

Comparative Performance Analysis of SVM and QSVM in Breast Cancer Diagnosis

Gita Kartika^{1*}, Muhamad Akrom^{2**}

* Teknik Informatika, Universitas Dian Nuswantoro

** Research Center for Quantum Computing and Materials Informatics, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro
111202113812@mhs.dinus.ac.id¹, m.akrom@dsn.dinus.ac.id²

Article Info

Article history:

Received 2026-09-04

Revised 2026-09-25

Accepted 2026-10-18

Keyword:

Support Vector Machine (SVM),
Quantum Support Vector Machine (QSVM),
Breast Cancer.

ABSTRACT

The application of machine learning has shown significant progress in various fields, including clinical data classification. This study presents a comparative analysis between the classical Support Vector Machine (SVM) and Quantum Support Vector Machine (QSVM) for breast cancer classification using the Wisconsin Diagnostic Breast Cancer (WDBC) dataset. Principal Component Analysis (PCA) was applied as dimension reduction strategy across three configurations (PCA = 2, PCA = 4, and PCA = 6) to evaluate the sensitivity of both models to the feature space dimension. Preprocessing included log1p transformation on 22 skewed features, Manual SMOTE balancing, and MinMaxScaler normalization. SVM was optimized via GridSearchCV (80 parameter combinations, cv = 5), while QSVM employed a ZZFeatureMap quantum kernel (reps = 2, entanglement = full) on qasm_simulator. Experimental results demonstrate that SVM achieved superior accuracy across all configurations, with peak performance at PCA = 6 (Accuracy= 98.83%, AUC = 0.9985, Specificity = 1.0000, Sensitivity = 0.9688, Precision = 1.0000, NPV = 0.9817, and F1-score for malignant = 0.98). QSVM achieved its highest accuracy at PCA = 4 (92.98%), AUC = 0.9622, Sensitivity = 0.9219, Precision = 0.8939, NPV = 0.9524, and F1-score for malignant = 0.91. Across the three PCA configurations, QSVM exhibited a non-monotonic relationship between PCA dimensions and quantum kernel performance. These findings provide a quantitative benchmark for comparing quantum and classical machine learning in breast cancer diagnosis.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Perkembangan pembelajaran mesin (*machine learning*) telah mendorong kemajuan signifikan di berbagai bidang, termasuk di bidang kesehatan, khususnya pada sistem diagnosis berbasis data. Dalam konteks medis, data yang digunakan umumnya bersifat berdimensi tinggi, memiliki korelasi antar fitur yang kompleks, dan sering kali mengandung pola *non-linear* yang menantang bagi proses pemodelan [1]. Kondisi ini menjadikan klasifikasi data medis memerlukan pendekatan analisis yang cermat dan metodologi evaluasi yang tepat.

Oleh karena itu, fokus utama dalam pengembangan sistem diagnosis pembelajaran mesin tidak semata-mata terletak pada pencapaian akurasi tinggi, melainkan pada evaluasi kemampuan model dalam melakukan generalisasi secara

konsisten terhadap data. Pemilihan algoritma klasifikasi dalam penelitian medis perlu mempertimbangkan keseimbangan antara kompleksitas model, stabilitas kinerja, serta relevansi hasil terhadap konteks aplikasi yang dituju.

Salah satu permasalahan medis yang banyak diteliti adalah diagnosis kanker payudara. Kanker payudara merupakan salah satu jenis kanker dengan insidensi tertinggi dan menjadi penyebab utama kematian pada perempuan di berbagai negara [2]. Diagnosis dini dan akurat sangat berperan dalam meningkatkan peluang keberhasilan pengobatan dan menurunkan tingkat mortalitas. Oleh sebab itu, pengembangan model klasifikasi yang mampu membedakan secara akurat antara tumor jinak (*benign*) dan ganas (*malignant*) menjadi prioritas utama dalam penelitian di bidang informatika medis.

Support Vector Machine (SVM) merupakan salah satu algoritma pembelajaran mesin klasik yang banyak digunakan dalam klasifikasi data medis karena stabilitas, ketahanan terhadap *overfitting*, serta kemampuannya dalam menangani data berdimensi tinggi. SVM bekerja dengan menentukan *hyperplane* optimal yang memisahkan dua kelas data dengan margin maksimum [3]. Untuk data yang tidak dapat dipisahkan secara linear, SVM menggunakan fungsi kernel. Kernel yang umum diterapkan dalam klasifikasi antara lain seperti Radial Basis Function (RBF) dan kernel polynomial (Poly), yang masing-masing memiliki karakteristik pemetaan ruang fitur berbeda. Berbagai penelitian sebelumnya menunjukkan bahwa SVM mampu memberikan performa yang kompetitif dan konsisten dalam tugas diagnosis kanker payudara [4], [5], [6], [7].

Di sisi lain, seiring dengan berkembangnya komputasi kuantum, muncul paradigma baru yang dikenal sebagai Quantum Machine Learning (QML). QML merupakan pendekatan yang mengintegrasikan prinsip-prinsip mekanika kuantum ke dalam algoritma pembelajaran mesin [8]. QML memanfaatkan fenomena kuantum seperti superposisi dan *entanglement* untuk memperluas ruang representasi data, sehingga berpotensi meningkatkan kemampuan model dalam menangkap hubungan kompleks antar fitur [9]. Salah satu algoritma QML yang mendapatkan perhatian luas adalah Quantum Support Vector Machine (QSVM), yang merupakan adaptasi dari SVM dengan memanfaatkan kernel kuantum sebagai ukuran kesamaan antar data dalam ruang Hilbert berdimensi tinggi [10].

Dalam QSVM, data klasik dipetakan ke dalam keadaan kuantum melalui *quantum feature map*, salah satunya adalah ZZFeatureMap, yang menggabungkan rotasi kuantum dan interaksi antar qubit untuk menangkap korelasi fitur yang kompleks [11]. Kesamaan antar data kemudian dihitung menggunakan quantum kernel, dan proses klasifikasi dilakukan dengan pendekatan yang serupa dengan SVM klasik. Secara teoretis, pendekatan ini memberikan potensi keunggulan dalam merepresentasikan data *non-linear* tanpa memerlukan transformasi kernel eksplisit seperti pada metode klasik.

Dalam beberapa tahun terakhir, sejumlah penelitian telah mengkaji penerapan QSVM pada berbagai permasalahan klasifikasi, termasuk diagnosis kanker payudara menggunakan dataset Wisconsin Diagnostic Breast Cancer Dataset (WDBC). Penelitian-penelitian tersebut, termasuk penelitian ini, pada umumnya dijalankan dalam lingkungan simulasi kuantum akibat masih terbatasnya akses terhadap perangkat keras kuantum saat ini [7], [12], [13], [14]. Oleh karena itu, hasil yang diperoleh merepresentasikan kinerja QSVM dalam konteks simulasi dan bersifat evaluatif, sehingga interpretasinya perlu mempertimbangkan keterbatasan lingkungan eksperimen serta karakteristik dataset yang digunakan.

Meskipun perbandingan antara QSVM dan SVM pada dataset WDBC telah banyak dilaporkan dalam penelitian sebelumnya, setiap studi umumnya mengadopsi rancangan metodologi yang disesuaikan dengan tujuan dan batasan masing-masing penelitian [7], [12], [13], [14]. Perbedaan

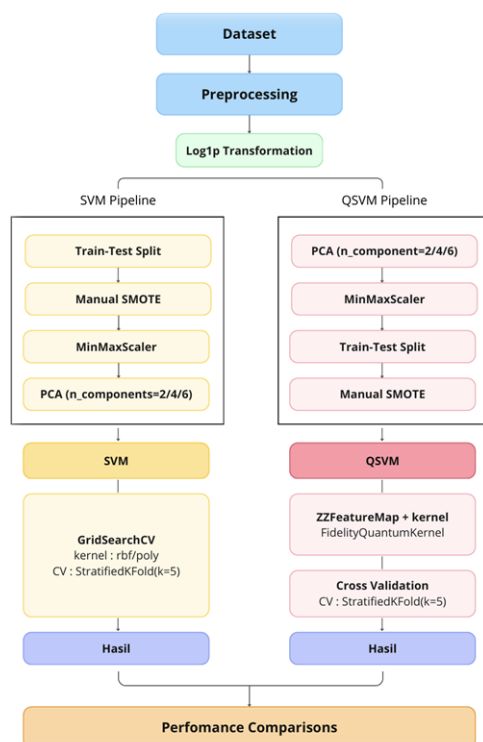
tersebut dapat mencakup tahapan *preprocessing*, pemilihan dan konfigurasi *quantum feature map*, serta skema evaluasi kinerja yang digunakan. Variasi metodologis ini menunjukkan bahwa masih terdapat ruang untuk kajian komparatif yang lebih sistematis antara SVM dan QSVM, khususnya yang mempertimbangkan tahapan *preprocessing*, konfigurasi *feature map*, dan skema evaluasi secara terpadu.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan mengkaji kinerja SVM dan QSVM dalam klasifikasi diagnosis kanker payudara menggunakan dataset WDBC. Principal Component Analysis (PCA) diintegrasikan sebagai tahap *preprocessing* untuk mereduksi dimensi data dan meningkatkan stabilitas pemodelan, sekaligus menyesuaikan representasi data dengan keterbatasan jumlah qubit pada lingkungan simulasi kuantum. Kinerja kedua pendekatan dianalisis secara kuantitatif menggunakan metrik accuracy, precision, recall, F1-score, Area Under Curve (AUC), specificity, dan NPV dengan tiga konfigurasi PCA yang diuji ($n=2, 4, 6$).

