employed to compute the quantum kernel matrix by estimating the similarity (fidelity) between quantum states encoded via selected feature maps [24]. Model performance was evaluated using 5-fold stratified cross-validation to ensure balanced representation of class distributions. For the Pegasos-QSVM model, the regularization parameter was set to $C = 1000$, and the number of optimization steps was set to $\tau = 100$, based on empirical tuning. Across all models and experiments, two key evaluation metrics were considered: classification accuracy and execution time, both averaged over the five folds.

A schematic representation of the complete methodology is presented in Figure 1, illustrating the entire workflow from data preprocessing to classification using both classical and quantum SVM algorithms.

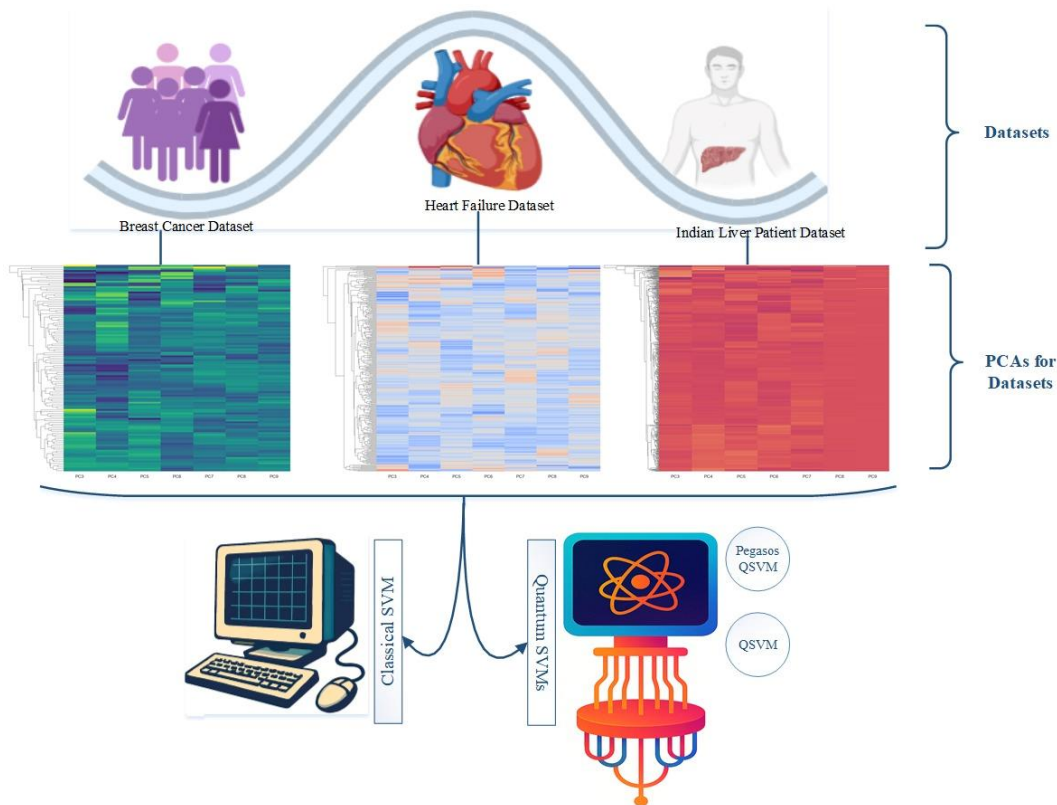


Figure 1. Classification workflow using classical and quantum SVM algorithms

2.1.Dataset

In this study, three distinct clinical datasets were utilized to develop and evaluate predictive models for different medical conditions. All datasets were obtained from the UCI Machine Learning Repository, a widely used open-access platform for benchmarking machine learning algorithms [25–27]. Each dataset includes structured, domain-specific features that enable classification tasks related to liver disease, breast cancer, and heart failure, respectively.

2.1.1.Indian liver patient dataset

A publicly available dataset containing the biochemical profiles of 584 individuals from the northeastern region of Andhra Pradesh, India, is employed in this study. The dataset comprises demographic variables such as age and gender, along with clinically significant biochemical indicators including total and direct bilirubin, alkaline phosphatase (ALP), alanine aminotransferase (ALT), aspartate aminotransferase (AST), total protein, albumin, and the albumin-to-globulin ratio. Of the participants, 441 are male and 142 are female. These features are widely recognized as metabolic biomarkers commonly used in the diagnosis of liver dysfunction. The dataset was constructed with the primary objective of supporting the development of predictive models capable of determining whether an individual is affected by liver disease based on these biomarkers [28].

2.1.2. Breast cancer coimbra dataset

An open-access dataset comprising 116 clinical records exclusively from female participants in the Coimbra region of Portugal is employed for breast cancer prediction. The dataset includes nine attributes, encompassing both basic anthropometric measurements such as age and body mass index (BMI) and metabolic or inflammatory biomarkers, including glucose, insulin, HOMA (Homeostatic Model Assessment), leptin, adiponectin, resistin, and MCP-1 (Monocyte Chemoattractant Protein-1). These variables are considered potential indicators associated with the biological mechanisms of breast cancer. Each instance is labeled based on the subject's diagnosis status, making the dataset suitable for supervised classification tasks. The dataset was curated to support the development of machine learning-based early detection systems for breast cancer [29].

