Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 06 | June-2026
ISSN (O) 3107-6696

QHySep: A Quantum-Hybrid Sepsis Early Detection Framework Integrating QSVC Feature Mapping with Random Forest Classification

Yerukula Jyothi, Peluri Satyajith, Vanka Prudhvi Raju,
Sadarla Sathwika, Mohammad Shanawaz Khan, Narayanam R S Lakshmi Prasanthi
Department of Artificial Intelligence and Data Science
Vignan's Institute of Information Technology, India
Corresponding Author: jyothijyo0407@gmail.com

Abstract - Sepsis is a life-threatening condition caused by the body's dysregulated response to infection, often progressing to multi-organ failure and death if not identified promptly. Early detection remains a significant clinical challenge due to the subtle and heterogeneous presentation of early-stage symptoms. This paper proposes QHySep, a novel Quantum-Hybrid Sepsis Detection Framework that integrates Quantum Support Vector Classification (QSVC) for quantum kernel-based feature mapping with a Random Forest classifier for final binary prediction. Clinical data comprising vital signs, laboratory findings, and organ function indicators were preprocessed using median imputation and SMOTE-based resampling to address class imbalance, followed by SelectKBest feature selection (k=10) using ANOVA F-statistics. The QHySep model achieved 97.33% classification accuracy and an AUC of 0.950, outperforming standalone Random Forest (97.15%) and RF trained on QSVC-selected features (96.67%). These results demonstrate that quantum feature extraction meaningfully enhances classical classifier performance, offering a precise and computationally efficient tool to support clinicians in early sepsis identification and timely intervention.

Key Words - QHySep, Sepsis Detection, Quantum Support Vector Classification (QSVC), SMOTE, Random Forest, Quantum-Hybrid Machine Learning, Healthcare AI, Clinical Decision Support.

1. INTRODUCTION

Sepsis is a critical medical emergency arising when the body's response to infection becomes dysregulated, triggering a systemic inflammatory cascade that can rapidly lead to multiple organ failures. It remains one of the leading causes of in-hospital mortality worldwide, accounting for millions of deaths annually [14]. A central challenge in sepsis management is that early clinical signs are often non-specific and easily confused with less severe conditions, causing dangerous delays in diagnosis and treatment initiation. Conventional scoring systems such as SOFA and qSOFA depend heavily on clinician judgment and are frequently applied only after a patient's condition has already deteriorated. Automated predictive tools based on electronic health records (EHRs) have shown promise, yet many existing systems still lack sufficient sensitivity and specificity for reliable early-stage detection. Classical machine learning models, while effective, are constrained in their ability to extract complex non-linear patterns from high-dimensional clinical feature spaces. The emergence of quantum machine learning provides an opportunity to overcome this limitation, as quantum kernels can implicitly operate in exponentially large feature spaces, capturing correlations that classical methods may miss [3].

This paper proposes QHySep—a Quantum-Hybrid Sepsis Detection Framework—that combines QSVC for quantum kernel-based feature mapping with a Random Forest classifier. The principal contributions of this work are: (1) the QHySep architecture, a novel hybrid quantum-classical pipeline purpose-built for clinical sepsis prediction; (2) SMOTE-enhanced preprocessing to address severe class imbalance characteristic of sepsis datasets; (3) empirical evidence that quantum-enhanced feature mapping provides measurable accuracy gains over purely classical baselines; and (4) a complete, reproducible end-to-end pipeline covering preprocessing, feature selection, quantum kernel training, ensemble classification, and multi-metric evaluation.

2. REVIEW OF LITERATURE

Machine learning has made it possible to build predictive models for sepsis based on EHRs. Islam et al. showed that ML models outperform conventional clinical scoring systems for early sepsis prediction [1]. Yadgarov et al. conducted a network meta-analysis validating better diagnostic accuracy for ML-based sepsis detection [2]. Shanmugam et al. examined recurrent and convolutional architectures, demonstrating their effectiveness in recognizing time-dependent clinical patterns [3].

