Grenze International Journal of Engineering and Technology, June Issue

Revolutionizing Lung Cancer Prognosis using a Quantum Approach for Uncertain Medical Data

Jayasheela S D[1], V N Manjunath Aradhya[2], Manoj Kumar C S[3] and Nikhil D Bharadwaj[4]
[1-4]Department of Computer Applications, JSS Science and Technology University, Mysuru, India
Email: aradhya@jssstuniv.in

***Abstract*— Traditional machine learning methods does not always work faster enough to identify lung cancer in patient because of the limited availability of clinical data. This study proposed a Quantum SVM based approach for lung cancer prediction using a publicly available dataset containing 1158 patients records with 14 clinical features and also comparing both traditional SVM and Quan-tum SVM performance. The 8-qubit QSVM gained the highest accuracy of 88.57%, while the classical SVM achieved 88.35% results in lung cancer prediction. This research shows the efficacy of the Quantum Support Vector Machine (QSVM) in overcoming the limitations of addressing the uncertain-ty issue while working with medical data and also shows how quantum approach is better than the classical SVM.**

***Keywords*— Lung Cancer Prediction, Quantum Machine Learning, Classical SVM, Quantum Support Vector Machine (QSVM), Quantum Kernel.**

I. INTRODUCTION

Lung cancer is a medical issue which has effected on the health situation of the world. Lung cancer is one of the leading causes of cancer death in the world, and timing of survival of the patient depends on the precision of diagnosis. The survival rate can be increased by the early detection of lung cancer at a young age. The signs of weakness in early diagnosis plays a vital role in diagnosing and treating with the accurate medication [1]. Even though machine learning (ML) implementation in medicine has promising solutions, its practical performance needs to be vigorously undermined in real time in terms of feasibility which is usually not validated. The traditional algorithms like Support Vector Machine (SVM), might fail to in data having uncertainty values. This weakness creates an obvious necessity for something more advanced, within the very huge, flexible model of computation, and high-dimensional feature spaces [2]. Quantum computing models shows exponential increase in the space of computation, and effect of quantum states in a high-dimensional space such as Hilbert space, which is huge than any space than a classical kernel.

A multiomic QML (MQML) framework that could combine both the data of genomics and proteomics to classify and discover biomarkers of cancer accurately. The incorporation of heterogeneous omic data into a quantum system resulted in a high dimensionality diagnosis results [3]. The importance of data normalization methods in the classical SVM models increased by 15% and showed the importance of data normalization by performing an effective quantum-classical comparison [4]. An ensemble SVM model with a SMOTE balancing and advanced feature selection gained an accuracy of 92% [5]. Quantum circuit design using ZZFeatureMaps with Hadamard, rotation, and controlled-Z gates. Balancing expressiveness and hardware feasibility on near-term NISQ devices was proposed to show the importance of feature mapping [6].

Grenze ID: 01.GIJET.12.2.18_25

Quantum Approximate Optimization Algorithm (QAOA) for feature selection with QSVM classifiers and standard PCA algorithms introduced a new feature selection that can be applied for Quantum Machine Learning (QML) [7].
A study addressed survival prediction using Quantum Support Vector Regres-sion (QSVR) which resulted in better prediction of continuous survival outcomes [8]. A comprehensive classical end-to-end optimized ML framework addressed feature extraction by fully integrating quantum methods for lung cancer predic-tion [9]. A quantum-enhanced feature spaces was introduced which demonstrat-ed the effectiveness of quantum embedding for complex biomedical datasets like lung cancer [10].
This paper proposes a hybrid Quantum Support Vector Machine (QSVM) for the prediction of lung cancer and also provides strong experimental evidence that use of quantum computing enhances the performance in disease predictabil-ity and presents evidence that hybrid Quantum SVM real working component.

II. METHODOLOGY

Unlike the classical methods of computing, the proposed methodology incorpo-rates a Quantum SVM workflow model that strengthens predictive capability of lung cancer diagnosis, utilizing quantum feature representation. Fig. 1 represents an overall diagram of the model with the data preprocessing and the evaluation of the model in a series.

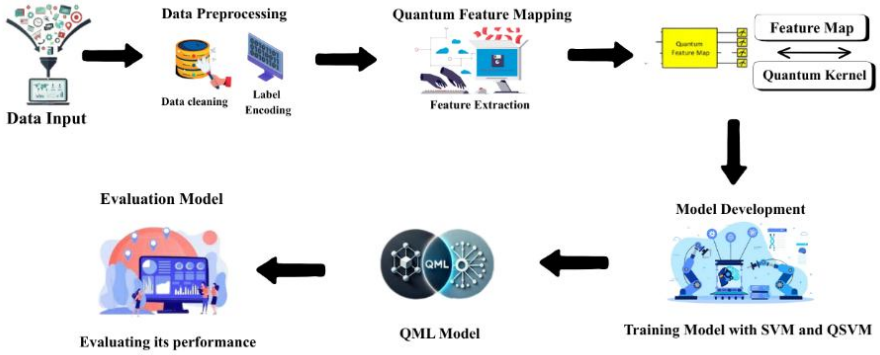

Fig. 1. Proposed Methodology

A. Data Preprocessing

Pre-processing of the data was a necessary to make the understandable data for Quantum machine learning algorithms. Numerical encoding of categorical data was applied to change GENDER into binary form (Male = 0, Female = 1). Z-score normalization was used to ensure that all features had equal contribution in learning by its models. The change was mathematically translated as:

$$Z = \frac{(x-\mu)}{\sigma} \qquad (1)$$

where x is the feature value, μ is the mean, and σ, is the standard deviation. This was done to have a zero mean and unit variance across features. After the data normalization, a stratified sampling method was used to divide the data into three subsets to preserve the proportional distribution of 60% training, 20% validation, and 20% testing.

B. Quantum Kernel and Quantum Feature Mapping

A quantum embedding is applied in order to standardize the clinical characteris-tics in a high-dimensional Hilbert space for quantum-enhanced learning. Equa-tion 2 provides the ZZFeatureMap circuit with two repeats.

$$K(x, x') = |\langle \phi(x) | \phi(x') \rangle|^2 \qquad (2)$$

The fidelity kernel is then used to define the quantum similarity between two inputs. In order to guarantee one component per qubit, the feature map object and the FidelityQuantumKernel evaluator were created. The output from this stage provided a configured quantum feature map with a quantum kernel routine to build the train-test matrices after dimensionality reduction for QSVM training and inference.

C. Dimensionality Reduction with PCA

Principal Component Analysis (PCA) was used to simplify computations and transform data to be simulated using quantum computers. PCA minimally re-duced the 15-dimensional input space to 2, 4, 6 and 8 principal components. Equation 3 shows the PCA transformation.

$$Y = W \cdot \mathrm{X} \qquad (3)$$

where X represent the original features and array. W signifies the principal com-ponent vectors and 0. Y represents the transformed feature representation.

D. Model Development and Training

The experimental design of the study consisted of two models a Classical Sup-port Vector machine (SVM) and a Quantum Support Vector machine (QSVM) to test the efficacy of quantum enhanced learning for prediction of lung cancer.

The two models are closely related as they are both based on the principle of margin-based classification but they have a fundamental difference in the input feature space representation.

Classical SVM

The classical implementation of SVM was done using SGDClassifier with a hinge loss function which is in line with the goal of a linear SVM.

The algorithm was expected to find the best hyperplane that will consist the largest margin be-tween two classes of positive and negative as shown in equation 4.

$$\min_{w,b,\xi} \frac{1}{2} \| \mathrm{w} \|^2 + \mathrm{C} \sum_{i=1}^{n} \xi_i \qquad (4)$$

where w is the weight vector of the decision boundary, b is the bias, ξi, are slack variables of tolerance to misclassification, and C is a constant of regularization.

Quantum SVM (QSVM)

The Quantum Support Vector Machine (QSVM) is the hybrid of classical and quantum approach. The fundamental classification task of a QSVM is the same one performed by its classical SVM, except that the transformation in quantum states of classical data takes place in a quantum feature space, where a parame-terized quantum circuit is used to do the transformation. This embedding allowed the model to represent the complicated nonlinear dependencies which might not be easy for the classical algorithms. Equation 5 shows how the QSVM was creat-ed.

$$U_{\Phi(\mathrm{x})} = \left(\prod_j H_j\right) exp\left(i \sum_j x_i \, Z_j\right)\left(\prod_{j<k} \exp(i(x_j - \pi)(x_k - \pi) Z_j Z_k\right) \qquad (5)$$

It creates complex, deeply entangled quantum states that can effectively encode nonlinear relationships for more extensive representational space than classical kernels.

$$u_q^{(n)(z_1,\ldots,z_n)} = \left(\sum_{j=1}^{n} \quad H_j\right)\left(\prod_{i=1}^{n} \quad e^{icz_i}\right)\left(\prod_{(i,j):\, i<j\leq k-n} \quad e^{(c(z_j - z_i) - z_n) E_i I}\right) \qquad (6)$$

Where n = 2,4,6,8 qubits

Quantum Kernel Computation was ultimately performed by using the equation 7, which estimates the similarity between two quantum states,

$$K(x, x') = |\langle \phi(x) | \phi(x') \rangle|^2 \qquad (7)$$

This fidelity measure is used to measure the proximity of two data points in the quantum feature space.

There were 2, 4, 6, and 8-qubit architectures to develop four different QSVM configurations, which correspond to PCA-reduced reduction dimensions of in-put, respectively.