Penelitian ini tidak dimaksudkan untuk mengklaim keunggulan absolut QSVM terhadap metode klasik, melainkan untuk mengevaluasi secara empiris sejauh mana pendekatan berbasis quantum kernel mampu mengimbangi kinerja SVM klasik pada skenario data medis berdimensi tinggi. Hasil penelitian ini diharapkan dapat memberikan kontribusi berupa pemahaman yang lebih realistis mengenai potensi dan keterbatasan QSVM dalam konteks diagnosis medis, serta menjadi landasan awal bagi penelitian lanjutan terkait pengembangan dan evaluasi algoritma pembelajaran mesin berbasis komputasi kuantum di bidang kesehatan.

II. METODE

Penelitian ini menggunakan pendekatan eksperimental kuantitatif, meliputi *preprocessing* data, pelatihan model SVM dan QSVM, serta evaluasi kinerja pada klasifikasi diagnosis kanker payudara. Kerangka metodologi diilustrasikan pada Gambar 1.



Gambar 1. Kerangka Metodologi

A. Dataset

Data penelitian bersumber dari Wisconsin Diagnostic Breast Cancer (WDBC) yang disediakan UCI Machine Learning Repository [15]. Dataset ini berisi 569 sampel dengan 30 fitur numerik hasil ekstraksi citra jaringan kanker

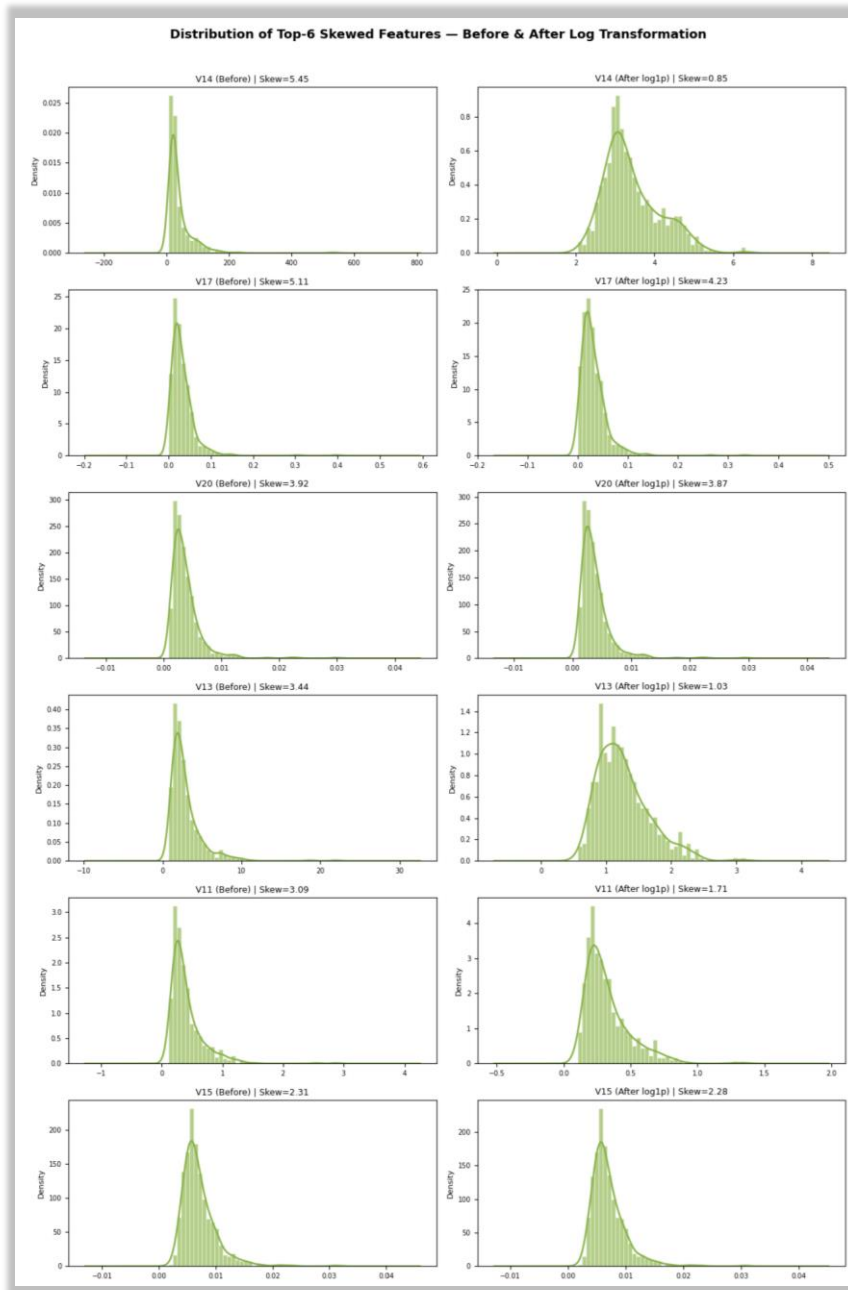
payudara, meliputi *mean*, *standard error*, dan *worst value* dari 10 atribut sel (*radius*, *texture*, *perimeter*, *area*, *smoothness*, *compactness*, *concavity*, *concave points*, *symmetry*, *fractal dimension*). Setiap sampel berlabel *Malignant* (ganas) sebanyak 212 sampel dan *Benign* (jinak) sebanyak 357 sampel, dengan rasio imbalance 1.68:1.

Pemilihan dataset WDBC didasarkan pada ketersediaan data yang terstandar dan luasnya penggunaan dataset ini dalam penelitian klasifikasi kanker payudara, sehingga memungkinkan perbandingan hasil secara konsisten dengan studi-studi sebelumnya. Namun demikian, penggunaan satu dataset dalam penelitian ini membatasi generalisasi temuan ke dataset atau populasi lain, sehingga hasil penelitian ini bersifat kontekstual terhadap karakteristik WDBC.

B. Preprocessing

Tahap *preprocessing* dirancang untuk menangani karakteristik dataset WDBC sebelum digunakan pada model SVM dan QSVM. Masing-masing tahapan diuraikan sebagai berikut:

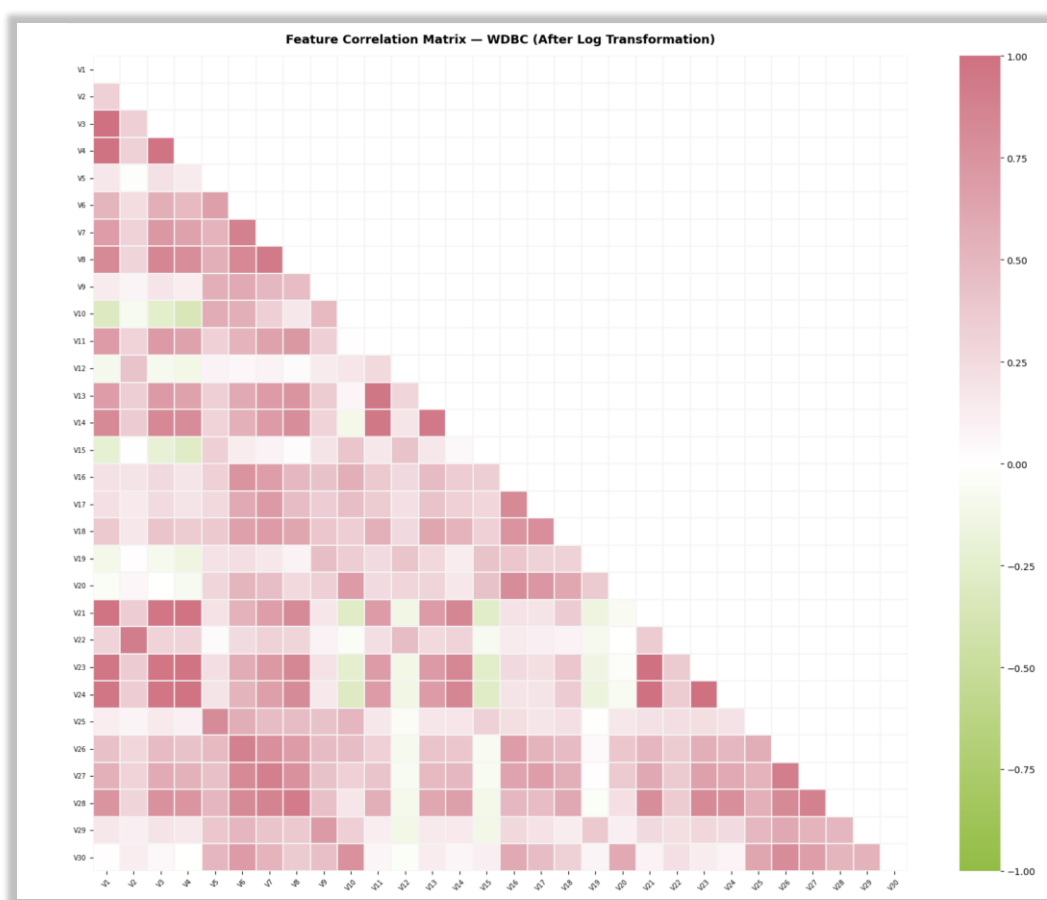
1) *Analisis Skewness dan Log1p Transformation*: Distribusi setiap fitur dianalisis menggunakan koefisien skewness Fisher. Sebanyak 22 dari 30 fitur menunjukkan nilai $|\text{skewness}| > 1$, yang mengindikasikan distribusi yang menyimpang signifikan dari normal. Distorsi ini dapat mempengaruhi kinerja fungsi kernel. Transformasi $\log_{1p}$, $f(x) = \log(1+x)$, diterapkan pada seluruh 22 fitur tersebut. Fungsi ini dipilih karena mampu menangani nilai nol ($\log(1+0) = 0$) yang terdapat pada sebagian fitur dataset. Nilai inf/NaN residual diatasi dengan imputasi median kolom. Distribusi sebelum dan sesudah transformasi ditampilkan pada Gambar 2.



Gambar 2. Distribusi 6 fitur skewness tertinggi sebelum dan sesudah log1p transformation

Analisis korelasi antar 30 fitur setelah transformasi ditampilkan pada Gambar 3. Heatmap mengungkapkan korelasi positif sangat tinggi (>0.9) pada kelompok fitur geometri sel (radius, perimeter, area) pada ketiga level pengukuran (mean, SE, worst), yang memperkuat justifikasi

penggunaan PCA sebagai tahap reduksi dimensi. Korelasi negatif moderat juga teridentifikasi antara fitur fractal_dimension dengan kelompok fitur geometri yang mengindikasikan bahwa fitur dimensi fraktal memiliki pola yang berlawanan dengan fitur ukuran sel secara umum.



Gambar 3. Heatmap korelasi 30 fitur setelah log1p transformation

TABEL I
PEMETAAN LABEL V1-V30

Kelompok	Kode	Atribut
Mean	V1-V10	radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, fractal_dimension
Standard Error (SE)	V11-V20	
Worst	V21-V30	

2) *Stratified Train-Test Split*: Dataset dibagi dengan 70:30 menggunakan stratified sampling (random_state=42) untuk memastikan proporsi kelas *benign* dan *malignant* terjaga secara proporsional di kedua subset. Pada pipeline SVM, *splitting* dilakukan sebelum SMOTE dan *scaling* (MinMaxScaler). Sedangkan pada pipeline QSVM, *splitting* dilakukan setelah PCA dan *scaling*.

3) *Manual SMOTE*: Manual SMOTE (Synthetic Minority Over-sampling Technique) diterapkan hanya pada data training untuk menyeimbangkan distribusi kelas tanpa mencemari evaluasi. Fungsi manual_smote() menggunakan k = 5 nearest neighbors untuk menginterpolasi sampel sintetis

kelas minoritas (*malignant*) hingga jumlahnya seimbang dengan kelas mayoritas (*benign*). Pada pipeline SVM, SMOTE diterapkan setelah *splitting*. Pada pipeline QSVM, SMOTE diterapkan setelah *splitting* yang dilakukan pada data yang telah melalui PCA dan *scaling*.

4) *MinMaxScaler*: MinMaxScaler digunakan untuk memetakan seluruh fitur ke dalam rentang [0, 1] berdasarkan formula:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

Normalisasi ke rentang $[0, 1]$ dipilih karena ZZFeatureMap pada QSVM mengkodekan nilai fitur sebagai parameter rotasi sudut pada sirkuit kuantum, sehingga magnitude fitur secara langsung mempengaruhi struktur *quantum state* yang dihasilkan. Rentang $[0, 1]$ memastikan parameter rotasi berada dalam domain yang stabil.

Pada pipeline SVM, MinMaxScaler di-fit pada data training setelah SMOTE untuk mencegah *data leakage*. Parameter x_{min} dan x_{max} diperoleh hanya dari distribusi *training set*, kemudian diterapkan (transform) pada *test set*.

Pada pipeline QSVM, MinMaxScaler diterapkan secara global pada seluruh data setelah PCA dan sebelum proses *splitting*. PCA global dijustifikasi oleh kebutuhan untuk memperoleh representasi komponen principal yang stabil sebelum encoding ke sirkuit kuantum, mengingat keterbatasan jumlah qubit pada lingkungan simulasi NISQ (Noisy Intermediate-Scale Quantum). Selanjutnya, MinMaxScaler di-fit pada seluruh 569 sampel hasil PCA, di mana parameter x_{min} dan x_{max} dihitung dari keseluruhan data sehingga setiap data point, baik yang nanti menjadi training maupun testing, dijamin berada dalam rentang $[0, 1]$ yang persis sama. Hal ini krusial karena ZZFeatureMap mengkodekan nilai fitur sebagai rotasi sudut pada gate kuantum. Nilai 0,0 menghasilkan rotasi 0; nilai 0,5 menghasilkan rotasi $0,5\pi$; dan nilai 1,0 menghasilkan rotasi π . Dengan *scaling* global, seluruh rotasi sudut berada dalam domain $[0, \pi]$ yang konsisten. Sebaliknya, jika MinMaxScaler di-fit hanya pada *training set*, *test set* berpotensi menghasilkan nilai di luar $[0, 1]$; hal ini menyebabkan rotasi kuantum melampaui batas $[0, \pi]$ dan membuat kernel fidelity antara training dan testing tidak sebanding. Berbeda dengan SVM klasik yang masih toleran terhadap perbedaan skala karena menggunakan kernel berbasis jarak Euclidean yang

proporsional dan dapat dikompensasi oleh hyperparameter (C, gamma), QSVM jauh lebih rentan karena rotasi sudut bersifat periodik, sehingga ketidakkonsistenan skala dapat merusak perhitungan kernel quantum. Meskipun pendekatan global *scaling* ini berpotensi menimbulkan *data leakage* yang sangat kecil (hanya berupa informasi statistik min/max dari *test set*), dampaknya terhadap generalisasi model dinilai minimal. Dalam konteks NISQ, stabilitas *quantum kernel* dianggap prioritas dibandingkan kepatuhan ketat terhadap konvensi no-leakage. Selain itu, dengan jumlah sampel terbatas (569 sampel), penggunaan seluruh data untuk menentukan rentang *scaling* menghasilkan parameter yang lebih stabil dan representatif.

5) *Reduksi Dimensi dengan PCA*: Principal Component Analysis (PCA) diterapkan untuk menyederhanakan kompleksitas data [16] dengan memproyeksikan fitur asli ke ruang berdimensi baru. PCA diuji dengan tiga konfigurasi jumlah komponen utama, yaitu $n_{components} = 2, 4, \text{ dan } 6$.

Pendekatan PCA, dipilih karena ZZFeatureMap pada QSVM memiliki keterbatasan teknis pada lingkungan simulasi NISQ, sehingga data berdimensi tinggi perlu direduksi terlebih dahulu. Reduksi menjadi jumlah komponen tersebut memudahkan penanganan oleh QSVM sekaligus menjaga sebagian besar informasi data. Selain itu, PCA membantu mengatasi *curse of dimensionality* dan mengurangi beban komputasi pada simulasi quantum.

Penerapan PCA pada pipeline SVM dilakukan setelah *scaling*. Sedangkan pada pipeline QSVM, PCA di-fit sebelum *scaling* dan *splitting*. Perbedaan urutan ini menghasilkan distribusi varian PCA yang berbeda secara signifikan, sebagaimana ditunjukkan pada Tabel II.