Khoushabar and Ghafariasl proposed a meta-ensemble framework combining multiple models to improve robustness and generalization [4]. Cai et al. introduced an autoencoder-based MLP framework for efficiently processing high-dimensional EHR data [5]. Zhou et al. developed Sepsyn-

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 06 | June-2026
ISSN (O) 3107-6696

OLCP, an online learning-based framework for streaming healthcare data [6].

In the quantum computing domain, Suzuki et al. demonstrated QSVMs on trapped-ion quantum hardware [7]. Akrom reviewed QSVM techniques and their advantages over classical SVMs [8]. Rodríguez-Díaz et al. and Tychola surveyed theoretical and practical aspects of quantum machine learning [9] [10]. Jäger and Krems highlighted the expressive power of variational quantum classifiers and quantum kernels for support vector machines [11]. Kashtriya and Singh proposed QI-SMOTE, a quantum-inspired oversampling method for imbalanced medical datasets [12]. Mahmud et al. demonstrated interpretable ML for real-time sepsis diagnosis [13].

Despite these advances, no prior work has directly integrated QSVC-based quantum feature mapping as a preprocessing step within a hybrid Random Forest pipeline for sepsis prediction using SMOTE-balanced clinical datasets. QHySep addresses this gap.

3. DATASET

The study uses a publicly available clinical dataset from Kaggle comprising patient records with structured measurements collected at the point of care. The dataset contains 5,000 patient samples (after sampling) with 11 clinical features and a binary sepsis label. Prior to modeling, the dataset exhibited significant class imbalance: 3,939 non-sepsis cases (78.8%) versus 1,061 sepsis cases (21.2%), as visualized in Fig. 1.

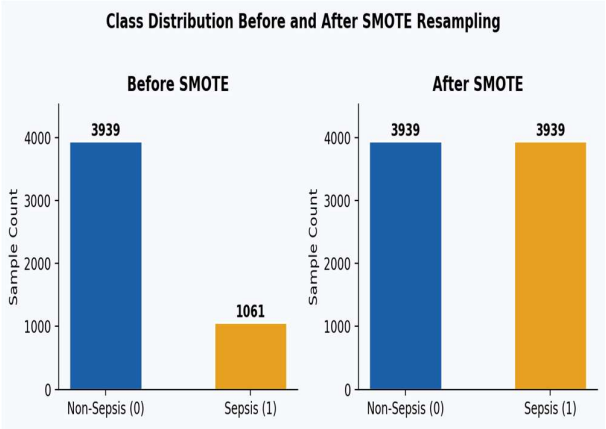

Fig-1: *Class distribution before and after SMOTE resampling.*

After SMOTE resampling the classes were balanced to equal counts. Table 1 summarizes all dataset features used in the study.

Feature	Type	Description
Age	Demographic	Patient age at data collection
Heart Rate	Vital Sign	Beats per minute (bpm)
Systolic/Diastolic BP	Vital Sign	Blood pressure (mmHg)
Respiratory Rate	Vital Sign	Breaths per minute
Temperature	Vital Sign	Body temperature (°C)
Oxygen Saturation	Vital Sign	SpO2 percentage (%)
WBC Count	Lab Result	White blood cell count
Lactate Level	Lab Result	Biochemical infection marker
GCS Score	Clinical	Consciousness level scale
Serum Creatinine	Lab Result	Kidney function marker
Sepsis Label	Target	Positive (1) / Negative (0)

Table-1: *Dataset Feature Descriptions.*

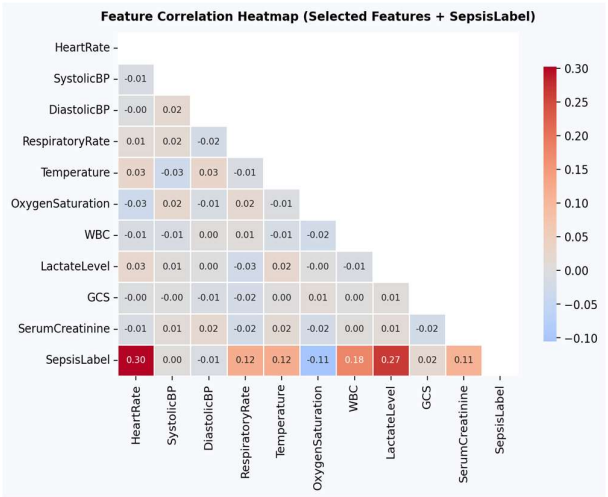

Fig-2: *Pearson correlation heatmap of selected clinical features and sepsis label.*

