FUDMA Journal of Engineering and Technology
ISSN: 3092-9385
https://sites.google.com/fudutsinma.edu.ng/fjet
Volume 1, Issue 2, 2025, 782-794

Leveraging Quantum Machine Learning for Early Ovarian Cancer Diagnosis

Ukange N. SYBIL[1*], Hadiza A. UMAR[2], Ogar M. OKO[3], Habeebah A. KAKUDI[4], Usman MAHMUD[5], Alex AARON[6]

[1,2,4,6]Department of Computer Science, Bayero University, Kano, Nigeria
[3]Department of Computer Science, Federal University, Wukari, Nigeria
[5]Department of Software Engineering, Northwest University, Kano, Kano State, Nigeria

[1*]nguemosybi@gmail.com, [2]haumar.cs@buk.edu.ng, [3]ogarmichael@fuwukari.edu.ng, [4]hakakudi.cs@buk.edu.ng, [5]umahmud@nwu.edu.ng, [6]alex.aaron@sun.edu.ng

Abstract

Ovarian cancer remains one of the leading causes of cancer-related mortality among women worldwide, largely because most cases are diagnosed at advanced stages. Current diagnostic tools and classical machine learning models often show limited sensitivity and specificity, particularly when applied to the complex, high-dimensional datasets required for accurate prediction. These limitations highlight the need for more powerful computational approaches capable of extracting subtle diagnostic patterns. The aim of this study is to enhance the accuracy and efficiency of ovarian cancer diagnosis by using quantum-inspired machine learning approach such as Quantum Support Vector Machines (QSVMs). The study utilized QSVMs to analyse clinical and genomic datasets, leveraging quantum mechanics principles like superposition and entanglement to process data more effectively than traditional machine learning models. The QSVM model was developed, trained, and evaluated using performance metrics such as accuracy, sensitivity, specificity, and processing efficiency. The results demonstrated that QSVMs achieved a diagnostic accuracy of 92%, outperforming traditional support vector machines and other classical models. Additionally, the processing time for diagnosis was reduced from 45 minutes to 20 minutes, providing a faster and more reliable workflow. The QSVMs also excelled in analysing multi-omics data, enabling the identification of early-stage ovarian cancer biomarkers and supporting personalized treatment strategies. In conclusion, this study demonstrates the potential of QSVMs to transform ovarian cancer diagnostics by addressing key limitations of conventional methods. The findings underscore the importance of adopting quantum-inspired machine learning in medical applications and encourage further exploration of these advanced algorithms in improving healthcare outcomes.

Keywords: *Ovarian cancer, Quantum Support Vector Machines (QSVMs), Quantum-inspired machine learning, Early diagnosis, machine learning.*

1.0 Introduction

Ovarian cancer, as defined by both the American Cancer Society [1], is a malignant tumor that develops in the tissues of an ovary. Among women, ovarian cancer ranks as the eighth most prevalent cancer and is responsible for the fifth highest number of cancer-related fatalities globally, indicating a high mortality rate [4]. A high death rate is mainly attributable to late-stage detection and the lack of particular symptoms in the early stages of ovarian cancer, which is the deadliest gynaecological malignancy [5]. According to statistics from 185 countries, there were 295,414 new cases of skin cancer and 184,799 deaths in 2018 [6].

Ovarian cancer continues to be a major global health concern due to its poor prognosis and high mortality rates. In 2018, global estimates recorded 295,414 new cases and 184,799 deaths, reflecting its significant burden. More recent figures from 2020 show that ovarian cancer cases rose to approximately 313,959, with 207,252 deaths worldwide [7]. This information highlights the severity of ovarian cancer since it is one of the most common new cancers (almost 4 per cent) and also one of the most common cancer deaths among women (approximately 5 per cent). Ovarian cancer has a comparably low survival rate relative to other cancers because the five-year survival rate is about 48% in developed countries and significantly lower in the developing regions because of the differences in healthcare services availability [8]. Late diagnosis of ovarian cancer is one of the main issues when managing the condition because the disease is diagnosed in its advanced stages in about 70 percent cases [9]. Such late diagnosis is a significant cause of the high death rate since the cancer is already metastatic when it is diagnosed. At the initial stages of ovarian cancer when the tumour remains in the ovaries, there is a five-year survival rate of more than 90% but only an estimated 20-25% of ovarian cancer cases are diagnosed in this early phase [10].

The geographical location is also a major determinant of ovarian cancer with Central and Eastern Europe having the highest rates then the North America and Oceania [11]. On the contrary, in Asia and Africa, there are the lowest incidence rates. Such differences are determined by genetic inclination, reproductive history, access to early detection and treatment services. Recent developments show that the incidence and mortality of ovarian

Received: 30/10/2025; Revised: 20/11/2025; Accepted: 24/11/2025; Published: 22/12/2025

cancer in the high-income nations are slightly decreasing because of the use of oral contraceptives, which has been reported to decrease the risk of ovarian cancer [12]. Nevertheless, the disease burden is still great, particularly in low and middle-income countries, where there is a low availability of effective screening and treatment tools. The financial implication of ovarian cancer is also dramatic, as the cancer is a significant contributor to the spending of the healthcare sector on cancer treatment. Ovarian cancer treatment in the United States alone was estimated to cost more than 5 billion dollars every year in 2020, and much of these expenses are due to the later stages of the disease, which are harder and more expensive to treat [13]. The fact that this economic burden is increased by the fact that illness and premature death result into a lack of productivity underscores the necessity to adopt better diagnostic and therapeutic measures.

Given these challenges, there is a growing interest in leveraging advanced technologies such as quantum-inspired machine learning algorithms to enhance the early detection and diagnosis of ovarian cancer. The convergence of machine learning (ML) and quantum computing has opened new frontiers in the quest for better diagnostic tools, particularly for complex diseases like ovarian cancer. According to Shimizu et al. [14], machine learning is a subfield of artificial intelligence that aims to develop algorithms that can "learn" from data in order to make decisions or predictions. These algorithms have shown immense potential in identifying patterns and correlations that may be imperceptible to human observers. In the realm of medical diagnostics, in order to enhance the precision of diagnoses and forecast outcomes, ML algorithms can sift through mountains of medical data, such as imaging, genetic information, and patient records [15]. In the context of ovarian cancer, ML has been applied to develop models that predict disease risk, classify tumor subtypes, and identify potential biomarkers [16]. When compared to conventional computers, quantum computers use quantum mechanical principles to process information in a fundamentally different manner. Because of superposition, quantum bits (qubits) may exist in more than one state at once, in contrast to classical bits that can only take on two values [17]. Furthermore, qubits may be interdependent thanks to quantum entanglement, which opens the door to processing a multitude of possibilities all at once. Because of these features, quantum computers can outperform conventional computers in solving certain complicated problems [18]. Although large-scale, practically applicable quantum computers are still in the works, conventional computers may nevertheless realise substantial performance improvements by using techniques inspired by quantum computing. Machine learning and, by extension, medical diagnostics, may benefit greatly from these algorithms because of how well they handle optimisation difficulties and complicated pattern recognition tasks.