TABEL II
PERSENTASE VARIAN KUMULATIF PCA PER KONFIGURASI

Konfigurasi	PC1	PC2	PC3	PC4	PC5	PC6	Total
SVM PCA=2	55.97%	16.57%	-	-	-	-	72.54%
SVM PCA=4	55.97%	16.57%	7.67%	5.71%	-	-	85.92%
SVM PCA=6	55.97%	16.57%	7.67%	5.71%	3.84%	2.66%	92.42%
QSVM PCA=2	92.32%	7.29%	-	-	-	-	99.61%
QSVM PCA=4	92.32%	7.29%	0.33%	0.04%	-	-	99.98%
Q SVM PCA=6	92.32%	7.29%	0.33%	0.04%	0.01%	0.01%	100.00%

Pada pipeline SVM, PCA diterapkan setelah MinMaxScaler sehingga seluruh fitur telah dinormalisasi ke rentang $[0, 1]$. Normalisasi menyeimbangkan kontribusi tiap fitur terhadap matriks kovarians, sehingga menghasilkan distribusi varian yang lebih merata. PC1 hanya menjelaskan 55.97% varian dan diperlukan 6 komponen untuk mencapai 92.42%. Pada pipeline QSVM, PCA diterapkan pada data sebelum *scaling* sehingga fitur dengan skala absolut besar mendominasi matriks kovarians. Akibatnya, PC1

mendominasi 92.32% varian dan hanya 4 komponen sudah mencukupi 99.98% varian. Perbedaan distribusi varian ini secara langsung mempengaruhi kualitas representasi data yang masuk ke masing-masing model dan menjadi salah satu faktor pembeda perilaku kedua model terhadap konfigurasi PCA.

C. Support Vector Machine (SVM)

Model Support Vector Machine (SVM) merupakan salah satu algoritma supervised learning yang banyak digunakan

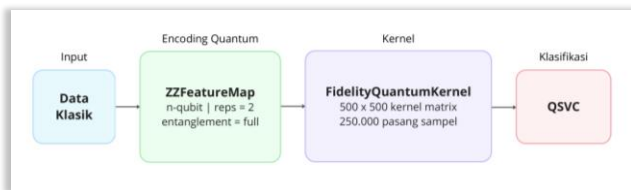
untuk tugas klasifikasi. SVM bekerja dengan menentukan *hyperplane* optimal yang memisahkan dua kelas dengan margin maksimum [3]. Untuk data yang tidak bisa dipisahkan secara linear, model menggunakan fungsi kernel yang menghitung *inner product* dalam ruang fitur berdimensi tinggi tanpa transformasi eksplisit, sehingga pemisahan linear dapat dilakukan secara efisien. Kernel yang diuji dalam penelitian ini adalah *Radial Basis Function (RBF)* dan *polynomial (poly)*, yang masing-masing memiliki karakteristik pemetaan ruang fitur yang berbeda.

Parameter C mengontrol *trade-off* antara margin yang lebar dan toleransi kesalahan klasifikasi, sedangkan parameter γ mengontrol lebar fungsi kernel RBF. Nilai γ yang besar membuat model lebih sensitif terhadap pola lokal, sedangkan nilai yang kecil menghasilkan pengaruh yang lebih global. Parameter degree berlaku khusus untuk kernel polynomial.

SVM dioptimalkan menggunakan GridSearchCV dengan 80 kombinasi parameter: $C \in \{0.01, 0.1, 1, 10, 100\}$, $kernel \in \{rbf, poly\}$, $\gamma \in \{scale, auto, 0.01, 0.1\}$, $degree \in \{2, 3\}$. Validasi silang menggunakan StratifiedKFold ($k=5$) dengan scoring accuracy pada data yang telah diseimbangkan SMOTE. Dari 80 kombinasi yang dievaluasi, parameter optimal yang diperoleh adalah $kernel = rbf$, $C = 100$, $\gamma = 0.1$, dengan best CV accuracy = 0.9680. Parameter class_weight tidak digunakan karena data training sudah seimbang. Setelah diperoleh model terbaik, dilakukan validasi silang tambahan menggunakan StratifiedKFold ($k = 5$) pada data training untuk mengevaluasi konsistensi generalisasi best_svm.

D. Quantum Support Vector Machine (QSVM)

Quantum Support Vector Machine (QSVM) diimplementasikan menggunakan framework Qiskit Machine Learning dengan pendekatan *quantum kernel* berbasis ZZFeatureMap. Seluruh proses QSVM dilakukan pada lingkungan simulasi kuantum ideal menggunakan qasm_simulator (Qiskit Aer), tanpa melibatkan perangkat keras kuantum nyata. Mekanisme kerja QSVM ditunjukkan pada Gambar 4.

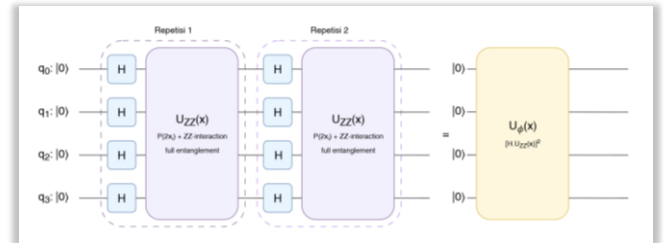


Gambar 4. Mekanisme kerja QSVM

1) **Quantum Feature Map (ZZFeatureMap):** Setiap vektor fitur hasil reduksi PCA, $x \in \mathbb{R}^n$, dipetakan ke dalam keadaan kuantum melalui ZZFeatureMap ($reps = 2$, $entanglement = full$).

$$|\psi(x)\rangle = U_\phi(x)|0\rangle^{\otimes n} \quad (2)$$

ZZFeatureMap $U_\phi(x)$ terdiri dari dua layer yang diulang sebanyak $reps = 2$. Konfigurasi ini dipilih untuk meningkatkan ekspresivitas pemetaan fitur kuantum sekaligus menjaga kedalaman sirkuit tetap moderat agar sesuai keterbatasan lingkungan simulasi NISQ. Circuit view ZZFeatureMap ditunjukkan pada Gambar 5.



Gambar 5. Tampilan sirkuit ZZFeatureMap

Secara matematis, operator $U_\phi(x)$ pada konfigurasi ini didefinisikan sebagai:

$$U_\phi(x) = [H \cdot U_{ZZ}(x)]^2 \quad (3)$$

$U_{ZZ}(x)$ terdiri dari dua sub-layer. Sub-layer pertama adalah rotasi satu qubit yang dinyatakan sebagai $P(2x_i)$, di mana $P(2x_i) \equiv R_z(2x_i)$. Gerbang PhaseGate $P(2x_i)$ ini mengkodekan nilai fitur x_i sebagai rotasi fase sebesar $2x_i$ radian pada qubit ke- i . Sub-layer kedua adalah ZZ entanglement antar pasangan qubit. Untuk setiap pasangan qubit (i, j) , entanglement ini dinyatakan sebagai $U_{ZZ}(x_i, x_j) = \exp(i \cdot 2(\pi - x_i)(\pi - x_j) \cdot Z \otimes Z)$, yang secara ekuivalen dapat diimplementasikan melalui rangkaian $CNOT \cdot P(2(\pi - x_i)(\pi - x_j)) \cdot CNOT$. Mekanisme ini mengkodekan korelasi antar fitur x_i dan x_j ke dalam state kuantum.

2) **Quantum Kernel (FidelityQuantumKernel):** Kesamaan antara dua sampel x_i dan x_j dihitung menggunakan *quantum kernel* berbasis fidelitas [17].

$$K(x_i, x_j) = |\langle \psi(x_i) | \psi(x_j) \rangle|^2 \quad (4)$$

Kernel ini digunakan sebagai masukan pada *Quantum Support Vector Classifier (QSVC)*.

3) **Pelatihan:** Sebelum pelatihan akhir, dilakukan validasi silang menggunakan StratifiedKFold ($k = 5$, $shuffle = True$, $random_state = 42$) pada data training. Inferensi dilakukan melalui `qsvc.predict()` untuk prediksi kelas dan `qsvc.decision_function()` sebagai skor diskriminasi kontinu untuk perhitungan AUC-ROC, mengingat QSVC tidak mendukung `predict_proba()`.

E. Evaluasi

Kinerja kedua model dievaluasi menggunakan metrik accuracy, precision, recall, F1-score, Area Under Curve (AUC).

1) *Accuracy*: Accuracy mengukur proporsi prediksi yang benar dari seluruh sampel [18]. Metrik ini paling informatif ketika distribusi kelas seimbang. Pada penelitian ini, distribusi kelas pada *test set* tidak sepenuhnya seimbang (107 *benign* dan 64 *malignant*), sehingga Accuracy diinterpretasikan bersama metrik lainnya.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

2) *Precision*: Precision mengukur prediksi positif yang benar dari seluruh prediksi positif yang dibuat model [19]. Precision tinggi mengindikasikan rendahnya false positive [20], sehingga hanya sedikit pasien *benign* yang salah didiagnosis sebagai *malignant* dalam konteks klinis.