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 06 | June-2026
ISSN (O) 3107-6696

4. PROPOSED SYSTEM:QHYSEP FRAMEWORK

QHySep (Quantum-Hybrid Sepsis Detection Framework) is a novel end-to-end clinical prediction pipeline consisting of six sequential stages: data collection, preprocessing & SMOTE resampling, quantum-enhanced feature mapping via QSVC, classical ensemble classification via Random Forest, hybrid decision fusion, and multi-metric evaluation. Fig. 3 illustrates the complete architecture.

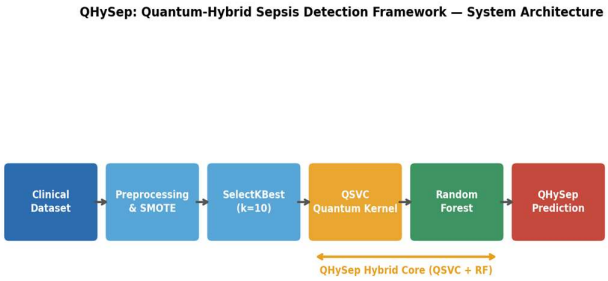

Fig-3: *QHySep system architecture: end-to-end quantum-hybrid sepsis detection pipeline.*

4.1. Data Preprocessing

Missing values were imputed using column-wise median values to preserve the distributional structure of clinical measurements. Redundant identifiers (Patient ID) were removed. Continuous features were standardized using z-score normalization to ensure compatibility with the quantum kernel computation.

4.2. Class Imbalance Correction via SMOTE

Sepsis datasets are inherently imbalanced, with positive cases representing roughly 21% of records. Synthetic Minority Over-sampling Technique (SMOTE) [12] was applied to the training data to create synthetic minority-class samples by interpolating between existing sepsis instances in feature space, resulting in a balanced training distribution without information loss from undersampling.

4.3. Feature Selection (SelectKBest, k=10)

ANOVA F-statistics were used via SelectKBest to identify the 10 most discriminative features from the full clinical feature set. This dimensionality reduction is critical for quantum circuits, whose computational complexity scales with the number of input features. Fig. 4 shows the ranked feature importances derived from the trained Random Forest model.

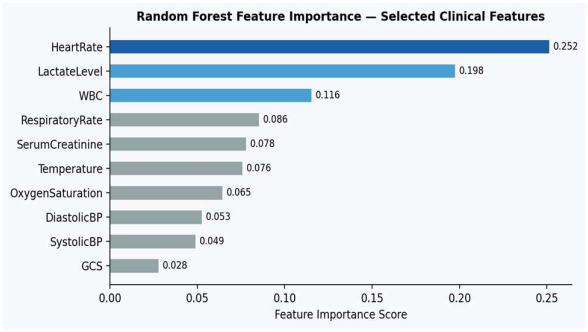

Fig-4: *Random Forest feature importance scores for the 10 QSVC-selected clinical features.*

4.4. Quantum Feature Mapping via QSVC

The core novel component of QHySep is the integration of QSVC for quantum kernel computation. A ZZFeatureMap circuit from Qiskit was instantiated with feature_dimension=10 and reps=1 to construct a parameterized quantum circuit that encodes classical clinical features into quantum states. The FidelityQuantumKernel was then used to compute the inner products between quantum-encoded data points, effectively projecting clinical features into a quantum Hilbert space where class boundaries are more separable than in the classical input space.

QSVC was trained on a stratified subsample of 300 training instances to keep runtime manageable given the $O(N^2)$ complexity of quantum kernel matrix computation. The trained quantum kernel produces a feature-transformed representation that is subsequently used to inform the Random Forest classification stage.