2.1.3. Heart failure clinical records dataset

Collected from 299 individuals diagnosed with heart failure, this clinical dataset provides detailed information on demographic, physiological, and laboratory parameters relevant to cardiovascular risk assessment. It includes 12 attributes such as age, sex, ejection fraction, serum creatinine, serum sodium, platelet count, creatinine phosphokinase levels, and the presence of conditions like anemia, diabetes, high blood pressure, and smoking habits. The dataset also contains a follow-up duration feature (in days) and a binary outcome indicating whether a death event occurred during the observation period. These variables are widely recognized as important predictors in the progression and prognosis of heart failure. The dataset serves as a resource for developing predictive models aimed at identifying high-risk patients and improving early intervention strategies in clinical practice [30].

2.2.Data preparation

In this study, the preprocessing pipeline consisted of three main stages. First, all features were standardized using the StandardScaler to ensure equal contribution by transforming them to have zero mean and unit variance. Next, dimensionality reduction was performed via Principal Component Analysis (PCA), aligning the number of features with the number of qubits used in the quantum classification models. Lastly, the reduced feature set was rescaled into the range $[0, \pi]$ using MinMaxScaler, making it compatible with quantum state encoding. This sequence of transformations prepared the data appropriately for both classical and quantum machine learning algorithms.

Principal Component Analysis is a widely adopted statistical technique for reducing the dimensionality of datasets, especially those with a high number of features. It aims to preserve the most relevant information by generating new components that capture the majority of the variance in the original data, with minimal information loss. These components are constructed to be mutually uncorrelated, allowing complex datasets to be restructured into more compact and interpretable forms [31].

By applying orthogonal transformations, PCA converts correlated variables into independent ones, which helps suppress noise and reveal meaningful relationships between features [32]. This simplified structure improves the performance of various learning models, including those based on quantum computation. Thanks to its effectiveness and interpretability, PCA remains one of the most widely preferred methods for dimensionality reduction in the literature [33].

2.3.Methods

2.3.1.Support vector machine

Support vector machines are robust and theoretically grounded supervised learning methods widely used for solving classification problems. The fundamental principle of SVM is to separate data instances belonging to different classes by constructing an optimal hyperplane within the feature space. The algorithm aims to maximize the margin between the classes while simultaneously minimizing classification errors [34].

After training the model on labeled data, the resulting decision function is applied to new instances to perform classification. The data points closest to the decision boundary are referred to as “support vectors” and they play a crucial role in determining the performance of the classifier. In linearly separable datasets, SVM constructs a hyperplane that is equidistant from both classes. For datasets that are not linearly separable, kernel functions are employed to transform the data into a higher-dimensional space where a linear separation becomes feasible. Commonly used kernels include linear, polynomial, sigmoid, and Radial Basis Function (RBF) kernels. The hyperplane is positioned perpendicular to the shortest distance between data points from each class, thereby enhancing the separability and improving the model's generalization capability. SVM is a powerful method rooted in statistical learning theory, capable of producing effective results in both binary and multiclass classification tasks [35,36].

2.3.2.Quantum support vector machine

Quantum support vector machines represent a quantum-enhanced adaptation of the classical Support Vector Machine (SVM) algorithm. While classical SVMs aim to find an optimal hyperplane that separates different classes by maximizing the margin between them, QSVM utilizes quantum state space as the feature domain. This allows for more expressive data representation and the potential to solve complex classification problems more efficiently.

The QSVM algorithm integrates Grover's search algorithm as a quantum subroutine to optimize non-convex cost functions, enabling it to reach global optima [37]. Classical data are encoded into quantum states, and kernel functions are employed to transform the data into a higher-dimensional space where separation becomes feasible. Using quantum superposition and parallelism, QSVM evaluates multiple states simultaneously, accelerating the training process.

In addition to improving classification accuracy, QSVM reduces time complexity from linear to sublinear, making it particularly promising for large-scale learning problems. However, its performance is highly dependent on quantum hardware quality, as quantum noise may impact reliability. Despite this, QSVM stands out for its potential to enhance machine learning efficiency by combining statistical learning theory with quantum computing principles [38].

2.3.3.Pegasos quantum support vector machine

Pegasos-QSVM (Primal Estimated sub-GrAdient SOLver for SVM-Classification) is an efficient algorithm developed to address nonlinear binary classification problems using SVM. Unlike traditional SVM solvers that rely on dual formulation, Pegasos-QSVM directly operates on the primal objective and employs stochastic sub-gradient descent to approximate a solution. Its optimization strategy aims to minimize hinge loss while applying regularization, ensuring both accuracy and generalizability against overfitting [39].

In quantum implementations, constructing accurate kernel matrices often requires a large number of quantum circuit executions, significantly increasing computational costs. Pegasos-QSVM is well-suited for such quantum settings, as it preserves the use of the kernel trick: instead of explicitly mapping data into high-dimensional feature space, it relies on computing inner products. This approach allows for the

construction of the separating hyperplane through kernel evaluations, and the decision function is expressed as a weighted sum of kernel values associated with the support vectors. Additionally, the bias term is implicitly handled by embedding a constant into the kernel function, enhancing model flexibility. The quantum Pegasos algorithm requires fewer kernel evaluations than dual QSVM, offering a significant advantage in settings with large datasets and limited quantum resources [40].