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

3) *Recall (Sensitivity)*: Recall menunjukkan kemampuan model untuk mendeteksi seluruh data positif secara benar [21]. Dalam konteks diagnosis kanker Recall mengukur kemampuan model mendeteksi kasus *malignant*, di mana nilai yang tinggi menunjukkan semakin sedikit kasus kanker yang terlewat (false negative).

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

4) *F1-score*: F1-score merupakan rata-rata harmonik antara precision dan recall [7], sehingga memberikan ukuran keseimbangan antara keduanya.

$$F_1\text{-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

5) *AUC-ROC (Area Under the Receiver Operating Characteristics Curve)*: mengukur kemampuan model untuk membedakan antara kelas positif dan negatif di seluruh rentang *threshold* klasifikasi [22]. ROC curve diplot dengan True Positive Rate (TPR = Recall) pada sumbu y dan False Positive Rate (FPR = 1 – Specificity) pada sumbu x [23]. Nilai AUC berkisar antara 0.5 (*random classifier*) hingga 1.0 (*perfect classifier*). AUC ≥ 0.9 mengindikasikan kemampuan diskriminasi yang sangat baik [24].

6) *Specificity*: mengukur proporsi sampel negatif actual yang berhasil diidentifikasi dengan benar sebagai negatif oleh

model [25]. Dalam penelitian ini, Specificity mengukur kemampuan model mengidentifikasi kasus *benign* secara benar, di mana nilai yang tinggi menunjukkan semakin sedikit false positive (kasus *benign* yang salah diklasifikasikan sebagai *malignant*).

$$Specificity = \frac{TN}{TN + FP} \quad (9)$$

7) *Negative Predictive Value (NPV)*: mengukur proporsi prediksi negatif (*benign*) yang benar-benar negatif [26]. NPV tinggi menunjukkan pasien yang diprediksi *benign* memang benar tidak menderita kanker. Hal ini penting untuk memastikan keamanan pasien yang dinyatakan bebas dari kanker oleh model.

$$NPV = \frac{TN}{TN + FN} \quad (10)$$

III. HASIL DAN PEMBAHASAN

A. SVM

Tabel III menyajikan classification report dan metrik evaluasi SVM pada tiga konfigurasi PCA (n=2, 4, 6). SVM menunjukkan pola peningkatan performa yang monoton dan konsisten seiring bertambahnya jumlah komponen PCA, mencerminkan karakteristik kernel RBF yang mampu mengeksplorasi setiap dimensi tambahan secara progresif. Secara keseluruhan, SVM dengan parameter optimal (kernel = rbf, C = 100, gamma = 0.1) berhasil mencapai performa yang sangat tinggi pada seluruh konfigurasi, mengonfirmasi bahwa data WDBC memiliki tingkat separabilitas yang tinggi dalam ruang fitur hasil *preprocessing* yang diterapkan.

Pada konfigurasi PCA = 2, SVM mencapai Accuracy sebesar 97.66% dan AUC sebesar 0.9973. Confusion matrix menunjukkan model menghasilkan TP = 61, TN = 106, FP = 1, FN = 3. Specificity 0.9907 (FP = 1) menunjukkan 106 dari 107 kasus *benign* berhasil diklasifikasikan dengan benar, sementara Sensitivity 0.9531 mengindikasikan 3 dari 64 kasus *malignant* tidak terdeteksi (FN = 3). Precision menunjukkan 98.39% prediksi *malignant* model adalah benar. NPV menunjukkan 97.25% pasien yang diprediksi *benign* memang benar-benar *benign*. F1-score kelas *malignant* tercatat sebesar 0.97. CV accuracy tercatat sebesar $93.40 \pm 1.62\%$, mengindikasikan tidak adanya *overfitting* pada konfigurasi ini.

Konfigurasi PCA = 4 menghasilkan Accuracy yang sama dengan PCA = 2, yaitu 97.66%, sekaligus mencapai AUC tertinggi di antara semua konfigurasi, yaitu sebesar 0.9993. Confusion matrix menunjukkan menghasilkan TP = 60, TN = 107, FP = 0, FN = 4. Specificity sempurna sebesar 1.0000 (FP = 0) menunjukkan tidak ada satu pun kasus *benign* yang salah diklasifikasikan sebagai *malignant*. Pencapaian Specificity sempurna ini mengindikasikan bahwa penambahan dimensi PC3 dan PC4 berkontribusi terhadap peningkatan

kemampuan pemisahan kelas oleh kernel RBF. Precision sempurna 1.000 menunjukkan tidak ada satu pun prediksi *malignant* yang keliru. Namun, sensitivity pada konfigurasi ini sedikit turun ke 0.9375 (FN = 4) dibanding PCA = 2, menunjukkan adanya *trade-off* antara presisi prediksi positif dan kemampuan mendeteksi seluruh kasus *malignant*. NPV 0.9640 mengindikasikan bahwa 4 kasus *malignant* yang tidak terdeteksi menurunkan NPV dari 0.9725 (PCA = 2). F1-score kelas *malignant* tercatat 0.97. CV accuracy meningkat ke $96.80 \pm 1.60\%$.

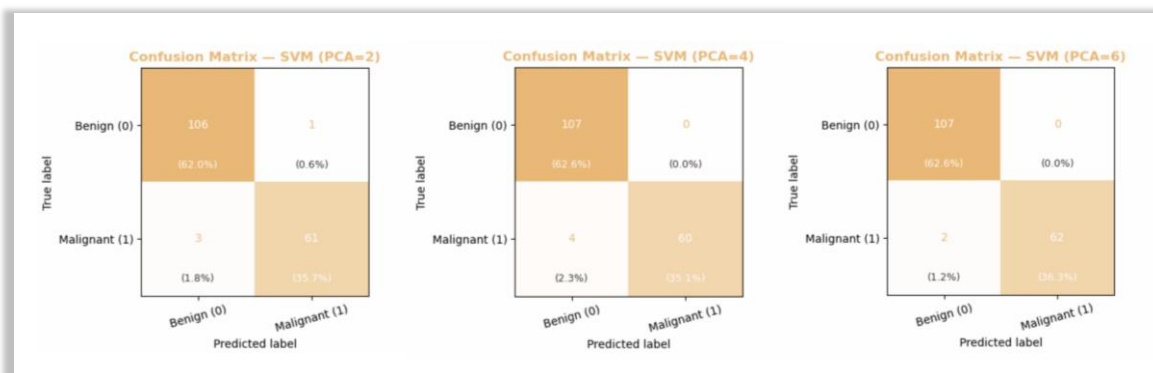
Pada konfigurasi PCA = 6, SVM mencapai puncak secara komprehensif. Confusion matrix menunjukkan menghasilkan TP = 62, TN = 107, FP = 0, FN = 2. Model mencapai Accuracy 98.83%, AUC 0.9985, Specificity 1.0000 (FP = 0),

Sensitivity 0.9688 (FN = 2), Precision 1.0000, NPV 0.9817, dan F1-score *malignant* 0.98. Kombinasi Specificity = 1.0000 dan Precision = 1.0000 sekaligus menunjukkan bahwa model tidak menghasilkan satu pun false positive, sementara NPV (0.9817) mengonfirmasi bahwa hanya 2 dari 109 prediksi *benign* yang keliru. CV accuracy $96.60 \pm 1.36\%$, meskipun rata-ratanya sedikit lebih rendah dibanding PCA = 4 (96.80%), standar deviasi justru paling kecil di antara ketiga konfigurasi, mengindikasikan generalisasi paling stabil. Meskipun AUC PCA = 6 (0.9985) sedikit lebih rendah dari PCA = 4 (0.9993), perbedaan ini tidak signifikan secara statistik dan tidak mengurangi keunggulan komprehensif PCA = 6, terutama dalam meminimalkan false negative (FN = 2) yang paling krusial dalam diagnosis kanker.

TABEL III
CLASSIFICATION REPORT DAN METRIK EVALUASI SVM

PCA	Class	Prec.	Rec.	F1	Acc.	AUC	Spec./Sens	CV Acc.
2	Benign	0.97	0.99	0.98	97.66%	0.9973	Spec: 0.9907 NPV: 0.9725	$93.40 \pm 1.62\%$
	Malignant	0.98	0.95	0.97			Sens: 0.9531 Prec: 0.9839	
4	Benign	0.96	1.00	0.98	97.66%	0.9993	Spec: 1.0000 NPV: 0.9640	$96.80 \pm 1.60\%$
	Malignant	1.00	0.94	0.97			Sens: 0.9375 Prec: 1.0000	
6	Benign	0.98	1.00	0.99	98.83%	0.9985	Spec: 1.0000 NPV: 0.9817	$96.60 \pm 1.36\%$
	Malignant	1.00	0.97	0.98			Sens: 0.9688 Prec: 1.0000	

*Prec. = Precision, Rec. = Recall, F1 = F1-score, Spec. = Specificity, Sens. = Sensitivity, NPV = Negative Predictive Value



Gambar 6. Confusion Matrix SVM pada Konfigurasi PCA n = 2, 4, dan 6.