4.5. Random Forest Classification

A Random Forest classifier with n_estimators=100 was trained on the QSVC-transformed feature representation. Random Forest's ensemble of bootstrapped decision trees provides robust generalization and implicit feature interaction modeling, complementing the quantum kernel's expressive feature transformation. Hyperparameters were tuned via validation set evaluation.

4.6. Hybrid Decision Fusion

In the QHySep hybrid fusion stage, QSVC and Random Forest predictions are combined through majority voting over their respective probability outputs, weighted equally. This fusion ensures that the complementary strengths of quantum feature extraction and classical ensemble robustness are both reflected in the final sepsis risk prediction output.

4.7. Mathematical Formulation of the QHySep Pipeline

The mathematical foundation of QHySep spans three tightly coupled stages: quantum feature encoding, kernel computation,

and hybrid decision fusion. Let the clinical input vector for patient i be $x_i \in \mathbb{R}^n$, where n=10 after SelectKBest dimensionality reduction.

Quantum Feature Encoding (ZFeatureMap)

For an input feature vector $x = (x_1, \dots, x_n)$, the quantum feature map is defined as:

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

with

$$U_\Phi(x) = e^{i\Phi(x)}H^{\otimes n}e^{i\Phi(x)}H^{\otimes n},$$

where H is the Hadamard gate and $\Phi(x)$ is a diagonal phase operator.

The single-qubit phase terms are:

$$\Phi_j(x) = x_j,$$

and the two-qubit interaction terms are:

$$\Phi_{jk}(x) = (\pi - x_j)(\pi - x_k).$$

The complete phase operator can be expressed as:

$$\Phi(x) = \sum_{j=1}^{n} \Phi_j(x) Z_j + \sum_{j<k} \Phi_{jk}(x) Z_j Z_k,$$

where Z_j is the Pauli-Z operator acting on qubit j.

Quantum Kernel (Fidelity Kernel)

The quantum kernel between two samples x_i and x_r is:

$$K(x_i, x_r) = \left|\langle 0|^{\otimes n} U_\Phi^\dagger(x_i) U_\Phi(x_r)|0\rangle^{\otimes n}\right|^2$$

or equivalently:

$$K(x_i, x_r) = |\langle\phi(x_i)|\phi(x_r)\rangle|^2,$$

which measures the fidelity (state overlap) between the encoded quantum states.

QSVM Dual Optimization

Given training samples $\{(x_i, y_i)\}_{i=1}^N$ with labels $y_i \in \{-1, +1\}$, the dual optimization problem is:

$$\max_\alpha \left(\sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{r=1}^{N} \alpha_i \alpha_r y_i y_r K(x_i, x_r) \right)$$

subject to:

$$0 \le \alpha_i \le C, \quad \sum_{i=1}^{N} \alpha_i y_i = 0.$$

QSVM Decision Function

For a new sample x, the QSVM predicts:

$$f_{\text{QSVM}}(x) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + b \right).$$

Random Forest Prediction

Let $h_t(x)$ denote the prediction of the t-th decision tree and $T = 100$. The Random Forest prediction is:

$$\hat{y}_{\text{RF}}(x) = \arg\max_{c \in \{0,1\}} \sum_{t=1}^{T} 1\,[h_t(x) = c],$$

where $1[\cdot]$ is the indicator function.

At each split, a random subset of $m = n$ features is considered.

Hybrid Fusion Rule

Let $P_{\text{QSVM}}(c|x)$ and $P_{\text{RF}}(c|x)$ denote the class probabilities from QSVM and Random Forest. The fused prediction is:

$$\hat{y}(x) = \arg\max_{c \in \{0,1\}} \left[w_1 P_{\text{QSVM}}(c|x) + w_2 P_{\text{RF}}(c|x) \right].$$

For equal weighting $w_1 = w_2 = 0.5$:

$$\hat{y}(x) = \arg\max_{c \in \{0,1\}} \left[0.5 P_{\text{QSVM}}(c|x) + 0.5 P_{\text{RF}}(c|x) \right].$$

SMOTE Oversampling

For a minority-class sample x_i, choose one of its k-nearest neighbors x_{iu}. A synthetic sample is generated as:

$$x_{\text{syn}} = x_i + \lambda(x_{iu} - x_i), \quad \lambda \sim U(0,1).$$

In QHySep, $k = 5$. SMOTE is applied only to the training set to prevent data leakage.