Furthermore, in QSVMs, classical data is embedded into the quantum feature space using parameterized quantum circuits, resulting in quantum kernel functions derived from Hilbert–Schmidt inner products. These quantum kernels are only approximately computable due to the probabilistic nature of quantum measurements governed by Born’s rule. This inherent statistical noise in quantum settings makes the Pegasos approach even more appealing, as it avoids constructing the full kernel matrix and yields a training complexity that is independent of dataset size. Therefore, Pegasos-QSVM is expected to outperform standard QSVMs in terms of scalability and efficiency when applied to large training sets under quantum constraints [10].

2.3.4. FeatureMap

Quantum feature maps are fundamental components that allow classical data to be encoded into quantum circuits in a form suitable for quantum processing. These maps project classical input vectors into quantum states by embedding them in a high-dimensional Hilbert space [41]. This process involves building quantum circuits composed of specific quantum gates to represent the data as qubits, making it usable by quantum machine learning algorithms.

The choice of feature map significantly influences the expressivity and entanglement structure of the circuit. Key factors include the circuit depth, encoding method, gate types, and the entanglement pattern. Commonly used feature maps include Z, ZZ, and Pauli-based variants, each with unique characteristics. The Z FeatureMap applies Hadamard gates followed by single-qubit unitary operations, encoding data without creating entanglement, thus resulting in a simpler and faster circuit design. In contrast, the ZZ feature map introduces entanglement through Controlled-Z gates, enhancing representational capacity at the cost of increased computational complexity [39]. The PauliFeatureMap offers even greater flexibility by combining Pauli X, Pauli Y, and Z operations, enabling the circuit to adapt to the structure of the input data and the specific classification task [42].

In this study, Pauli X, Pauli Y, Z, and ZZ feature maps were employed to encode the PCA-reduced classical data into quantum form. These maps were selected to assess the impact of different encoding strategies on the classification performance of QSVM and Pegasos-QSVM. Each map introduces different structural and computational complexities, allowing for comparative analysis. For all feature maps, experiments were conducted with both $reps=1$ and $reps=2$ to evaluate the influence of circuit depth. Furthermore, the entanglement parameter was set to “full” in the Pauli X, Pauli Y, and ZZ feature maps to investigate how various entanglement configurations affect the models’ performance. The circuit structures of the Z, ZZ, Pauli X, and Pauli Y feature maps used in this study are illustrated in Figure 2. These visualizations provide a clear comparison of the different entanglement schemes and circuit complexities introduced by each mapping strategy.

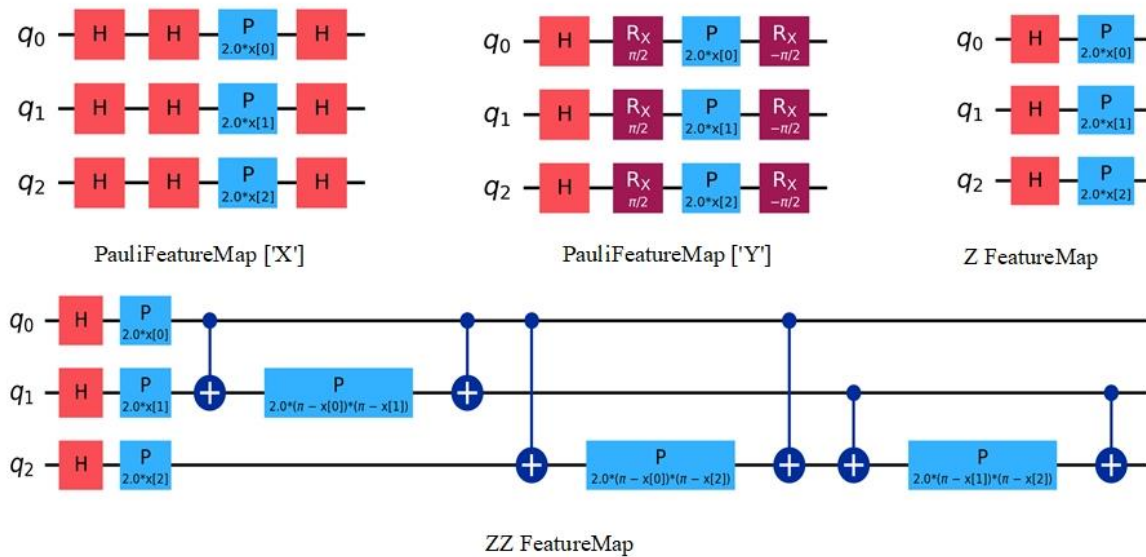


Figure 2. Quantum circuits of the Z, ZZ, Pauli X, and Pauli Y feature maps constructed with 3 qubits and reps = 1

3. Research Findings

In this section, the results of analyses conducted on three different healthcare datasets are presented. The performance of classical SVM and quantum-based models QSVM and Pegasos-QSVM was evaluated based on classification accuracy. To ensure the reliability of the models on both training and testing sets, a 5-fold cross-validation strategy was employed. For the sake of reproducibility, the random seed was fixed at 42 during data splitting.