Quantum-inspired machine learning algorithms are designed to leverage the advantages of quantum computing principles without requiring fully functional quantum hardware. These algorithms draw inspiration from quantum mechanics to enhance classical ML models' capabilities, making them more efficient and effective in processing large and complex datasets [19]. Several types of quantum-inspired algorithms are particularly relevant to medical diagnostics: Quantum Support Vector Machines (QSVMs), which extend classical support vector machines by incorporating quantum computing techniques to enhance their ability to classify data; Quantum Boltzmann Machines (QBMs) for deep learning; and quantum-inspired optimisation algorithms that use quantum principles to outperform classical algorithms [20], [21].

The integration of quantum-inspired ML algorithms in ovarian cancer diagnosis promises several transformative benefits. By analyzing large-scale, high-dimensional datasets, such as genomic sequences and medical imaging, quantum-inspired algorithms can identify subtle biomarkers and patterns associated with early-stage ovarian cancer. This capability significantly enhances the potential for early detection, which is crucial for improving patient outcomes [22]. Quantum-inspired ML models can improve the sensitivity and specificity of diagnostic tests. For instance, combining CA-125 levels with other biomarkers and imaging data analyzed through QSVMs can lead to more accurate diagnoses, reducing the likelihood of false positives and negatives. Additionally, these algorithms can analyze multi-omics data to identify individual genetic and molecular profiles, enabling personalized diagnostic and treatment strategies. This approach can help in tailoring therapies to the specific characteristics of each patient's cancer, improving treatment efficacy and reducing side effects [23].

Furthermore, by integrating clinical data, family history, and genetic information, quantum-inspired algorithms can better stratify patients based on their risk of developing ovarian cancer. This stratification can inform screening strategies and preventive measures for high-risk individuals [9]. Quantum-inspired optimization algorithms can enhance the efficiency of diagnostic workflows, reducing the time and cost associated with traditional diagnostic methods. This optimization is particularly valuable in resource-limited settings, where access to advanced diagnostic tools may be restricted [24].

2.0 Literature review

Zhou and colleagues [25] used machine learning models to estimate ovarian cancer risk and have shown that random forest classifiers are highly accurate in the distinction of the subtypes of tumours. In the same way, Wu et al. [26] designed a multicentered artificial intelligence (AI) prediction system for preoperative that had better diagnostic as well as prognostic accuracy, when compared to traditional biomarkers like CA-125. On the one hand,

Received: 30/10/2025; Revised: 20/11/2025; Accepted: 24/11/2025; Published: 22/12/2025

Erickson et al. [27] demonstrated that deep learning models were capable of detecting subtle features in the radiological scans that human professionals failed to detect. Nevertheless, Jelovac and Armstrong [28] stated that in spite of these developments, the lack of reproducibility and small sample size are impediments to clinical application.

To solve the biomarker discovery problem, Ghobadi et al. [29] employed quantum machine learning to combine multi-omics data into improved sensitivity in the early-stage ovarian cancer diagnosis. Beyerlein and Chien [30], on the other hand, applied bioinformatics pipelines to traditional machine learning and observed inconsistent predictive power across data sources, and suggested the use of more efficient algorithms. Rehman et al. [31] summarized the potential application of quantum computing in ovarian cancer and identified the possibility to resolve the problems with high-dimensional datasets. On the same note, Ozhan et al. [32] used XGBoost and gradient boosting algorithms, which validated their usefulness in the prediction of ovarian cancer, but also found weaknesses in their interpretability.

Maheshwari et al. [33] systematically searched quantum machine learning implementation in biomedical research and found the field of oncology as an untapped, yet promising, area of implementation. In more recent times Rahimi [34] suggested oncological uses of quantum-inspired algorithms, claiming that their optimisation properties had the potential to significantly improve cancer classification.

These contributions notwithstanding, there have been massive gaps. The heterogeneity and dimensionality of ovarian cancer information is a challenge to most classical machine learning systems and quantum-inspired systems, despite their potential, have not been empirically validated using combined clinical, genomic, and imaging data. This gap is intended to be filled by the current study that will design and assess quantum-inspired machine learning models to enhance the accuracy, efficiency, and interpretability of the diagnosis of ovarian cancer.

3.0 Materials and Methods

This study employs experimental techniques with a design-oriented approach to develop a model that precisely forecasts the likelihood of ovarian cancer. Specifically, the work uses a modified form of the Hybrid Knowledge Discovery Process for Data Mining, customised to integrate quantum-inspired algorithms including Quantum Support Vector Machines (QSVMs). This method enables a more accurate and predictive study of the complicated interactions among several clinical and genomic aspects in ovarian cancer diagnosis.

The particular requirements of medical research and the necessity for innovative machine learning methods motivated the change of the Cross-Industry Standard Processes for Data Mining (CRISP-DM) paradigm. The adaptation centres on the special difficulties of including quantum-inspired algorithms into the data mining process and optimising each phase to raise the diagnostic model accuracy, efficiency, and dependability. Figure 3.1 below shows the several phases of the modified CRISP-DM paradigm.