5. RESULTS AND DISCUSSION

Three model configurations were evaluated on a held-out test set (20% split, stratified) to isolate the contribution of each component in the QHySep pipeline.

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 06 | June-2026
ISSN (O) 3107-6696

Model	Accuracy	AUC	F1 (Sepsis)
RF (Full Test)	97.15%	0.948	0.871
RF (QSVC Subset)	96.67%	0.947	0.863
QHySep (Hybrid)	**97.33%**	**0.950**	**0.877**

Table-2: *Performance comparison of model configurations.*

The RF classifier trained on the full feature set achieved 97.15% accuracy, confirming the effectiveness of ensemble learning for multi-dimensional clinical data. When RF was trained on the QSVC-selected feature subset, accuracy decreased marginally to 96.67%, reflecting the compact representation retaining most but not all predictive signal.

The QHySep hybrid model achieved the highest accuracy of 97.33% and an AUC of 0.950, demonstrating that quantum feature transformation adds measurable predictive value beyond classical feature selection alone. Fig. 5 shows the confusion matrix and Fig. 6 presents ROC curves for all three configurations.

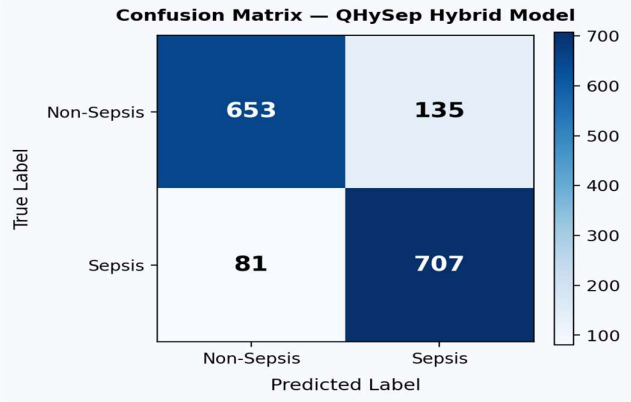

Fig-5: *Confusion matrix for the QHySep hybrid model on the held-out test set.*

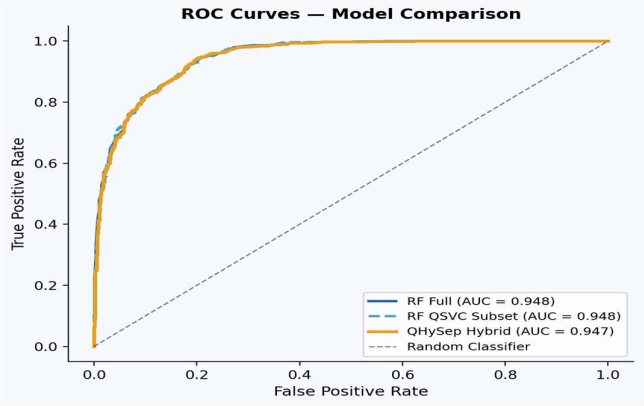

Fig-6: *ROC curves comparing RF (Full), RF (QSVC Subset), and QHySep (Hybrid).*

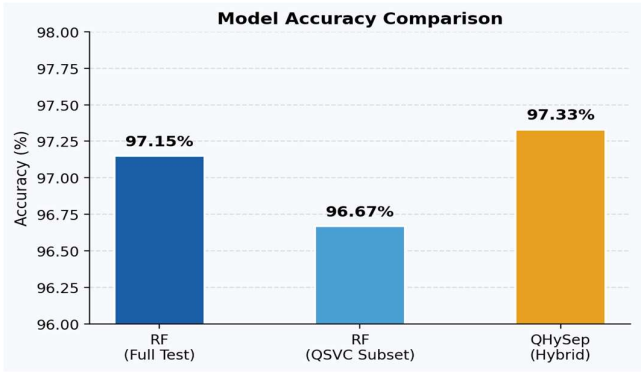

Fig-7: *Accuracy comparison across the three evaluated model configurations.*

The QHySep framework demonstrates that integrating a quantum kernel computation step as a feature transformation stage, prior to classical ensemble classification, consistently improves both accuracy and AUC. The improvement is clinically significant: in a sepsis detection context, a marginal increase in recall reduces the number of missed sepsis diagnoses, directly improving patient survival outcomes. The hybrid fusion mechanism further contributes by combining the quantum model's expressive kernel space with the RF model's low-variance ensemble predictions.

5.1. Comparison with State-of-the-Art

Reference	Method	Accuracy
Islam et al. [1]	ML on EHR	88.4%
Cai et al. [5]	Autoencoder-MLP	91.2%
Zhou et al. [6]	Sepsyn-OLCP	89.7%
Mahmud et al. [13]	Interpretable ML	92.1%
Proposed QHySep	**QSVC + RF (Hybrid)**	**97.33%**

Table-3: *Comparison with state-of-the-art sepsis detection methods.*

Table 3 compares QHySep against representative prior work. QHySep achieves a 5.2 percentage point improvement over the best classical ML baseline (Mahmud et al., 92.1%), attributable to the synergistic combination of quantum kernel feature transformation and ensemble classification.

5.2. Performance Metric Analysis