To enable a comprehensive comparison of quantum algorithms, the experiments were carried out using seven different qubit sizes and four distinct quantum feature maps. Prior to classification, all datasets underwent a series of preprocessing steps. PCA was applied for dimensionality reduction, resulting in datasets with varying numbers of features suitable for quantum circuit input constraints. All algorithms were developed using the Python programming language and implemented with the Qiskit library. Both classical and quantum machine learning models were executed in the Google Colab cloud environment. For the execution of quantum algorithms, the statevector simulator environment was utilized.

The findings for each dataset are presented in the following subsections, starting with the Indian Liver Patient Dataset. The results obtained using the Pegasos-QSVM and QSVM algorithms on the “Indian Liver Patient Dataset,” based on different qubit values and feature maps with circuit repetition counts set to 1 and 2, are presented in Table 1.

Table 1. Accuracy analysis for the ILPD using Pegasos-QSVM and QSVM algorithms

Qubits	Pegasos-QSVM				QSVM			
	Accuracy (%) Reprs=1							
	X	Y	Z	ZZ	X	Y	Z	ZZ
3	62.98	66.91	66.91	65.53	71.36	71.36	71.36	71.36
4	62.98	64.51	64.51	62.12	71.36	71.36	71.36	70.33
5	71.36	70.67	70.67	63.29	71.36	71.36	71.36	71.18
6	71.36	66.03	66.03	62.74	71.36	70.84	70.84	68.61
7	62.98	65.19	65.19	66.40	71.36	71.36	71.36	71.18
8	62.74	55.09	55.09	57.09	71.36	71.53	71.53	71.01
9	62.74	62.75	62.75	66.87	71.36	71.70	71.70	71.01

Accuracy (%) Reps=2								
Qubits	X	Y	Z	ZZ	X	Y	Z	ZZ
3	66.91	66.74	69.30	66.39	71.36	71.36	71.36	71.36
4	64.51	65.54	65.87	68.79	71.36	71.36	71.36	71.01
5	70.67	71.36	68.62	71.36	71.36	71.36	71.18	72.38
6	66.03	61.53	64.23	57.56	71.36	71.36	70.67	70.67
7	65.19	62.98	66.56	64.66	71.36	71.36	71.70	71.53
8	55.09	61.90	60.06	62.61	71.53	71.36	71.70	71.70
9	62.75	62.56	67.22	50.09	71.70	71.36	71.70	71.70

The QSVM algorithm demonstrated consistently high and stable performance across nearly all qubit and feature map combinations, achieving an accuracy rate of approximately 71%. The highest recorded accuracy was 72.38%, obtained with 5 qubits, the ZZ feature map, and Reps=2. These results indicate that the QSVM algorithm is relatively insensitive to variations in feature map types or qubit numbers and provides robust and consistent outcomes, especially when circuit repetitions are increased. In contrast, the performance of the Pegasus-QSVM algorithm appeared to be more variable across different configurations. Although it achieved accuracy levels comparable to QSVM in certain scenarios such as 5 and 6 qubits with Reps=1 (both reaching 71.36%) its overall sensitivity to the choice of feature map and qubit count was notably higher. Particularly with the Pauli X and Pauli Y feature maps and at 8 or 9 qubits, accuracy values dropped below 60%. These findings suggest that while Pegasus-QSVM can deliver results similar to QSVM under specific settings, QSVM generally yields more stable and reliable performance across a broader range of parameters.

Table 2 presents the accuracy results obtained from the Breast Cancer Coimbra dataset using Pegasus-QSVM and QSVM algorithms across varying qubit numbers (from 3 to 9), different feature maps (Pauli X, Pauli Y, Z, ZZ), and circuit repetition counts (Reps=1 and Reps=2).

Table 2. Accuracy analysis for the Breast Cancer dataset using Pegasus-QSVM and QSVM algorithms

Pegasus-QSVM					QSVM			
Accuracy (%) Reps=1								
Qubits	X	Y	Z	ZZ	X	Y	Z	ZZ
3	52.64	53.41	54.42	51.78	55.17	61.21	61.21	65.52
4	51.70	65.54	66.41	48.95	55.17	67.24	67.24	50.00
5	52.57	67.32	67.32	59.42	55.17	72.41	72.41	56.03
6	53.51	65.51	65.51	53.44	55.17	72.41	74.14	59.48
7	54.28	65.51	65.51	54.28	55.17	75.86	75.86	59.48
8	55.18	62.10	63.01	61.27	55.17	73.28	73.28	58.62
9	59.42	61.20	66.38	59.42	55.17	75.00	75.00	57.76
Accuracy (%) Reps=2								
Qubits	X	Y	Z	ZZ	X	Y	Z	ZZ
3	53.41	55.11	55.98	51.67	61.21	55.17	54.31	54.31
4	65.54	49.96	58.62	49.09	67.24	55.17	67.24	51.72
5	67.32	52.72	60.22	52.64	72.41	55.17	70.69	56.90
6	65.51	53.55	65.58	55.22	74.14	55.17	68.97	55.17
7	65.51	50.04	58.51	49.13	75.86	55.17	70.69	56.90
8	62.10	57.57	63.77	56.88	73.28	55.17	73.28	55.17
9	61.20	51.70	64.57	58.51	75.00	55.17	74.14	55.17