These matrices of kernels code the comparisons of all training samples in pairs to give the input to the classical SVM algorithm to do the final classification step.

III. Results and Discussion

To evaluate the performance comparison of classical Support Vector Machine (SVM) and quantum-enhanced QSVM versions against each other through clini-cally relevant metrics. The experiment was performed on a common dataset and same split of 60% training, 20% validation, and 20% testing.

A. Dataset Description

The study used the lung cancer dataset which is publicly available in Kaggle released by Chandan Mishra/Somesh, which contains 1,158 records of patients with 15 features (14 predictive + 1 target). Table I presents the complete charac-teristics of the dataset and parameters in the experimental setup. The dataset has realistic clinical imbalances between classes with 32.1% positive and 67.9% negative mirroring the actual prevalence of lung cancer and sensitivity and spec-ificity must be optimized to reduce the number of false negatives and false posi-tives.

TABLE I. DATASET OVERVIEW

Characteristics	Details
Total Samples	1158
Positive Cases (y = 1)	372 (Lung Cancer)
Negative Cases (y = 0)	786 (No Lung Cancer)
Total Features	15 (14 feature + 1 target)
Target Feature	1 (LUNG_CANCER)

B. Experimental Setup

All experiments were performed in the cloud platform of Google Colab with GPU enabled acceleration, using Python 3.10 and its libraries and the dataset was stored in Google Drive and majorly the study utilized IBM Qiskit Library to per-form the entire Quantum Simulation and also utilized Quantum Machine Learn-ing and Quatum Aer libraries to accomplish the entire experiment.

C. Experiemntal Results

All detailed metrics pertaining to the performance of the classical SVM baseline and four QSVM versions working with 2, 4, 6, and 8 qubits are illustrated by Table 2. Th models were evaluated with metrics shown in Table 2. Figure 2 shows how the accuracy progresses through the five models, the 8-qubit QSVM has a maximum accuracy measure of 88.57% for overall testing outpacing the respective benchmark set by classical SVM at 88.35%. This proves that quantum feature mapping with sufficient qubit resources can equal and even slightly sur-pass classical performance. The clearer upward slope demonstrates the trend in trajectory with the increase in qubit counts from 2 to 8. The 2-qubit configura-tion has underperformed remarkably at 79.74%. The underperformance of the 2-qubit system is also attributed to the severe reduction in dimensions. Therefore, the 6-qubit QSVM (86.20%) makes a good compromise.

TABLE II. COMPARISONS BETWEEN SVM AND PERFORMANCE METRICS. QSVM CHARACTERISTICS

Model	Accuracy	Sensitivity	Specificity	PPV	NPV	F1 Score
SVM	88.35	75.52	93.5	84.99	89.47	88.19
QSVM (2 qubits)	79.74	87.59	69.97	78.47	81.81	79.5
QSVM (4 qubits)	84.91	95.34	71.84	80.92	92.5	84.58
QSVM (6 qubits)	86.2	92.24	78.64	84.39	89.01	86.08
QSVM (8 qubits)	88.57	77.77	93.68	85.36	89.89	88.42

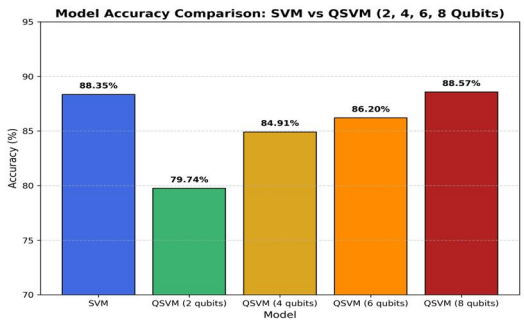

Fig. 2. Comparison of Models Accuracy: SVM vs QSVM (2, 4, 6, 8 Qubits).

D. Analysis

It is in sensitivity performance that the most evident clinical finding comes about; the 4-qubit QSVM attained an outstanding 95.34% sensitivity, far higher than that of the classical SVM (75.52%) and 8-qubit QSVM (77.77%),

with a reasonable translation: 19 additional cancer cases identified per each 100 pa-tients. In the meantime, classical SVM and 8-qubit QSVM performed better at specificity, with both cases judging correctly non-cancer situations in 94 out of every 100 sampled healthy individuals, resulting in fewer false positive diagno-ses. The 8-qubit QSVM had a positive predictive value of 85.36%, which was about similar as that of classical SVM (84.99%), but the best was achieved by 4-qubit QSVM with a negative predictive value of 92.50%. The highest F1-Score of 88.42% could thus be associated with the 8-qubit QSVM and indicated the best balance between precision and recall.