B. QSVM

Kinerja QSVM pada tiga konfigurasi PCA (n=2, 4, 6) ditampilkan pada Tabel IV. Tidak seperti SVM, QSVM memperlihatkan karakteristik *non-monoton* yang mencerminkan sensitivitas kernel kuantum terhadap dimensionalitas ruang fitur.

Pada konfigurasi PCA = 2, QSVM mencapai AUC tertinggi (0.9734) dengan Accuracy 90.64%. Confusion matrix menghasilkan TP = 55, TN = 100, FP = 7, FN = 9. Specificity 0.9346 (FP = 7) menunjukkan 7 kasus *benign* salah diklasifikasikan. Sensitivity sebesar 0.8594 (FN = 9) menunjukkan 9 kasus *malignant* tidak terdeteksi. Precision 0.8871 mengindikasikan bahwa 88.71% prediksi *malignant*

adalah benar. NPV menunjukkan 91.74% prediksi *benign* adalah benar. F1-score kelas *malignant* tercatat 0.87. CV accuracy $90.20 \pm 3.49\%$ dengan standar deviasi terbesar di antara ketiga konfigurasi, mengindikasikan instabilitas generalisasi pada ruang 2-dimensi.

QSVM pada konfigurasi PCA = 4 mencapai performa terbaik secara keseluruhan dengan Accuracy tertinggi (92.98%) dan AUC sebesar 0.9622. Confusion matrix menghasilkan TP = 59, TN = 100, FP = 7, FN = 5. Sensitivity meningkat ke 0.9219 (FN = 5), merupakan nilai tertinggi di antara ketiga konfigurasi, yang menunjukkan peningkatan kemampuan model dalam mendeteksi kasus *malignant*. Precision *malignant* tercatat 0.8939. NPV meningkat ke 0.9524, mengindikasikan bahwa lebih sedikit prediksi *benign* yang keliru dibanding PCA = 2. Specificity tetap 0.9346 (FP = 7). F1-score kelas *malignant* tercatat 0.91. CV accuracy tercatat $88.00 \pm 1.67\%$, dengan standar deviasi yang lebih kecil dibanding PCA = 2, mengindikasikan generalisasi yang lebih baik.

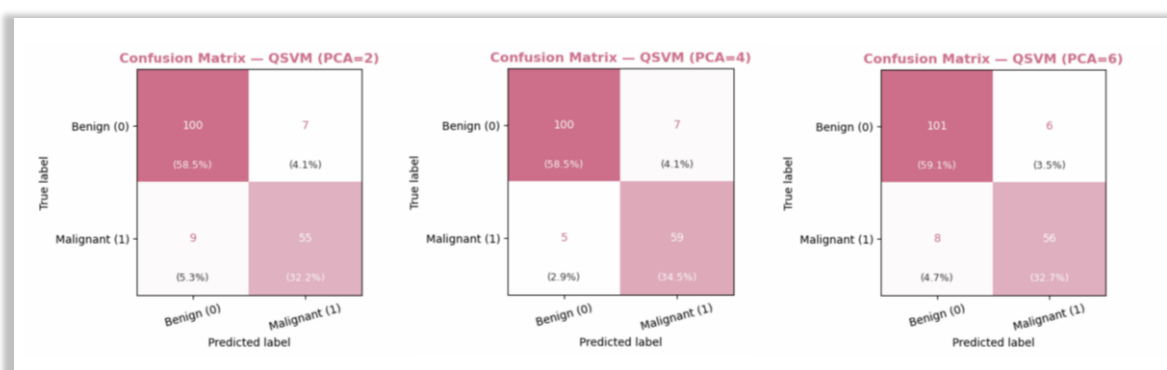
Pada konfigurasi PCA = 6, terjadi degradasi performa. Confusion matrix menghasilkan TP = 56, TN = 101, FP = 6, FN = 8. Accuracy turun ke 91.81%, AUC sebesar 0.9671, Sensitivity turun ke 0.8750 (FN = 8), sementara Specificity sedikit meningkat ke 0.9439 (FP = 6). Precision *malignant* tercatat 0.9032, dan NPV turun ke 0.9266. F1-score kelas *malignant* 0.89. Menariknya, CV accuracy pada PCA = 6 justru mencapai nilai tertinggi ($92.40 \pm 1.96\%$), melebihi test accuracy (91.81%), fenomena ini mengindikasikan bahwa meskipun rata-rata performa validasi silang lebih tinggi, terjadi ketidakstabilan pada *test set* tertentu. Penurunan Sensitivity dan NPV secara bersamaan mengonfirmasi bahwa PC5 – PC6 yang mengandung varian sangat kecil ($< 0.2\%$) bertindak sebagai noise dalam sirkuit kuantum, mendistorsi kemampuan deteksi *malignant*. Fenomena *non-monoton* ini membedakan perilaku kernel kuantum dari kernel RBF klasik yang robust terhadap komponen bervarian rendah.

TABEL IV
CLASSIFICATION REPORT DAN METRIK EVALUASI QSVM

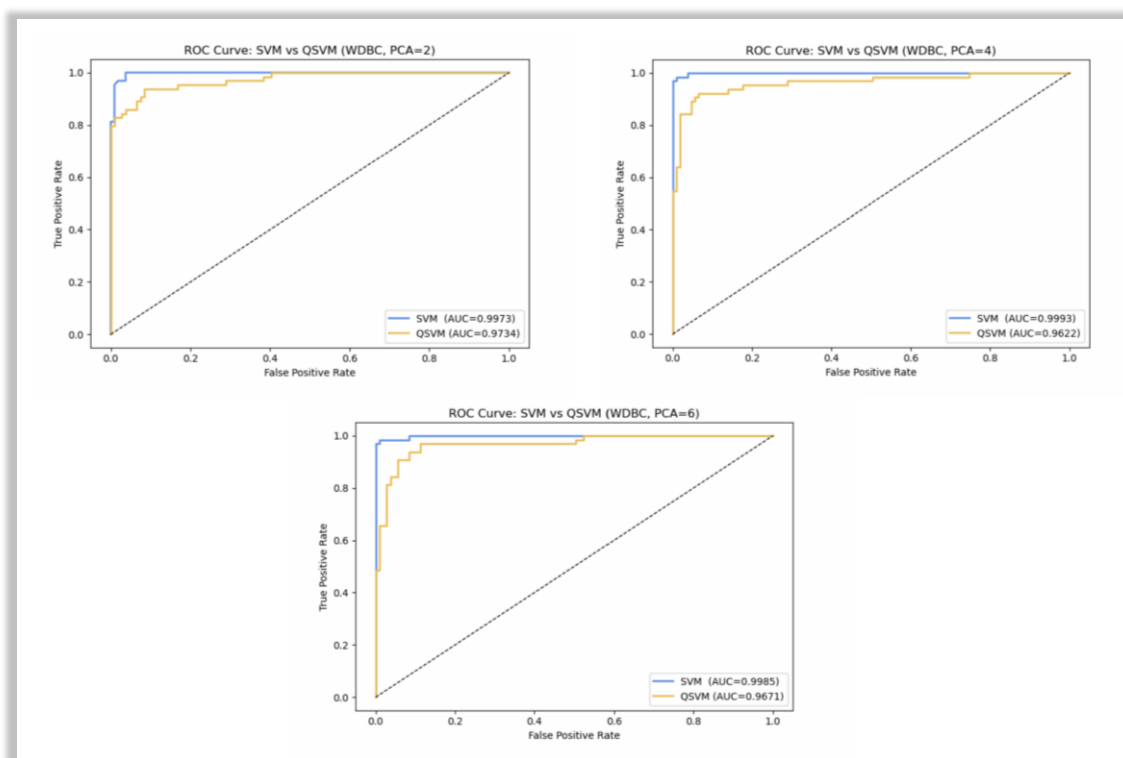
PCA	Class	Prec.	Rec.	F1	Acc.	AUC	Spec./Sens	CV Acc.
2	Benign	0.92	0.93	0.93	90.64%	0.9734	Spec: 0.9346 NPV: 0.9174	$90.20 \pm 3.49\%$
	Malignant	0.89	0.86	0.87			Sens: 0.8594 Prec: 0.8871	
4	Benign	0.95	0.93	0.94	92.98%	0.9622	Spec: 0.9346 NPV: 0.9524	$88.00 \pm 1.67\%$
	Malignant	0.89	0.92	0.91			Sens: 0.9219 Prec: 0.8939	
6	Benign	0.93	0.94	0.94	91.81%	0.9671	Spec: 0.9439 NPV: 0.9266	$92.40 \pm 1.96\%$
	Malignant	0.90	0.88	0.89			Sens: 0.8750 Prec: 0.9032	

*Spesifikasi perangkat : Intel Core i9-12900H, 32GB RAM.