The results show that the QSVM algorithm consistently outperformed Pegasus-QSVM across nearly all configurations, especially in settings with 5 to 7 qubits and when using the Z feature map. The highest accuracy value for QSVM was 75.86%, achieved with 7 qubits, the Pauli X or Y feature map, and Reps=1. Furthermore, even with low repetition counts, QSVM maintained relatively stable and high

accuracy, generally remaining above 70% in its best configurations. In contrast, the Pegasos-QSVM algorithm exhibited more fluctuating and less predictable performance. While it reached competitive accuracy values in certain configurations (e.g., 67.32% with 5 qubits and the Pauli Y feature map, Reprs=1), its performance tended to vary significantly across different qubit and feature map combinations. For instance, accuracy dropped below 55% in multiple cases, especially for ZZ feature maps and low qubit counts. Table 3 presents the accuracy rates obtained using the Pegasos-QSVM and QSVM algorithms on the Heart Failure dataset.

Table 3. Accuracy analysis for the Heart Failure dataset using Pegasos-QSVM and QSVM algorithms

Pegasos-QSVM					QSVM			
Accuracy (%) Reprs=1								
Qubits	X	Y	Z	ZZ	X	Y	Z	ZZ
3	67.89	63.58	67.59	52.52	67.89	74.58	74.58	69.23
4	39.44	61.59	62.58	53.53	67.89	77.26	77.26	68.90
5	60.77	62.89	57.54	60.56	67.89	74.92	74.92	67.89
6	60.56	67.89	68.56	58.86	67.89	76.92	76.92	69.90
7	60.56	70.57	69.55	65.23	67.89	76.59	76.59	67.56
8	67.89	69.59	62.28	58.54	67.89	74.92	74.92	67.89
9	67.89	61.56	65.56	59.77	67.89	75.59	75.59	67.89
Accuracy (%) Reprs=2								
Qubits	X	Y	Z	ZZ	X	Y	Z	ZZ
3	65.59	62.56	68.89	54.20	74.58	67.89	73.91	68.23
4	63.23	67.89	58.47	52.54	77.26	67.89	75.25	66.22
5	66.56	67.89	65.56	51.77	74.92	67.89	73.91	67.22
6	64.46	60.44	65.21	52.55	76.92	67.89	75.59	67.56
7	63.59	59.56	64.24	57.56	76.59	67.89	73.24	67.89
8	55.11	65.89	65.88	55.00	74.92	67.89	70.90	67.89
9	61.23	58.45	60.55	59.53	75.59	67.89	68.23	68.23

The QSVM algorithm consistently delivered higher accuracy rates compared to Pegasos-QSVM across both repetition scenarios. For Reprs=1, the highest accuracy achieved by QSVM was 77.26%, obtained using 4 qubits with either the Pauli X or Y feature map. In the Reprs=2 scenario, QSVM maintained similarly strong performance, particularly with qubit counts between 5 and 7 and the Z feature map, achieving accuracy rates ranging from 73% to 75%. In contrast, the accuracy rates of the Pegasos-QSVM algorithm exhibited a wider variability. For instance, a notably low accuracy of 39.44% was recorded with 4 qubits and the Pauli X feature map, whereas higher accuracies up to 67.89% were achieved in certain scenarios with 8 and 9 qubits. This indicates that Pegasos-QSVM is more sensitive to configuration choices and struggles to maintain consistent performance. In conclusion, the QSVM algorithm demonstrated more stable and higher-accuracy classifications on this dataset, particularly when using fewer repetitions (Reprs=1) and a moderate number of qubits (4 to 7).

Table 4 presents the classification accuracy results of the classical SVM algorithm applied to the Indian Liver Patient, Breast Cancer Coimbra, and Heart Failure datasets. The performances are reported based on varying numbers of features, following dimensionality reduction through PCA. This allows for a direct comparison of how the number of features impacts the accuracy of classical SVM across different medical datasets.

Table 4. Accuracy results of SVM across three datasets with varying feature counts

Features	Indian Liver Patient	Breast Cancer Coimbra	Heart Failure
	SVM Accuracy (%)		
3	71.36	62.93	71.58
4	71.18	74.06	71.25

5	71.18	72.39	72.24
6	70.84	71.56	72.56
7	70.84	75.00	73.57
8	71.19	73.26	74.58
9	71.19	72.39	74.58

When comparing the results in Table 4 with those presented in the previous tables for QSVM and Pegasos-QSVM models, several noteworthy observations emerge. Firstly, the QSVM algorithm demonstrates consistent and high performance across all datasets and qubit configurations. Particularly in the Indian Liver Patient and Heart Failure datasets, QSVM frequently achieves equal or higher accuracy values compared to classical SVM. This suggests that QSVM is capable of operating effectively on datasets with limited feature dimensions. On the other hand, the performance of the Pegasos-QSVM algorithm exhibits greater variability. Significant changes in accuracy are observed depending on the feature map and qubit configuration. Although in certain scenarios (e.g., the Breast Cancer dataset with 5 qubits and Reprs=2), Pegasos-QSVM achieves accuracy values comparable to or even surpassing QSVM, its overall performance appears to be less consistent. This indicates that Pegasos-QSVM is more sensitive to hyperparameter selection and requires careful tuning to achieve optimal results.