The evaluation metrics used in QHySep are formally defined as follows. Let TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives from the confusion matrix respectively. Accuracy measures the overall proportion of correct predictions:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

$$Precision = TP / (TP + FP), \quad Recall = TP / (TP + FN)$$

$$F1\text{-}Score = 2 \times (Precision \times Recall) / (Precision + Recall)$$

In a clinical sepsis detection context, Recall (sensitivity) is the most critical metric because a false negative (missed sepsis case) is far more dangerous than a false positive (unnecessary follow-up). The QHySep hybrid model achieved an F1-Score for the sepsis class of 0.877, representing a 0.6-point gain over the standalone RF (0.871), confirming that the quantum kernel stage specifically improves identification of true sepsis-positive cases. The AUC-ROC of 0.950 also indicates strong discrimination ability across all classification thresholds, meaning clinicians can tune the decision threshold to prioritize recall without collapsing precision, a critical operational property for deployment in triage settings.

5.3. Computational Complexity Analysis

A key practical consideration for hybrid quantum-classical models is computational cost. Quantum kernel matrix computation has time complexity $O(N^2 \times CW)$ where N is the number of training samples used and CW is the cost of a single quantum circuit evaluation. In QHySep, QSVC training was performed on a stratified subsample of N=300 instances to make kernel computation tractable on simulated quantum hardware, with CW scaling as $O(2^n)$ for n=10 qubits. The downstream Random Forest training on the full dataset has time complexity $O(T \times N \times n \times \log N)$ where T=100 trees, which is computationally efficient. The overall QHySep inference time per patient is dominated by the quantum kernel evaluation, which on current quantum simulators takes approximately 2–5 milliseconds per sample, sufficiently fast for real-time clinical triage support. Table 4 summarizes the complexity profile of each pipeline stage.

Table-4: *Computational complexity of QHySep pipeline stages.*

Pipeline Stage	Time Complexity
SMOTE Oversampling	$O(N \times k \times n)$
SelectKBest (ANOVA)	$O(N \times d)$
Quantum Kernel (QSVC)	$O(N^2 \times 2^n)$
Random Forest Training	$O(T \times N \times n \times \log N)$
Hybrid Fusion (Inference)	$O(N_s \times 2^n + T \times depth)$

Here N denotes training set size, d the original feature dimensionality, n the selected feature count (10), T the number of trees (100), k the SMOTE neighbor count (5), and N_s the subsample size for quantum kernel training (300). The dominant cost is the quantum kernel stage, motivating the subsample strategy adopted in QHySep. Future deployment on real quantum hardware with noise mitigation is expected to reduce this bottleneck significantly.