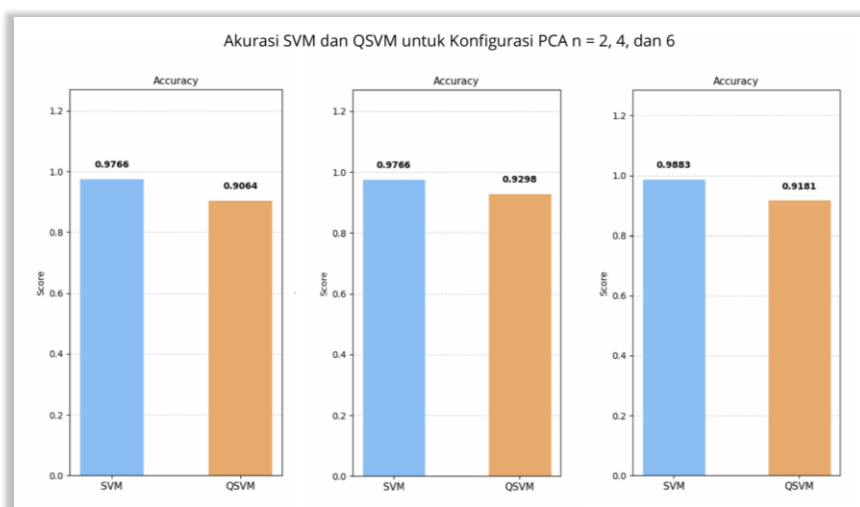
Prec. = Precision, Rec. = Recall, F1 = F1-score, Spec. = Specificity, Sens. = Sensitivity, NPV = Negative Predictive Value



Gambar 7. Confusion Matrix QSVM pada Konfigurasi PCA n = 2, 4, dan 6.



Gambar 8. ROC Curve SVM dan QSVM pada Konfigurasi PCA n = 2, 4, dan 6.



Gambar 9. Akurasi SVM dan QSVM pada Konfigurasi PCA n = 2, 4, dan 6.

C. Pembahasan

SVM dan QSVM menunjukkan perilaku yang sangat berbeda terhadap penambahan dimensi PCA, mencerminkan perbedaan antara kernel RBF dan kernel kuantum. SVM dengan kernel RBF menunjukkan peningkatan performa monoton dari PCA = 2 hingga PCA = 6, karena setiap dimensi tambahan memperkaya representasi geometri data dan

meningkatkan resolusi pengukuran kesamaan antar sampel. Dengan regularisasi yang longgar (Parameter $C = 100$) memungkinkan model membangun batas keputusan yang semakin kompleks tanpa risiko *overfitting* yang signifikan, hal ini dikonfirmasi oleh test accuracy yang konsisten melampaui CV accuracy pada seluruh konfigurasi. Stabilitas generalisasi model juga didukung oleh standar deviasi CV yang terus menurun (1.62% $\rightarrow$ 1.60% $\rightarrow$ 1.36%).

Performa tinggi SVM, terutama pada PCA = 6 dengan Accuracy 98.83%, F1-score tinggi pada kedua kelas (*benign* 0.99, *malignant* 0.98), serta AUC 0.9985 yang mengonfirmasi diskriminasi genuine, mencerminkan struktur data WDBC yang memang memiliki separabilitas tinggi, terutama setelah *preprocessing* yang diterapkan (log1p + SMOTE + MinMaxScaler + PCA). Hal ini juga sejalan dengan laporan penelitian lain pada WDBC yang melaporkan performa SVM yang tinggi [4], [5], [6], [7].

Sebaliknya, QSVM menunjukkan pola *non-monoton* yang kompleks terhadap dimensi PCA, mencerminkan mekanisme yang berbeda secara mendasar. Kernel kuantum mengukur *quantum fidelity* bukan jarak Euclidean seperti RBF. Pada PCA = 2, AUC tertinggi (0.9734) diraih namun Sensitivity terendah (0.8594) dan F1-score terendah (0.87), mengindikasikan *threshold* keputusan default belum optimal. PCA = 4 menjadi titik optimal dengan Accuracy 92.98%, Sensitivity terbaik (0.9219), F1-score terbaik (0.91), NPV terbaik (0.9524), dan CV std paling kecil (1.67%). Konsistensi ini dikarenakan PC1 – PC4 telah menangkap 99.98% varian sehingga tidak ada informasi diskriminatif yang tersisa di PC5 – PC6. Pada PCA = 6, penurunan Sensitivity (0.8750), F1-score (0.89), dan NPV (0.9266) secara bersamaan mengonfirmasi bahwa PC5 – PC6 (<0.02% varian) bertindak sebagai noise dalam sirkuit kuantum, di mana gate ZZ-interaction full entanglement mengkodekan semua pasang qubit tanpa seleksi relevansi, sehingga noise turut terkodekan dan mendistorsi kernel matrix. Fenomena ini tidak dijumpai pada kernel RBF yang secara alami menekan komponen bervariasi rendah melalui parameter gamma.

Dari perspektif klinis, metrik Specificity, Sensitivity, Precision, NPV, dan F1-score memberikan gambaran komprehensif tentang keandalan sistem diagnosis. SVM PCA = 6 menghasilkan Specificity = 1.0000 dan Precision = 1.0000, Sensitivity = 0.9688 (hanya 2 dari 64 *malignant* tidak terdeteksi), NPV = 0.9817, serta F1-score *malignant* = 0.98. QSVM PCA = 4, meskipun Sensitivity kompetitif (0.9219) dan F1-score = 0.91, masih mencatat 7 false positive dari 107 kasus *benign* dengan Specificity 0.9346 dan NPV 0.9524, yang mengindikasikan potensi overdiagnosis yang lebih tinggi dibandingkan SVM.

Perbedaan pipeline antara SVM (Split → SMOTE → MinMaxScaler → PCA) dan QSVM (PCA → Split → MinMaxScaler → SMOTE) dalam penelitian ini merupakan keputusan metodologis yang dirancang dengan justifikasi teknis masing-masing. Pada SVM, urutan konvensional mencegah *data leakage* secara ketat dalam kerangka GridSearchCV. Pada QSVM, PCA global sebelum split diperlukan untuk menjaga stabilitas komponen principal pada dataset kecil (569 sampel), sementara MinMaxScaler global memastikan konsistensi domain rotasi $[0, \pi]$ pada ZZFeatureMap untuk seluruh data point. Perbedaan pipeline ini menghasilkan distribusi varian PCA yang berbeda (SVM: 72.54% – 92.42%; QSVM: 99.61% – 100.00%), yang turut berkontribusi pada perbedaan perilaku kedua model terhadap dimensi PCA. Dengan demikian, perbandingan langsung nilai

metrik antara SVM dan QSVM pada konfigurasi PCA yang sama bersifat komparatif dalam konteks masing-masing pipeline, bukan perbandingan pada representasi data yang identik. Keterbatasan ini perlu dipertimbangkan dalam interpretasi hasil, dan penelitian mendatang disarankan untuk menginvestigasi dampak penyeragaman pipeline terhadap perbandingan kinerja SVM dan QSVM secara lebih terkontrol.

IV. KESIMPULAN

Berdasarkan eksperimen sistematis pada WDBC dataset, penelitian ini menghasilkan tiga kesimpulan utama.

SVM dengan kernel RBF secara konsisten mengungguli QSVM pada semua konfigurasi PCA, dengan performa puncak pada PCA = 6: Accuracy = 98.83%, AUC = 0.9985, Specificity = 1.0000, Sensitivity = 0.9688, Precision = 1.0000, NPV = 0.9817, dan F1-score *malignant* = 0.98. Kombinasi Specificity dan Precision sempurna mengkonfirmasi zero false positive, menjadikan konfigurasi ini paling optimal secara klinis.