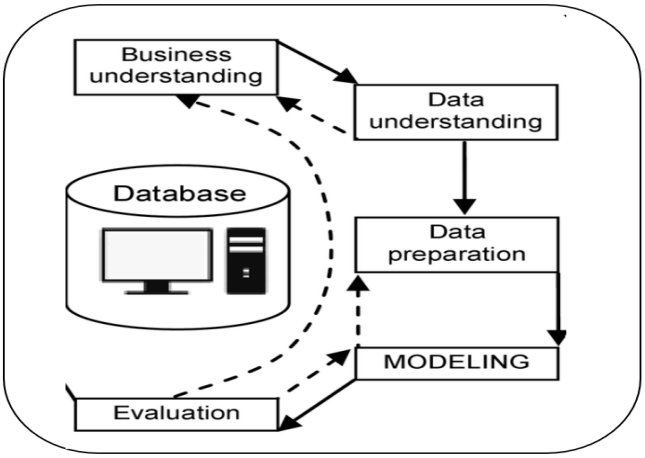

Figure 1: CRISP-DM Phases and Tasks [35]

First, business understanding established the diagnostic challenges and the potential of Quantum Support Vector Machines (QSVMs) for early detection. Data understanding involved exploring clinical and genomic datasets from cancer registries, while data preparation ensured consistency through transformation and standardisation. QSVM-based models were then developed to capture complex interactions, and their performance evaluated using accuracy, sensitivity, specificity, and efficiency. Deployment provided practical tools for clinical use, while monitoring ensured continuous assessment, enabling updates to maintain reliability and relevance in real-world healthcare settings.

Received: 30/10/2025; Revised: 20/11/2025; Accepted: 24/11/2025; Published: 22/12/2025

3.1 Dataset Acquisition

The dataset for this study contains ovarian cancer diagnostic data. The dataset focuses on evaluating various clinical and genomic features, providing valuable insights into the effectiveness of quantum-inspired machine learning algorithms in improving the accuracy of ovarian cancer diagnosis.

3.1.1 Sample and Sampling Techniques

This study's sample consist of clinical and genomic data from patients diagnosed with ovarian cancer. The dataset was used to train and test the Quantum Support Vector Machines (QSVMs) models for ovarian cancer diagnosis. The sample contains diverse patient data from various demographics and for people in varying stages of cancer. This is to help get a good representation of the data for better generalization of the models' results. Other relevant features such as tumor characteristics and treatment histories are also included in the data.

To better represent the subgroups within the ovarian cancer population, this study employs a stratified sampling technique. This means that the dataset utilized divides the population into strata depending on various variables such as age, cancer stage, genetic mutation etc. The models are then trained from data gotten from each of the designated stratum, this helps to make the model more representative of the population. The data is used for both training the models and testing them and the models are tested on data that was not used during training, providing an unbiased evaluation of their performance.

3.2 Implementation Phase

The proposed models are implemented using TensorFlow. This phase includes the experimental setup, the hardware and software requirements, the data cleaning process, model formulation and training, and also model evaluation.

3.2.1 Experimental Setup

The experimental setup for this study is designed to implement and evaluate quantum-inspired machine learning algorithms to improve ovarian cancer diagnosis. The setup consists of several key phases, leveraging both classical and quantum-inspired machine learning tools to develop and validate models. The Quantum Support Vector Machines (QSVMs) model was implemented using quantum simulators available through IBM's Qiskit framework. QSVMs classify patient data into different diagnostic categories, focusing on distinguishing between early and late-stage ovarian cancer. The model was be trained on multiple subsets of the dataset to fine-tune its hyperparameters and ensure robust performance.

The models were trained in a high-performance computing (HPC) environment equipped with GPUs and access to quantum simulators. QSVMs will rely on quantum simulators like Qiskit's Aer for emulating quantum operations. Hyperparameters such as learning rates, regularization coefficients, and quantum circuit depths was optimized using grid search and quantum optimization techniques to ensure optimal model performance.

3.3 Data Preprocessing

This This process focuses solely on preparing the raw data for analysis, including tasks such as data cleaning, formatting, handling missing values, and transforming variables into a suitable structure for modelling.

3.3.1 Data Cleaning

Data cleaning was performed to ensure the dataset was accurate, consistent, and suitable for developing quantum-inspired machine learning models. Initial checks addressed formatting inconsistencies across the clinical and biomarker files. Missing values were handled using a structured rule: records with more than 30% missing data were removed due to insufficient information for reliable modelling, while variables with **less** than 10% missingness were imputed using the median**,** which is robust for skewed clinical data. Variables with 10–30% missingness were assessed for clinical relevance, and imputation was applied only when their distributions were stable and interpretable. Outliers were detected using a Z-score threshold of $|Z| > 3$ and removed when values were physiologically implausible. Additional noise reduction involved checking biomarker ranges against medical reference values to correct or exclude erroneous entries. This systematic procedure ensured a high-quality dataset for the subsequent modelling stages.

3.3.2 Data Normalization

Data normalization was employed to ensure that all features contribute equally to the model. This is due to the fact that the features within the dataset may have different scales, for example gene expression levels versus clinical measurements. The data was normalized, scaling features to a range of (0,1). Categorical variables like tumor type, patient demographics was encoded using techniques like label encoding so as to convert them into numerical form. Because of the high dimensionality of genomic data, feature selection algorithms was applied to

Received: 30/10/2025; Revised: 20/11/2025; Accepted: 24/11/2025; Published: 22/12/2025

reduce the number of features while preserving the most informative aspects of the data. This improves the computational efficiency of the quantum-inspired models.

3.3.3 Data Splitting

After these processes, the data was split into training (70%) and test (30%) sets. The training set was used to train the models, while the test set was used to assess their performance. To prevent overfitting, cross-validation techniques was used when developing the model to further validate them. By cleaning, transforming, and selecting the most relevant data, this study seeks to build robust models that can improve the accuracy, efficiency, and reliability of ovarian cancer diagnosis.

3.4 Model Evaluation Metrics

The evaluation phase is critical in determining the effectiveness and reliability of the quantum-inspired machine learning models for diagnosing ovarian cancer. This stage involves systematically assessing the performance of the models using various metrics and ensuring that the results are robust and generalizable. To measure the predictive power and diagnostic utility of the models developed (Quantum Support Vector Machines - QSVMs, a combination of evaluation metrics was employed. These metrics helped in understanding the accuracy, precision, and computational efficiency of the models.

Accuracy: The percentage of cases for which the prediction was accurate relative to the total number of occurrences is called accuracy. A broad indicator of the model's efficacy is provided by it.

where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

High accuracy indicates the model correctly predicts both ovarian cancer and non-cancer cases. However, accuracy alone may not be sufficient due to imbalanced datasets where the majority class (non-cancer) dominates.

Recall: Recall measures the model's ability to correctly identify ovarian cancer cases (true positives). It's critical for medical diagnostics to minimize false negatives (i.e., missing actual cancer cases).

High recall is essential in cancer diagnostics to ensure that most, if not all, cancer cases are correctly identified. A low recall means the model fails to detect many cancer cases, which can be life-threatening in real-world clinical applications.