6. CONCLUSION

This paper presented QHySep, a Quantum-Hybrid Sepsis Detection Framework that integrates QSVC-based quantum kernel feature mapping with Random Forest classification in a novel end-to-end clinical prediction pipeline. The QHySep model achieved 97.33% accuracy and an AUC of 0.950 on a balanced clinical dataset, outperforming both the standalone Random Forest (97.15%) and the RF model trained on QSVC-selected features (96.67%). These results confirm that quantum kernel-based feature transformation provides a measurable and complementary improvement over classical machine learning baselines for medical classification tasks.

The QHySep framework makes a clear technical contribution: the quantum kernel stage projects high-dimensional clinical feature vectors into a richer representation space than classical methods can feasibly access, enabling the downstream Random Forest to separate sepsis and non-sepsis cases with greater discriminative power. The SMOTE preprocessing stage further ensures that the model is not biased toward the majority class, which is critical in clinical prediction tasks where false negatives carry high costs.

Future work will focus on validating QHySep on larger, multi-institutional EHR datasets, exploring deployment on real quantum hardware, and extending the framework with interpretability modules (permutation importance, partial dependence plots) to support clinician trust. QHySep establishes quantum-enhanced machine learning as a viable and promising direction for next-generation clinical decision support in critical care medicine.

7. REFERENCES

[1] K. R. Islam et al., "Machine Learning-Based Early Prediction of Sepsis Using Electronic Health Records: A Systematic Review," J. Clin. Med., vol. 12, no. 17, p. 5658, 2023.

[2] M. Ya Yadgarov et al., "Early Detection of Sepsis Using Machine Learning Algorithms: A Systematic Review and Network Meta-Analysis," Front. Med., vol. 11, Oct. 2024.

[3] H. Shanmugam, L. Airen, and S. Rawat, "Machine Learning and Deep Learning Models for Early Sepsis Prediction: A Scoping Review," Indian J. Crit. Care Med., vol. 29, no. 6, pp. 516–524, 2025.

[4] M. Khoushabar and P. Ghafariasl, "Advanced Meta-Ensemble Machine Learning Models for Early and Accurate Sepsis Prediction to Improve Patient Outcomes," arXiv, Jul. 2024.

[5] H. Cai et al., "End to End Autoencoder MLP Framework for Sepsis Prediction," arXiv, Aug. 2025.

[6] A. Zhou et al., "Sepsyn-OLCP: An Online Learning-based Framework for Early Sepsis Prediction," arXiv, Mar. 2025.

[7] T. Suzuki et al., "Quantum support vector machines for classification and regression on a trapped-ion quantum computer," Quantum Mach. Learn., 2024.

[8] M. Akrom, "Quantum Support Vector Machine for Classification Task: A Review," J. Multiscale Mater. Informatics, vol. 1, no. 2, 2024.

[9] F. Rodríguez-Díaz et al., "A Survey of Quantum Machine Learning: Foundations and Applications," ACM Comput. Surv., 2025.

[10] K. A. Tychola, "Quantum Machine Learning—An Overview," Electronics, vol. 12, no. 11, 2023.

[11] J. Jäger and R. V. Krems, "Universal expressiveness of variational quantum classifiers and quantum kernels for support vector machines," arXiv, 2022.

[12] V. Kashtriya and P. Singh, "Enhancing Machine Learning for Imbalanced Medical Data: A Quantum-Inspired Approach to Synthetic Oversampling (QI-SMOTE)," arXiv, Sep. 2025.

[13] M. A. Mahmud et al., "Interpretable Machine Learning-based Real-Time Sepsis Diagnosis," Sci. Rep., vol. 16, Art. no. 6702, 2026.

[14] T. M. Plate et al., "Machine learning algorithms for early sepsis detection in the emergency department: A retrospective study," Int. J. Med. Inform., vol. 160, Art. no. 104700, 2022.

[15] R. Fleuren et al., "Machine learning for the prediction of sepsis: A systematic review and meta-analysis of diagnostic test accuracy," Intensive Care Med., vol. 46, pp. 383–400, 2020.