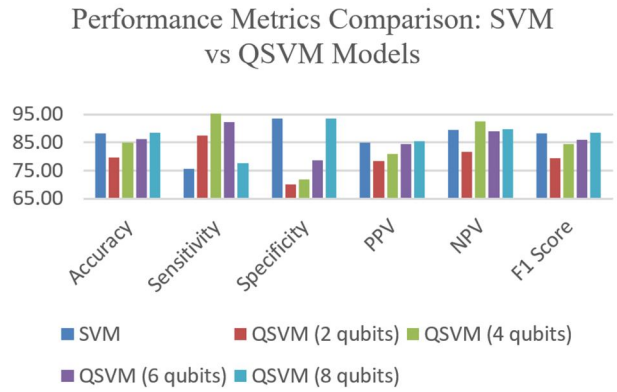

Fig. 3. Performance Metrics Comparison: SVM vs QSVM Models.

As seen in Figure 3, the performance characteristics can be classified into three categories: disease detection maximization with the 4-qubit QSVM (95.34% re-call) best for population screening, the classical SVM and 8-qubit QSVM focus-ing on correctly rejecting false non-cancer cases (93% specificity) best for con-firmatory diagnostics, and the 6-qubit QSVM falls can be utilized for general-purpose clinical tests. This shows that the Qubit selection needs to be tuned ac-cording clinical requirements. The gradual improvement in accuracy from 2 to 8 qubits indicates that the quantum advantage scales with the increase of Qubits. In terms of computational time the QSVM takes approximately 15-20 mins on NVIDIA T4 hardware.

IV. Conclusion

Lung cancer is a major global health concern and one of the leading causes of cancer related deaths, Accurate and early lung cancer detection is paramount for effective patient management requires a major improvement as the traditional machine learning models struggle to address the uncertainties in clinical da-tasets. This research presents the efficacy of Quantum Support Vector Machine (QSVM) to overcome these limitations by gaining better results in comparison to classical SVM. Future works will focus on deploying models on actual quantum computers such as IBM Quantum and IonQ with error mitigation techniques and utilize CT scan imaging, genomic biomarkers, and multi-class cancer staging classification.

References

[1] S. S. Dutta, S. Sandeep, S. S. Sridevi, M. R. Sridhar, and C. Singh, “Quantum Kernel-based Support Vector Machine with Quantum Approximate Optimization Algorithm Embedding for improved Lung Cancer Prediction,” in 2024 15th In-ternational Conference on Computing Communication and Networking Technol-ogies (ICCCNT), 2024.

[2] Bhat, Hilal Ahmad, Farooq Ahmad Khanday, and Khurshed Ahmad Shah. "Optimal quantum circuit decomposition of reversible gates on IBM quantum computer." 2022 5th International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT). IEEE, 2022.

[3] M. K. Saggi and S. Kais, “MQML: Multi-omic Quantum Machine Learning based Cancer Classification, Biomarker Identification in Human Lung Adenocar-cinoma,” in 2024 IEEE International Conference on Quantum Computing and Engineering (QCE), 2024.

[4] T. Kanimozhi and S. Sridevi, “SVM Kernel Methods with Data Normalization for Lung Cancer Survivability Prediction Application,” in 2024 15th Internation-al Conference on Computing Communication and Networking Technologies (ICCCNT), 2024.

[5] J. Alam, S. Alam, and A. Hossan, “Multi-stage lung cancer detection and pre-diction using multi-class SVM classifier,” in Proceedings of the 2018 IEEE In-ternational Conference on Computational Intelligence and Communication Net-works (IC4ME2), 2018, pp. 118–122.

[6] T. S. Humble, F. A. Mohiyaddin, H. Thapliyal, E. M. Coreas, and R. S. Ben-nink, “Quantum computing circuits and devices,” IEEE Des. Test, vol. 36, no. 3, pp. 69–94, Jun. 2019.
[7] G. Turati, M. Ferrari Dacrema, and P. Cremonesi, “Feature Selection for Clas-sification with QAOA,” in 2022 IEEE International Conference on Quantum Computing and Engineering (QCE), 2022, pp. 782–785
[8] J. A. Bartholomai and H. B. Frieboes, “Lung Cancer Survival Prediction via Machine Learning Regression, Classification, and Statistical Techniques,” in 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Dec. 2018, pp. 632–637.
[9] S. S. Raoof, M. A. Jabbar, and S. A. Fathima, “Lung Cancer Prediction Using Machine Learning: A Comprehensive Approach,” in 2020 2nd International Con-ference on Innovative Mechanisms for Industry Applications (ICIMIA), Mar. 2020, pp. 108–115.
[10] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, “Supervised learning with quantum-enhanced feature spaces,” Nature, vol. 567, pp. 209–212, 2019; preprint: arXiv:1804.11326, 2018.
[11] C. Mishra (Somesh), "More Accurate Lung Cancer Dataset," Kaggle,2025.https://www.kaggle.com/datasets/chandanmsr/more-accurate-lung-cancer-dataset.