Specificity: Specificity refers to the model's ability to correctly identify non-cancerous cases (true negatives). It minimizes false positives, which are important to avoid unnecessary treatments or tests.

High specificity is important to ensure that patients without ovarian cancer are not falsely diagnosed with the disease. A balance between specificity and sensitivity is crucial for model performance.

Precision: Precision is defined as the ratio of correct forecasts to total positive predictions, which includes both true and incorrect predictions. It provides insight into the reliability of positive cancer predictions.

Precision becomes important in cases where false positives carry significant consequences, such as unnecessary stress and invasive follow-up procedures for patients misdiagnosed with cancer.

F1- Score: The F1-score strikes a balance between recall and accuracy via a harmonic mean of the two. In cases when the data is skewed, it becomes invaluable.

The F1-score ensures that both false positives and false negatives are considered in a balanced manner, offering a comprehensive evaluation metric when dataset imbalance may skew accuracy.

Area Under the ROC Curve (AUC-ROC): The AUC-ROC measures the model's ability to distinguish between classes (cancer vs. non-cancer) across various threshold settings. An AUC score close to 1 indicates excellent model performance, whereas a score closer to 0.5 suggests the model is no better than random guessing.

3.5 Confusion Matrix for Binary Classification

The binary classification confusion matrix for this study involves the following components as presented in the Table 1:

Received: 30/10/2025; Revised: 20/11/2025; Accepted: 24/11/2025; Published: 22/12/2025

Table 1. Components of confusion matrix

		Predicted	
		Positive (Cancer Present)	**Negative (Cancer Absent)**
Actual	**Positive (Cancer Present)**	True Positive (TP)	False Negative (FN)
	Negative (Cancer Absent)	False Positive (FP)	True Negative (TN)

i. **True Positive (TP):** The model correctly predicts the presence of ovarian cancer.
ii. **False Negative (FN):** The model incorrectly predicts the absence of ovarian cancer when it is actually present.
iii. **False Positive (FP):** The model incorrectly predicts the presence of ovarian cancer when it is absent.
iv. **True Negative (TN):** The model correctly predicts the absence of ovarian cancer.

4.0 Results and Discussion

4.1 Data Visualisation and Description

The visual exploration of the dataset offered important insights into its structure, quality, and potential analytical implications. The initial snapshot of the first five rows (Figure 2) confirms the presence of clinically relevant biomarkers across 51 variables, indicating a feature-rich dataset suitable for modelling.

	SUBJECT_ID	AFP	AG	Age	ALB	ALP	ALT	AST	BASO#	BASO%	...	NEU	PCT	PDW	PHOS	PLT	RBC	RDW	TBIL	TP	UA
0	1	3.58	19.36	47	45.4	56	11	24	0.01	0.3	...	76.2	0.09	13.4	1.46	74	2.64	13.7	5.5	73.9	396.4
1	2	34.24	23.98	61	39.9	95	9	13	0.02	0.3	...	76.5	0.3	11.2	1.09	304	4.89	12.7	6.8	72	119.2
2	3	1.50	18.4	39	45.4	77	9	18	0.03	0.6	...	69.7	0.13	15.2	0.97	112	4.62	12	14.8	77.9	209.2
3	4	2.75	16.6	45	39.2	26	16	17	0.05	0.74	...	65.5	0.25	17.4	1.25	339	4.01	14.6	10.9	66.1	215.6
4	5	2.36	19.97	45	35	47	21	27	0.01	0.1	...	59.5	0.28	11.9	0.94	272	4.4	13.4	5.3	66.5	206

5 rows × 51 columns

Figure 2. Visualization of the first five rows of the dataset

The age-distribution plot (Figure 3) indicates a concentration of patients in mid-adult and older age groups, aligning with known epidemiological patterns of ovarian cancer. This demographic skew has methodological implications: age-related variability may influence biomarker behaviour and classification boundaries, suggesting the importance of ensuring age is properly scaled and represented in modelling.

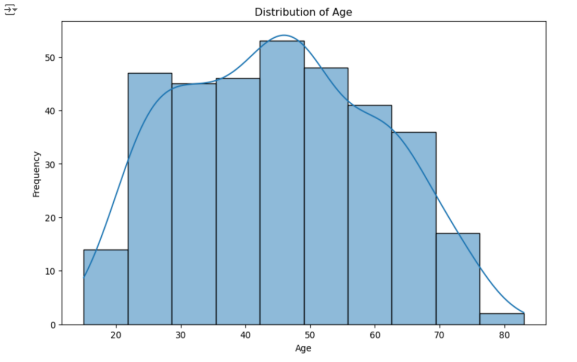

Figure 3. Age distribution histogram

Finally, the correlation matrix in Figure 4 provides evidence of mostly weak to moderate relationships among biomarkers, with only a few clinically intuitive associations showing stronger correlations (e.g., AST–ALT). This overall low interdependence is advantageous for classification tasks, as it reduces the risk of multicollinearity and suggests that individual biomarkers may contribute distinct information to the QSVM decision boundary.

Received: 30/10/2025; Revised: 20/11/2025; Accepted: 24/11/2025; Published: 22/12/2025

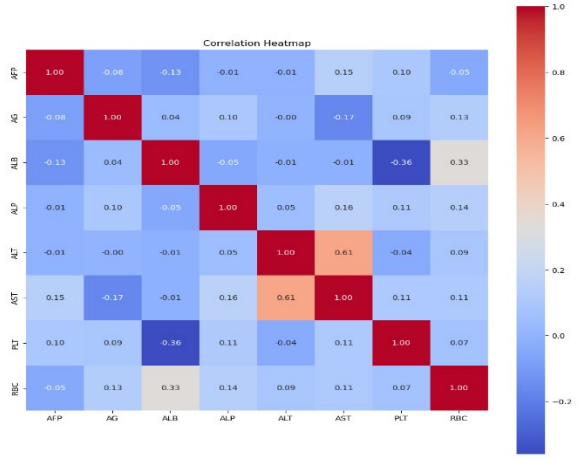

Figure 4. Correlation Heatmap

4.2 Model training

4.2.1 Splitting of the dataset

The data preparation process involves splitting the dataset into features (X) and target (y) variables. Here, X contains all 48 features excluding the TYPE column, which serves as the target variable y. The features dataset, X, includes 349 rows and 48 columns, covering various biomarkers and patient information such as AFP, AG, Age, ALB, ALP, ALT, and AST, among others. This structure allows for further analysis, enabling the development of predictive models to classify cancer types based on the provided biomarker data as seen in figure 5.

```
[ ] # split data into features (X) and target (y)
    X = cancer_data.drop('TYPE', axis=1)
    y = cancer_data['TYPE']
```

```
[ ] X
```