QSVM menunjukkan pola *non-monoton* terhadap dimensi PCA, konfigurasi optimalnya berada pada PCA = 4: Accuracy = 92.98% AUC = 0.9622, Sensitivity = 0.9219, Precision = 0.8939, NPV = 0.9524, dan F1-score *malignant* = 0.91. PCA = 4 merupakan titik optimal karena PC1 – PC 4 telah menangkap 99.98% varian data pada pipeline QSVM, sehingga komponen tambahan (PC5 – PC6) hanya membawa noise yang mendistorsi kernel kuantum dan menurunkan seluruh metrik termasuk Sensitivity, NPV, dan F1-score. Temuan ini mengidentifikasi bahwa pemilihan dimensi PCA untuk QSVM harus mempertimbangkan ambang varian kumulatif, bukan sekadar menambah komponen. Fenomena *non-monoton* ini menunjukkan bahwa kernel kuantum memiliki karakteristik yang berbeda dari kernel RBF.

Secara keseluruhan, kedua model terbukti mampu mengklasifikasikan data WDBC dengan baik, namun dengan pendekatan dan *preprocessing* yang berbeda secara fundamental. Penelitian mendatang dapat mengeksplorasi seleksi komponen PCA yang lebih adaptif untuk QSVM, variasi konfigurasi ZZFeatureMap, validasi pada perangkat keras kuantum nyata dan dataset medis lain, serta penyeragaman pipeline *preprocessing* antara SVM dan QSVM untuk memungkinkan perbandingan yang lebih terkontrol.

DAFTAR PUSTAKA

- [1] U. Ullah and B. Garcia-Zapirain, "Quantum Machine Learning Revolution in Healthcare: A Systematic Review of Emerging Perspectives and Applications," IEEE Access, vol. 12, pp. 11423–11450, 2024, doi: 10.1109/ACCESS.2024.3353461.
- [2] E. T. Sedeta, B. Jobre, and B. Avezbakiev, "Breast cancer: Global patterns of incidence, mortality, and trends,," Journal of Clinical Oncology, vol. 41, no. 16_suppl, pp. 10528–10528, Jun. 2023, doi: 10.1200/JCO.2023.41.16_SUPPL.10528.
- [3] H. Li, "Support Vector Machine,," Machine Learning Methods, pp. 127–177, 2024, doi: 10.1007/978-981-99-3917-6_7.
- [4] S. Ray, D. Pradhan, and N. K. Ray, "Breast Cancer Classification Using Machine Learning and Feature Subset Selection Techniques,,"

- Proceedings - 2024 OITS International Conference on Information Technology, OCIT 2024, pp. 488–494, 2024, doi: 10.1109/OCIT65031.2024.00091.
- [5] M. M. Rahman, S. Sarker, M. M. Hasan Shaikat, S. Das, and M. R. K. Khan, "Machine Learning for Breast Cancer Classification: A Comparative Study of Grid Search-Optimized SVM, Random Forest, and XGBoost," 2025 IEEE International Conference on Quantum Photonics, Artificial Intelligence, and Networking, QPAIN 2025, 2025, doi: 10.1109/QPAIN66474.2025.11172245.
- [6] H. Kurniawan, A. Aminuddin, T. Hidayat, N. Norhikmah, K. R. Hidayat, and N. Larasati, "A Comparative Performance Analysis of SVM Kernels in Automated Breast Cancer Diagnosis," 2024 International Conference on Information Technology Systems and Innovation, ICITSI 2024 - Proceedings, pp. 230–235, 2024, doi: 10.1109/ICITSI65188.2024.10929383.
- [7] S. Hariharan, V. Madhivanan, M. Krisnamoorthy, and A. Kumar Cherukuri, "Enhanced QSVM with elitist non-dominated sorting genetic optimisation algorithm for breast cancer diagnosis," inspirehep.net, 2024, doi: 10.1049/qtc2.12113.
- [8] S. Khurana and M. Nene, "Quantum machine learning: unraveling a new paradigm in computational intelligence," authorea.com, 2024, doi: 10.22541/AU.171037552.27623320.
- [9] S. Tomar, R. Tripathi, and S. Kumar, "Comprehensive Survey of QML: From Data Analysis to Algorithmic Advancements," Jan. 2025, Accessed: Jan. 12, 2026. [Online]. Available: <https://arxiv.org/pdf/2501.09528>
- [10] M. Akrom, "Quantum Support Vector Machine for Classification Task: A Review," Journal of Multiscale Materials Informatics, vol. 1, no. 2, pp. 1–8, Jul. 2024, doi: 10.62411/JIMAT.V1I2.10965.
- [11] A. Routh, P. Gopakumar, and G. Rubell Marion Lincy, "An Empirical Study on Enhancing Explainability Through Quantum Machine Learning for Credit Risk Analysis," Proceedings - 2024 IEEE International Conference on Future Machine Learning and Data Science, FMLDS 2024, pp. 443–450, 2024, doi: 10.1109/FMLDS63805.2024.00084.
- [12] T. H. Dang, V. M. Nguyen, D. Do Van, T. Le Tien, and L. D. Son, "Integrating Dimensionality Reduction and Quantum Support Vector Machine (QSVM) for Cancer Diagnosis," pp. 1–6, Mar. 2026, doi: 10.1109/ICHST66555.2025.11428348.
- [13] E. Akpinar, S. M. N. Islam, and M. Oduncuoglu, "Evaluating the Impact of Different Quantum Kernels on the Classification Performance of Support Vector Machine Algorithm: A Medical Dataset Application," Jul. 2024, Accessed: Sep. 14, 2026. [Online]. Available: <https://arxiv.org/pdf/2407.09930>
- [14] P. Joshi and A. Saxena, "Quantum Machine Learning for Breast Cancer Detection: A Comparison with Classical Machine Learning Techniques," 2025 2nd International Conference on Integration of Computational Intelligent System, ICICIS 2025, 2025, doi: 10.1109/ICICIS65613.2025.11371094.
- [15] "Breast Cancer Wisconsin (Diagnostic) - UCI Machine Learning Repository." Accessed: Nov. 21, 2025. [Online]. Available: <https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic>
- [16] W. R. Nurputra, W. Witanti, and A. Komarudin, "Principal Component Analysis (PCA) Untuk Meningkatkan Hasil Klasterisasi Penjualan Video Game Menggunakan Algoritma K-Means," Jurnal Locus Penelitian dan Pengabdian, vol. 4, no. 8, pp. 4683–4703, Aug. 2025, doi: 10.58344/LOCUS.V4I8.4151.
- [17] G. Agliardi et al., "Mitigating exponential concentration in covariant quantum kernels for subspace and real-world data," npj Quantum Information 2025, Dec. 2025, doi: 10.1038/s41534-025-01154-2.
- [18] G. Naidu, T. Zuva, and E. M. Sibanda, "A Review of Evaluation Metrics in Machine Learning Algorithms," Lecture Notes in Networks and Systems, vol. 724 LNNS, pp. 15–25, 2023, doi: 10.1007/978-3-031-35314-7_2.
- [19] A. Fatumnisa and H. Marcos, "Prediksi Kelulusan Tepat Waktu Siswa SMK Teknik Komputer Menggunakan Algoritma Random Forest," Jurnal Manajemen Informatika (JAMIKA), vol. 14, no. 1, pp. 101–111, Apr. 2024, doi: 10.34010/JAMIKA.V14I1.12114.
- [20] W. Arshad et al., "Cancer Unveiled: A Deep Dive Into Breast Tumor Detection Using Cutting-Edge Deep Learning Models", doi: 10.1109/ACCESS.2023.3335604.
- [21] S. Sathyanarayanan, "Confusion Matrix-Based Performance Evaluation Metrics," African Journal of Biomedical Research, pp. 4023–4031, Nov. 2024, doi: 10.53555/AJBR.V27I4S.4345.
- [22] A. Hade and M. Elhia, "Predicting mortgage credit defaults in morocco using machine learning approaches," Discover Artificial Intelligence 2025 5:1, vol. 5, no. 1, pp. 113–, Jun. 2025, doi: 10.1007/S44163-025-00303-Y.
- [23] S. Roumeliotis et al., "ROC curve analysis: a useful statistic multi-tool in the research of nephrology," International Urology and Nephrology 2024 56:8, vol. 56, no. 8, pp. 2651–2658, Mar. 2024, doi: 10.1007/S11255-024-04022-8.
- [24] L. M. Esteban et al., "Integrating radiological and clinical data for clinically significant prostate cancer detection with machine learning techniques," Scientific Reports 2025 15:1, vol. 15, no. 1, pp. 4261–, Feb. 2025, doi: 10.1038/s41598-025-88297-6.
- [25] D. Gaspar, P. Silva, and C. Silva, "Explainable AI for Intrusion Detection Systems: LIME and SHAP Applicability on Multi-Layer Perceptron," IEEE Access, vol. 12, pp. 30164–30175, 2024, doi: 10.1109/ACCESS.2024.3368377.
- [26] Z. Yue et al., "Predictive Diagnosis of Malignant Pulmonary Nodules Using Machine Learning Techniques," Aug. 2023, doi: 10.21203/RS.3.RS-3137418/V1.