In addition to classification accuracy, the training and prediction times of the models were also evaluated to analyze their practical applicability. All algorithms were executed on the same hardware. Time measurements were conducted using the time module, covering the duration from the start of training to the completion of prediction. For quantum algorithms, the statevector_simulator was used, allowing for the assessment of processing times under ideal conditions.

When considering the execution times of the applied models, it was observed that the classical SVM algorithm produced results in the shortest amount of time. This is attributed to the lower computational complexity of classical machine learning methods. On the other hand, while the QSVM algorithm often achieved higher accuracy, it exhibited the longest execution time due to the additional computational burden of quantum kernel evaluations. The Pegasos-QSVM algorithm was positioned between these two in terms of runtime. Unlike QSVM, this method does not compute the full quantum kernel matrix; instead, it performs calculations over selected subsets, significantly reducing the overall execution time. This finding indicates that although quantum-based methods have notable potential, there remains considerable room for improvement in terms of computational efficiency. Furthermore, with respect to the feature maps used, it was observed that the Pauli X, Pauli Y, and Z maps required similar execution times, whereas the ZZ feature map resulted in longer processing durations due to its more complex structure. For instance, when the number of circuit repetitions (reps) was set to 1 in the Pegasos-QSVM algorithm, it was observed that the Pauli X, Pauli Y, Z, and ZZ feature maps completed their analyses in 775, 807, 614, and 1933 seconds, respectively. Similarly, in the QSVM algorithm under the same conditions, the execution times for the Pauli X, Pauli Y, Z, and ZZ feature maps were recorded as 36290, 39853, 38588, and 73954 seconds, respectively.

4. Results and Discussion

In this study, classical and quantum support vector machines (SVM, QSVM, Pegasos-QSVM) were compared using the Indian Liver Patient, Breast Cancer Coimbra, and Heart Failure datasets. The evaluations were based on classification accuracy, considering various feature sizes (3–9), different quantum feature maps (Pauli X, Pauli Y, Z, ZZ), and circuit repetition counts (Reprs=1 and Reprs=2).

The results demonstrate that the QSVM algorithm generally provides more consistent and higher accuracy rates compared to classical SVM and Pegasos-QSVM. Especially in the Indian Liver Patient and Heart Failure datasets, QSVM often achieved accuracy values equal to or higher than those of classical SVM. This indicates that QSVM performs effectively on datasets with limited feature dimensions. On the other hand, the performance of the Pegasos-QSVM algorithm showed more variability and appeared to be more sensitive to the choice of feature map and qubit configuration. Although high accuracy values were obtained in some combinations (e.g., 5 qubits with Reprs=2 in the Breast Cancer dataset), the overall performance was less consistent compared to QSVM. This suggests that Pegasos-QSVM requires more careful tuning of hyperparameters to achieve optimal results. The

classical SVM model, meanwhile, remained competitive particularly with a higher number of features. However, QSVM, benefiting from the quantum kernel advantage, generally outperformed classical SVM across different settings.

In conclusion, the QSVM algorithm stands out as an effective quantum machine learning approach for small to mid-sized datasets, while the Pegasos-QSVM algorithm may require further optimization to fully realize its potential. Additionally, beyond classification accuracy, the runtime analysis revealed significant insights into the practical applicability of the models. While QSVM consistently delivered higher accuracy, it incurred the longest training and prediction times due to the computational burden of full quantum kernel evaluations. In contrast, Pegasos-QSVM demonstrated a notable advantage in terms of speed, requiring substantially less computation time. This efficiency stems from its ability to avoid full kernel matrix construction by selectively computing only necessary kernel evaluations. Therefore, Pegasos-QSVM may be considered a more practical choice in scenarios where computational resources or execution time are limited, despite its relatively lower and less stable accuracy.

In future work, hyperparameter optimization techniques (such as grid search or Bayesian optimization) can be employed to enhance the classification accuracy of the Pegasos-QSVM algorithm. Moreover, experimenting with combinations of different quantum feature maps or implementing hybrid classical-quantum kernel strategies may help improve the balance between accuracy and stability. These advancements could make Pegasos-QSVM a more competitive alternative in terms of both speed and predictive performance.