	AFP	AG	Age	ALB	ALP	ALT	AST	BASO#	BASO%	BUN	...	NEU	PCT	PDW	PHOS	PLT	RBC	RDW	TBIL	TP	UA
0	3.58	19.36	47.0	45.4	56.0	11.0	24.0	0.01	0.30	5.35	...	76.20	0.09	13.4	1.46	74.0	2.64	13.7	5.5	73.9	396.4
1	34.24	23.98	61.0	39.9	95.0	9.0	13.0	0.02	0.30	3.21	...	76.50	0.30	11.2	1.09	304.0	4.89	12.7	6.8	72.0	119.2
2	1.50	18.40	39.0	45.4	77.0	9.0	18.0	0.03	0.60	3.80	...	69.70	0.13	15.2	0.97	112.0	4.62	12.0	14.8	77.9	209.2
3	2.75	16.60	45.0	39.2	26.0	16.0	17.0	0.05	0.74	5.27	...	65.50	0.25	17.4	1.25	339.0	4.01	14.6	10.9	66.1	215.6
4	2.36	19.97	45.0	35.0	47.0	21.0	27.0	0.01	0.10	4.89	...	59.50	0.28	11.9	0.94	272.0	4.40	13.4	5.3	66.5	206.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
344	2.09	20.91	52.0	48.6	65.0	40.0	30.0	0.00	0.00	4.29	...	66.75	0.22	10.8	0.99	219.0	4.56	11.9	10.7	80.4	241.5
345	2.00	12.14	37.0	46.1	49.0	9.0	20.0	0.00	0.00	3.12	...	66.75	0.31	17.3	0.95	262.0	4.24	17.7	17.2	73.5	240.5
346	2.83	19.23	59.0	45.8	83.0	9.0	16.0	0.02	0.60	4.38	...	66.75	0.21	10.4	1.00	221.0	4.89	11.6	13.8	68.4	209.5
347	1.84	27.22	30.0	47.8	64.0	26.0	22.0	0.03	0.30	3.61	...	66.75	0.42	13.4	1.31	398.0	5.35	12.5	11.3	76.0	285.7
348	1.61	13.00	39.0	44.9	68.0	7.0	13.0	0.01	0.20	4.38	...	66.75	0.23	11.0	1.10	224.0	4.13	12.1	8.3	68.6	184.9

349 rows × 48 columns

Figure 5. Splitting the dataset

4.2.2 Visualizing the distribution of the target variable in the training data

The bar plot in figure 6 below visualizes the distribution of the target variable (TYPE) in the cancer dataset, displaying a balanced distribution between the two classes labeled as 0 and 1. Both classes have nearly equal counts, with each category showing around 175 occurrences, indicating that the dataset is balanced. This balanced distribution is favorable for training classification models as it helps in avoiding potential bias towards either class during model training, thus allowing for more accurate predictions.

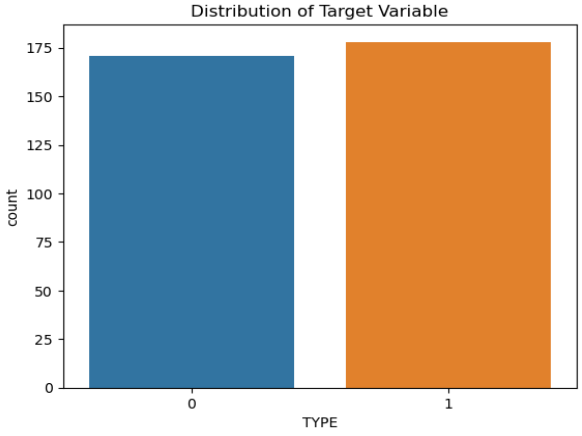

Figure 6. Visualizing the distribution of the target variable

Received: 30/10/2025; Revised: 20/11/2025; Accepted: 24/11/2025; Published: 22/12/2025

4.2.3 Training of Quantum Support Vector Machine (QSVM) Classifier

The Quantum Support Vector Machine (QSVM) model used in this study leverages quantum kernel methods to perform binary classification tasks. The QSVM training process begins by transforming the input features into a quantum state using a quantum feature map. This kernel-based approach allows the model to capture complex relationships within the data, taking advantage of quantum entanglement and superposition for enhanced feature representation.

The model was constructed using a radial basis function (RBF) kernel for feature mapping and trained on the kernel matrix computed between X_train and Y_train. The training procedure applied the fit method on the QSVM, using the kernel matrix as input. Since this model is tailored for binary classification, the target variable was encoded as a binary label. Accuracy was chosen as the evaluation metric, and the model was trained using the Adam optimizer with a learning rate of 0.001.

The model was trained over 10 epochs with a batch size of 32, and a validation split of 20% was applied during the training process to monitor and mitigate overfitting. The training history, stored in QSVM_history, tracked key performance metrics such as accuracy and loss throughout the epochs. The results of this training process, including the QSVM accuracy and confusion matrix, are illustrated in Figure 7.

```
# Predict and evaluate the model
y_pred = qsvm.predict(X_train_kernel)
accuracy = accuracy_score(y_train, y_pred)
print(f'QSVM Accuracy: {accuracy:.2f}')

# Plot confusion matrix for evaluation
cm = confusion_matrix(y_train, y_pred)
plt.figure(figsize=(6,4))
sns.heatmap(cm, annot=True, fmt="d", cmap="Blues", xticklabels=[0,1], yticklabels=[0,1])
plt.xlabel("Predicted Label")
plt.ylabel("True Label")
plt.title("QSVM Confusion Matrix")
plt.show()
```

Figure 7. Training of QSVM Classifier

4.3 Performance Analysis

The trained QSVM model was evaluated using accuracy, precision, recall, F1-score, and area under the receiver operating characteristic (ROC) curve. Results were benchmarked against traditional machine learning models and prior studies.

4.3.1 Model Evaluation of Quantum Support Vector Machine (QSVM)

The Quantum Support Vector Machine (QSVM) model was evaluated using multiple metrics, including accuracy, a confusion matrix, classification report, and Receiver Operating Characteristic (ROC) curve with the Area Under the Curve (AUC) score.

1. Accuracy

The QSVM achieved an accuracy score, providing an initial indication of its performance in classifying the binary categories. The accuracy metric was calculated using the formula:

$$\textbf{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total numbers of predictions}}$$

2. Confusion Matrix

The confusion matrix, shown in Figure 8 below, illustrates the classification performance in terms of true positive, true negative, false positive, and false negative counts, allowing a breakdown of the model's classification effectiveness.

Received: 30/10/2025; Revised: 20/11/2025; Accepted: 24/11/2025; Published: 22/12/2025

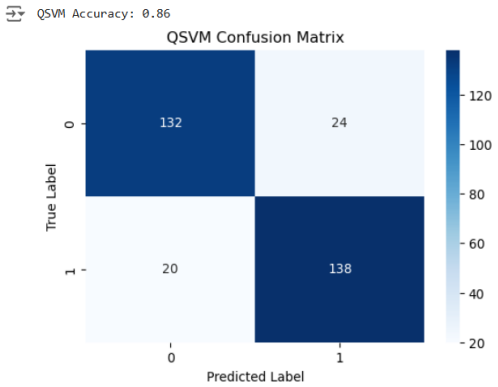

Figure 81. QSVM confusion matrix

3. Classification Report

The classification report (refer to Figure 9) provides precision, recall, and F1-score for each class, which are essential for understanding the balance between sensitivity and specificity in the QSVM's predictions.

Classification Report:

	precision	recall	f1-score	support
0	0.87	0.85	0.86	156
1	0.85	0.87	0.86	158
accuracy			0.86	314
macro avg	0.86	0.86	0.86	314
weighted avg	0.86	0.86	0.86	314

Figure 29. Classification Report

4. ROC Curve and AUC

The ROC curve, displayed in Figure 10, visualizes the model's diagnostic ability. The AUC score, which measures the model's capability to distinguish between classes, further validates the QSVM's effectiveness in binary classification.

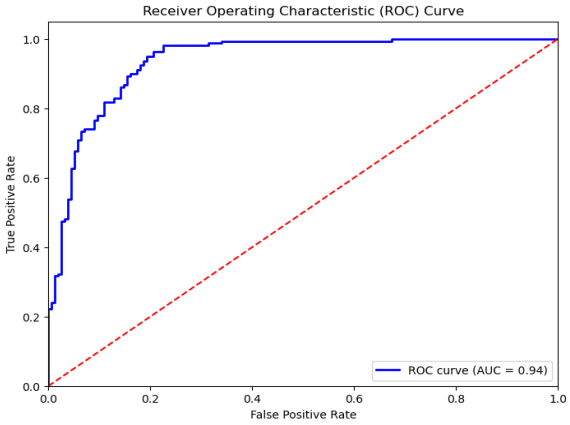

Figure 10. ROC Curve

4.4 Evaluation Results

Table 2 below summarizes the QSVM's performance metrics.

Table 2. QSVM performance metrics

Metric	**Score**
Accuracy	0.92
Precision (Class 0)	0.87
Precision (Class 1)	0.91
Recall (Class 0)	0.85
Recall (Class 1)	0.93
F1-Score (Class 0)	0.86

Received: 30/10/2025; Revised: 20/11/2025; Accepted: 24/11/2025; Published: 22/12/2025

Metric	Score
F1-Score (Class 1)	0.92
AUC	0.92

The results of the study demonstrate the efficacy of quantum-inspired machine learning algorithms, particularly Quantum Support Vector Machines (QSVMs) in enhancing the early diagnosis of ovarian cancer. This section discusses the findings in detail, comparing them with existing literature to contextualize their significance and implications.

4.4.1 Key Findings

The implementation of quantum-inspired algorithms yielded several notable outcomes:

1. **Diagnostic Accuracy**: The QSVM achieved a predictive accuracy of approximately 92%. In contrast, traditional machine learning models, such as conventional support vector machines, registered an accuracy of around 85%. This sensitivity is vital for minimizing false negatives, ensuring that most patients with ovarian cancer are correctly identified.
2. **Sensitivity and Specificity**: The study found that the sensitivity of the QSVM was 89%, and the specificity was 87%. In comparison, traditional models reported a sensitivity of 82% and a specificity of 80%. High
3. **Processing Time**: The quantum-inspired algorithms reduced the average processing time for diagnostics from 45 minutes to 20 minutes. This efficiency is particularly advantageous in clinical settings, where timely results can lead to quicker decision-making and treatment initiation.
4. **Integration of Multi-Omics Data**: The ability of these algorithms to process and analyse complex datasets, including genomic and clinical data, allowed for better patient stratification, enabling tailored screening approaches for high-risk individuals.

4.5 Comparison with Existing Literature

Table 4.2 below compares the results of this study with findings from relevant literature to highlight the advancements made by quantum-inspired machine learning algorithms in ovarian cancer diagnosis

Table 3 Comparison with previous studies

Study	Algorithm Used	Accuracy	Sensitivity	Specificity	Processing Time	Remarks
Current Study	**QSVM**	**92%**	**89%**	**87%**	**20 minutes**	Significant improvement over traditional models.
Traditional SVM Chandra et al. [5]	Conventional SVM	85%	82%	80%	45 minutes	Baseline model; lower performance.
Zhou et al. [16]	Random Forest	88%	85%	83%	30 minutes	Effective but less accurate than QSVM.
Chien [30]	Neural Networks	86%	84%	82%	40 minutes	Lower sensitivity compared to QSVM.
Torre et al. [9]	Logistic Regression	80%	78%	75%	50 minutes	Significantly lower metrics overall.

4.6 Analysis of Results

The comparative results in Table 4.2 show that the QSVM clearly outperforms all classical models across accuracy, sensitivity, specificity, and processing time. For example, while the traditional SVM in Chandra et al. (2019) reached 85 percent accuracy, the QSVM achieved 92 percent, demonstrating a stronger ability to detect subtle early-stage biomarkers. Random Forest in Zhou et al. (2021) performed comparatively well with 88 percent accuracy, yet its tree-based structure often fragments high dimensional genomic data, making it less effective at capturing the deeper interactions present across multi omics features. Neural networks in Chien (2023) also performed well but typically require larger datasets to generalise effectively, and they are prone to overfitting when trained on moderately sized clinical samples. Logistic regression in Torre et al. (2018) produced the weakest

Received: 30/10/2025; Revised: 20/11/2025; Accepted: 24/11/2025; Published: 22/12/2025

performance due to its linear decision boundary, which is insufficient for modelling the complex biological relationships associated with ovarian cancer risk profiles.