5. References

- [1] T.B. Alakus, M. Baykara, Comparison of Monkeypox and Wart DNA Sequences with Deep Learning Model, *Applied Sciences* 12 (2022) 10216. <https://doi.org/10.3390/app122010216>.
- [2] Ö. Yildirim, A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification, *Computers in Biology and Medicine* 96 (2018) 189–202. <https://doi.org/10.1016/j.combiomed.2018.03.016>.
- [3] N. Jeyaraman, M. Jeyaraman, S. Yadav, S. Ramasubramanian, S. Balaji, Revolutionizing Healthcare: The Emerging Role of Quantum Computing in Enhancing Medical Technology and Treatment, *Cureus* (2024). <https://doi.org/10.7759/cureus.67486>.
- [4] R. Ur Rasool, H.F. Ahmad, W. Rafique, A. Qayyum, J. Qadir, Z. Anwar, Quantum Computing for Healthcare: A Review, *Future Internet* 15 (2023) 94. <https://doi.org/10.3390/fi15030094>.
- [5] Z. Li, Analysis of the Principles of Quantum Computing and State-of-the-Art Applications, *Theoretical and Natural Science* 41 (2024) 65–71. <https://doi.org/10.54254/2753-8818/41/2024CH0155>.
- [6] D. Dhinakaran, L. Srinivasan, S.M. Udhaya Sankar, D. Selvaraj, Quantum-based privacy-preserving techniques for secure and trustworthy internet of medical things an extensive analysis, *QIC* 24 (2024) 227–266. <https://doi.org/10.26421/QIC24.3-4-3>.
- [7] A.M. Dalzell, S. McArdle, M. Berta, P. Bienias, C.-F. Chen, A. Gilyén, C.T. Hann, M.J. Kastoryano, E.T. Khabiboulline, A. Kubica, G. Salton, S. Wang, F.G.S.L. Brandão, Quantum algorithms: A survey of applications and end-to-end complexities, (2023). <https://doi.org/10.48550/arXiv.2310.03011>.
- [8] T.M. Khan, A. Robles-Kelly, Machine Learning: Quantum vs Classical, *IEEE Access* 8 (2020) 219275–219294. <https://doi.org/10.1109/ACCESS.2020.3041719>.
- [9] P. Lamichhane, D.B. Rawat, Quantum Machine Learning: Recent Advances, Challenges, and Perspectives, *IEEE Access* 13 (2025) 94057–94105. <https://doi.org/10.1109/ACCESS.2025.3573244>.
- [10] V. Havlíček, A.D. Córcoles, K. Temme, A.W. Harrow, A. Kandala, J.M. Chow, J.M. Gambetta, Supervised learning with quantum-enhanced feature spaces, *Nature* 567 (2019) 209–212. <https://doi.org/10.1038/s41586-019-0980-2>.

- [11] R. Guido, S. Ferrisi, D. Lofaro, D. Conforti, An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review, *Information* 15 (2024) 235. <https://doi.org/10.3390/info15040235>.
- [12] A. Kodipalli, S. Devi, Prediction of PCOS and Mental Health Using Fuzzy Inference and SVM, *Front. Public Health* 9 (2021). <https://doi.org/10.3389/fpubh.2021.789569>.
- [13] H.F. Kareem, M.S. AL-Huseiny, F.Y. Mohsen, E.A. Khalil, Z.S. Hassan, Evaluation of SVM performance in the detection of lung cancer in marked CT scan dataset, *Indonesian Journal of Electrical Engineering and Computer Science* 21 (2021) 1731–1738. <https://doi.org/10.11591/ijeecs.v21.i3.pp1731-1738>.
- [14] P.S. Oztekin, O. Katar, T. Omma, S. Erel, O. Tokur, D. Avci, M. Aydogan, O. Yildirim, E. Avci, U.R. Acharya, Comparison of Explainable Artificial Intelligence Model and Radiologist Review Performances to Detect Breast Cancer in 752 Patients, *Journal of Ultrasound in Medicine* 43 (2024) 2051–2068. <https://doi.org/10.1002/jum.16535>.
- [15] K. Ramu, S. Patthi, Y.N. Prajapati, J.V.N. Ramesh, S. Banerjee, K.B.V.B. Rao, S.I. Alzahrani, R. ayyasamy, Hybrid CNN-SVM model for enhanced early detection of Chronic kidney disease, *Biomedical Signal Processing and Control* 100 (2025) 107084. <https://doi.org/10.1016/j.bspc.2024.107084>.
- [16] A. Tudisco, D. Volpe, G. Turvani, Quantum Machine Learning in Healthcare: Evaluating QNN and QSVM Models, (2025). <https://doi.org/10.48550/arXiv.2505.20804>.
- [17] R. Khushal, D.U. Fatima, Fuzzy quantum machine learning (FQML) logic for optimized disease prediction, *Computers in Biology and Medicine* 192 (2025) 110315. <https://doi.org/10.1016/j.combiomed.2025.110315>.
- [18] D. Maheshwari, U. Ullah, P.A.O. Marulanda, A.G.-O. Jurado, I.D. Gonzalez, J.M.O. Merodio, B. Garcia-Zapirain, Quantum Machine Learning Applied to Electronic Healthcare Records for Ischemic Heart Disease Classification, *Human-Centric Computing and Information Sciences* 13 (2023) 1–15. <https://doi.org/10.22967/HCIS.2023.13.006>.
- [19] J. P, S. Hariharan, V. Madhivanan, S. N, M. Krisnamoorthy, A.K. Cherukuri, Enhanced QSVM with elitist non-dominated sorting genetic optimisation algorithm for breast cancer diagnosis, *IET Quantum Communication* 5 (2024) 384–398. <https://doi.org/10.1049/qt2.12113>.
- [20] S. Chatterjee, A. Das, An ensemble algorithm using quantum evolutionary optimization of weighted type-II fuzzy system and staged Pegasos Quantum Support Vector Classifier with multi-criteria decision making system for diagnosis and grading of breast cancer, *Soft Comput* 27 (2023) 7147–7178. <https://doi.org/10.1007/s00500-023-07939-x>.
- [21] Y. Nasir, K. Kadian, V. Kumar, A. Wary, Harnessing Quantum Computing: A Comparative Study in Skin Disease Detection with Traditional ML, in: T. Senjyu, C. So-In, A. Joshi (Eds.), *Smart Trends in Computing and Communications*, Springer Nature, Singapore, 2024: pp. 361–370. https://doi.org/10.1007/978-981-97-1323-3_30.
- [22] N. Singh, S.R. Pokhrel, Modeling Quantum Machine Learning for Genomic Data Analysis, (2025). <https://doi.org/10.48550/arXiv.2501.08193>.
- [23] M. Munshi, R. Gupta, N.K. Jadav, S. Tanwar, A. Nair, D. Garg, Quantum Machine Learning-based Lung Cancer Prediction Framework for Healthcare 4.0, in: *2024 Asia Pacific Conference on Innovation in Technology (APCIT)*, 2024: pp. 1–6. <https://doi.org/10.1109/APCIT62007.2024.10673456>.
- [24] FidelityQuantumKernel - Qiskit Machine Learning 0.8.2, (n.d.). <https://qiskit-community.github.io/qiskit-machine-learning/stubs/qiskit_machine_learning.kernels.FidelityQuantumKernel.html#> (accessed 15.06.2025).
- [25] J.P. Miguel Patricio, Breast Cancer Coimbra, (2018). <https://doi.org/10.24432/C52P59>.
- [26] N.V. Bendi Ramana, ILPD (Indian Liver Patient Dataset), (2022). <https://doi.org/10.24432/C5D02C>.
- [27] Unknown, Heart Failure Clinical Records, (2020). <https://doi.org/10.24432/C5Z89R>.