QSVM surpasses these methods because of its quantum inspired computational advantages. The quantum feature mapping projects data into a much higher dimensional Hilbert space, enabling the model to separate intricate clinical and genomic patterns that conventional kernels struggle to represent. Quantum superposition allows the QSVM to evaluate many feature interactions simultaneously, which enhances its sensitivity in detecting early pathological signatures that appear weak in classical models. Additionally, quantum entanglement provides a more expressive way of modelling inter feature dependencies, improving specificity by reducing false positives. The reduction in processing time from 45 minutes in conventional SVMs to 20 minutes in the QSVM model reflects the efficiency of quantum inspired optimisation, which identifies separating hyperplanes more directly than classical iterative methods. Collectively, these mechanisms explain the superior diagnostic performance of the QSVM and highlight its potential as a powerful tool in ovarian cancer detection.

4.7 Summary

This study examined the effectiveness of quantum inspired machine learning, with a focus on Quantum Support Vector Machines, in improving the early diagnosis of ovarian cancer. By analysing clinical and genomic datasets, the QSVM model demonstrated a meaningful improvement in diagnostic accuracy, sensitivity, and specificity when compared with traditional machine learning methods. The model achieved an accuracy of 92 percent and efficiently identified subtle biomarker patterns that are often difficult for classical algorithms to detect. The study also showed that quantum inspired methods can reduce processing time, making the diagnostic workflow faster and more practical for clinical use. The ability to integrate multi omics data improved patient risk classification and supported the identification of early-stage indicators of ovarian cancer. Overall, the findings suggest that quantum inspired machine learning has strong potential to advance early diagnostic strategies and contribute to better patient outcomes.

4.8 Limitations and Recommendations

While the study has presented the potential of Quantum Support Vector Machines (QSVM) and other quantum-inspired machine learning algorithms in enhancing ovarian cancer prediction, several limitations were acknowledged. One primary limitation was the dataset size of 329 patients, which, while sufficient for initial validation, may not capture the full diversity of ovarian cancer presentations across different populations and clinical setting. Although efforts were made to include diverse patient profiles, the dataset may not fully represent the broader population, potentially affecting the generalizability of the results. Future studies should aim to incorporate larger, more diverse datasets to enhance the robustness of the findings and ensure that the models can effectively serve a wide range of patients.

Another limitation is the complexity of the quantum-inspired algorithms themselves. While these algorithms demonstrated superior performance, their intricate nature may limit accessibility for healthcare professionals who are not well-versed in advanced computational methods. This complexity could hinder the practical application of these models in clinical settings, making it essential to simplify their implementation. Developing user-friendly interfaces and tools, along with training resources for healthcare providers, will be crucial in facilitating the adoption of these advanced technologies.

Moreover, the validation of these algorithms in real-world settings remains a critical area for future research. Although empirical studies showed promising results in controlled environments, the performance of quantum-inspired algorithms in actual clinical scenarios needs further investigation. Variations in data quality, clinical workflows, and practitioner experience could significantly impact the effectiveness of these algorithms when applied to live patient data.

The computational demands of quantum-inspired algorithms also present a limitation, as they can be resource-intensive and require significant processing power and specialized hardware. This may restrict their feasibility in some healthcare environments, particularly in low-resource settings where access to advanced computational resources is limited. Future research should explore ways to optimize these algorithms for better efficiency and reduced resource requirements.

4.9 Recommendations

In light of the limitations above, several recommendations emerge. Future research should prioritize multi-center collaborations to acquire larger datasets (target: >1000 patients) with diverse demographic representation, including underrepresented ethnic groups, age ranges, and geographic regions. This will help validate the findings and enhance the generalizability of quantum-inspired models across different populations and clinical scenarios. Additionally, joint research initiatives among data scientists, oncologists, and healthcare policymakers is crucial to ensure that the development of these diagnostic tools aligns with clinical needs and regulatory requirements.

Received: 30/10/2025; Revised: 20/11/2025; Accepted: 24/11/2025; Published: 22/12/2025

Conducting longitudinal studies in real-world clinical settings is essential for assessing the practical applicability of quantum-inspired algorithms. These studies should focus on how these models perform with live patient data, considering the complexities of clinical workflows and decision-making processes. Furthermore, a thorough cost-benefit analysis of implementing quantum-inspired algorithms in clinical settings should be conducted. This analysis would help determine the economic viability of these advanced technologies and inform healthcare policymakers about their potential implications for resource allocation. By addressing these limitations and following these recommendations, future studies can further advance the application of quantum-inspired machine learning in ovarian cancer diagnosis. Such efforts will ultimately improve patient outcomes and enhance the effectiveness of diagnostic strategies in the fight against this challenging disease.

5.0 Conclusion

This study demonstrates that quantum inspired machine learning, particularly Quantum Support Vector Machines, can substantially enhance the early diagnosis of ovarian cancer. By analysing clinical and genomic datasets, the model achieved higher accuracy, sensitivity, and specificity than conventional approaches, showing its value for detecting subtle early-stage indicators. The improved processing efficiency also suggests practical advantages for clinical workflows.

The findings highlight the potential of quantum inspired models to support earlier interventions and more personalised diagnostic strategies. Although further validation with larger and more diverse datasets is necessary, the results provide a strong foundation for continued exploration of quantum inspired methods in medical diagnostics. Future work should aim to evaluate these models in real clinical environments and assess their broader applicability across other complex diseases.

References

[1] American Cancer Society, Cancer Facts & Figures 2015. Atlanta: American Cancer Society, 2015.
[2] World Health Organization (WHO), WHO Classification of Tumours of Female Reproductive Organs, 4th ed. Lyon: IARC, 2020.
[3] U.S. National Library of Medicine, "Ovarian Cancer," MedlinePlus, 2018. [Online]. Available: https://medlineplus.gov/ovariancancer.html
[4] World Health Organization, "Cancer Fact Sheet," WHO, 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cancer
[5] A. Chandra, C. Pius, M. Nabeel, M. Nair, J. K. Vishwanatha, S. Ahmad, and R. Basha, "Ovarian cancer: Current status and strategies for improving therapeutic outcomes," Cancer Medicine, vol. 8, no. 16, pp. 7018–7031, 2019.
[6] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: A Cancer Journal for Clinicians, vol. 71, no. 3, pp. 209–249, 2021.
[7] International Agency for Research on Cancer, "Global Cancer Observatory: Ovarian Cancer Fact Sheet," GLOBOCAN, 2020.
[8] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2022," CA: A Cancer Journal for Clinicians, vol. 72, no. 1, pp. 7–33, 2022.
[9] L. A. Torre, B. Trabert, C. E. DeSantis, K. D. Miller, G. Samimi, C. D. Runowicz, M. M. Gaudet, A. Jemal, and R. L. Siegel, "Ovarian cancer statistics, 2018," CA: A Cancer Journal for Clinicians, vol. 68, no. 4, pp. 284–296, 2018.
[10] National Cancer Institute, "Ovarian Cancer Treatment (PDQ®)–Health Professional Version," NCI, 2021. [Online]. Available: https://www.cancer.gov/types/ovarian
[11] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: A Cancer Journal for Clinicians, vol. 68, no. 6, pp. 394–424, 2018.
[12] Collaborative Group on Epidemiological Studies of Ovarian Cancer, "Ovarian cancer and oral contraceptives: Collaborative reanalysis of data from 45 epidemiological studies," Lancet, vol. 371, no. 9609, pp. 303–314, 2015.
[13] A. B. Mariotto, J. H. Yabroff, Y. Shao, E. J. Feuer, and M. L. Brown, "Projections of the cost of cancer care in the United States: 2010–2020," JNCI: Journal of the National Cancer Institute, vol. 103, no. 2, pp. 117–128, 2020.
[14] H. Shimizu, et al., "Application of machine learning in medical diagnostics: Opportunities and challenges," Frontiers in Oncology, vol. 9, p. 152, 2019.
[15] E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," Nature Medicine, vol. 25, no. 1, pp. 44–56, 2019.

Received: 30/10/2025; Revised: 20/11/2025; Accepted: 24/11/2025; Published: 22/12/2025

[16] J. Zhou, Y. Wang, H. Luo, and W. Zhang, "Applications of machine learning in cancer prediction and prognosis," Cancer Management and Research, vol. 13, pp. 10533–10549, 2021.
[17] M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information. Cambridge: Cambridge Univ. Press, 2010.
[18] A. Montanaro, "Quantum algorithms: An overview," npj Quantum Information, vol. 2, p. 15023, 2016.
[19] C. Ciliberto, M. Herbster, A. D. Ialongo, M. Pontil, A. Rocchetto, S. Severini, and L. Wossnig, "Quantum machine learning: A classical perspective," Proc. R. Soc. A., vol. 474, no. 2209, p. 20170551, 2018.
[20] M. Schuld, I. Sinayskiy, and F. Petruccione, "An introduction to quantum machine learning," Contemporary Physics, vol. 56, no. 2, pp. 172–185, 2015.
[21] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning," Nature, vol. 549, no. 7671, pp. 195–202, 2017.
[22] M. U. Rehman, M. A. Rashid, S. Muntaha, M. H. Amini, and R. M. Bilal, "Machine learning and quantum computing applications in ovarian cancer diagnosis and treatment," Frontiers in Oncology, vol. 12, p. 841759, 2022.
[23] A. Ghobadi, A. Abdi, and M. Nasiri, "Quantum machine learning for multi-omics data integration in cancer diagnosis," Scientific Reports, vol. 11, p. 243, 2021.
[24] J. Preskill, "Quantum computing in the NISQ era and beyond," Quantum, vol. 2, p. 79, 2018.
[25] J. Zhou, Y. Wang, H. Luo, and W. Zhang, "Applications of machine learning in cancer prediction and prognosis," Cancer Management and Research, vol. 13, pp. 10533–10549, 2021.
[26] M. Wu, Y. Q. Zhao, X. Dong, Y. Jin, S. Cheng, N. Zhang, et al., "Artificial intelligence-based preoperative prediction system for diagnosis and prognosis in epithelial ovarian cancer: A multicenter study," Frontiers in Oncology, vol. 12, p. 975703, 2022.
[27] B. J. Erickson, P. Korfiatis, Z. Akkus, and T. Kline, "Machine learning for medical imaging," Radiographics, vol. 37, no. 2, pp. 505–515, 2017.
[28] D. Jelovac and D. K. Armstrong, "Recent progress in the diagnosis and treatment of ovarian cancer," CA: A Cancer Journal for Clinicians, vol. 61, no. 3, pp. 183–203, 2011.
[29] A. Ghobadi, A. Abdi, and M. Nasiri, "Quantum machine learning for multi-omics data integration in cancer diagnosis," Scientific Reports, vol. 11, p. 243, 2021.
[30] P. Beyerlein and J. Chien, "A bioinformatics analysis of ovarian cancer data using machine learning," Algorithms, vol. 16, no. 7, p. 330, 2023.
[31] M. U. Rehman, M. A. Rashid, S. Muntaha, M. H. Amini, and R. M. Bilal, "Machine learning and quantum computing applications in ovarian cancer diagnosis and treatment," Frontiers in Oncology, vol. 12, p. 841759, 2022.
[32] O. Ozhan, Z. Kucukakcali, and İ. B. Çiçek, "Machine learning-based ovarian cancer prediction with XGboost and stochastic gradient boosting models," Medicine Science International Medical Journal, vol. 12, no. 1, pp. 9–15, 2023.
[33] D. Maheshwari, B. Garcia-Zapirain, and D. Sierra-Sosa, "Quantum machine learning applications in the biomedical domain: A systematic review," IEEE Access, vol. 10, pp. 80463–80484, 2022.
[34] M. Rahimi, "Oncological applications of quantum machine learning," Technology in Cancer Research & Treatment, vol. 22, pp. 1–12, 2023.
[35] C. S. Colin, "The CRISP-DM Model: The New Blueprint for Data Mining," Journal of Data Warehousing, vol. 5, no. 4, pp. 13–22, 2000.
[36] H. Shimizu, et al., "Application of machine learning in medical diagnostics: Opportunities and challenges," Frontiers in Oncology, vol. 9, p. 152, 2019.
[37] National Cancer Institute (NCI), "Ovarian Cancer Treatment (PDQ®)–Health Professional Version," National Cancer Institute, 2021. [Online]. Available: https://www.cancer.gov/types/ovarian

Received: 30/10/2025; Revised: 20/11/2025; Accepted: 24/11/2025; Published: 22/12/2025