- [28] I. Straw, H. Wu, Investigating for bias in healthcare algorithms: a sex-stratified analysis of supervised machine learning models in liver disease prediction, *BMJ Health Care Inform* 29 (2022) e100457. <https://doi.org/10.1136/bmjhci-2021-100457>.
- [29] M. Patrício, J. Pereira, J. Crisóstomo, P. Matafome, M. Gomes, R. Seiça, F. Caramelo, Using Resistin, glucose, age and BMI to predict the presence of breast cancer, *BMC Cancer* 18 (2018) 29. <https://doi.org/10.1186/s12885-017-3877-1>.
- [30] D. Chicco, G. Jurman, Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone, *BMC Med Inform Decis Mak* 20 (2020) 16. <https://doi.org/10.1186/s12911-020-1023-5>.
- [31] I.T. Jolliffe, J. Cadima, Principal component analysis: a review and recent developments, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374 (2016) 20150202. <https://doi.org/10.1098/rsta.2015.0202>.
- [32] G.T. Reddy, M.P.K. Reddy, K. Lakshmana, R. Kaluri, D.S. Rajput, G. Srivastava, T. Baker, Analysis of Dimensionality Reduction Techniques on Big Data, *IEEE Access* 8 (2020) 54776–54788. <https://doi.org/10.1109/ACCESS.2020.2980942>.
- [33] Z. Özpolat, Kuantum Tabanlı Boyut İndirgeme ve Sınıflandırıcı Gerçekleştirilmesi, MSc Thesis, Firat University, Elazığ, Turkey, 2023.
- [34] M. Kaur, K. Jain, A. Singla, K. Kadian, Quantum Exploration in Ransomware Detection with Conventional Machine Learning Approaches, in: 2024 IEEE International Conference on Contemporary Computing and Communications (InC4), IEEE, 2024, pp. 1–8. <https://doi.org/10.1109/InC460750.2024.10649082>.
- [35] V.N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, 2000. <https://doi.org/10.1007/978-1-4757-3264-1>.
- [36] K.C. Chua, V. Chandran, U.R. Acharya, C.M. Lim, Application of Higher Order Spectra to Identify Epileptic EEG, *Journal of Medical Systems* 35 (2011) 1563–1571. <https://doi.org/10.1007/s10916-010-9433-z>.
- [37] D. Anguita, S. Ridella, F. Riviello, R. Zunino, Quantum optimization for training support vector machines, *Neural Networks* 16 (2003) 763–770. [https://doi.org/10.1016/S0893-6080\(03\)00087-X](https://doi.org/10.1016/S0893-6080(03)00087-X).
- [38] A. Zeguendry, Z. Jarir, M. Quafafou, Quantum Machine Learning: A Review and Case Studies, *Entropy* 25 (2023) 287. <https://doi.org/10.3390/e25020287>.
- [39] M. Aly, S. Fadaaq, O.A. Warga, Q. Nasir, M.A. Talib, Experimental Benchmarking of Quantum Machine Learning Classifiers, in: 2023 6th International Conference on Signal Processing and Information Security (ICSPIS), 2023: pp. 240–245. <https://doi.org/10.1109/ICSPIS60075.2023.10343811>.
- [40] A. Thomsen, Comparing Quantum Neural Networks and Quantum Support Vector Machines, MSc Thesis, ETH Zurich, 2021, 97 p. <https://doi.org/10.3929/ETHZ-B-000527559>.
- [41] S. Altares-López, A. Ribeiro, J.J. García-Ripoll, Automatic design of quantum feature maps, *Quantum Sci. Technol.* 6 (2021) 045015. <https://doi.org/10.1088/2058-9565/ac1ab1>.
- [42] A. Daspal, OptiPauli: An algorithm to find a near-optimal Pauli Feature Map for Quantum Support Vector Classifiers, in: 2022 IEEE International Conference on Quantum Computing and Engineering (QCE), 2022: pp. 828–830. <https://doi.org/10.1109/QCE53715.2022.00133>.