A QUANTUM SUPPORT VECTOR MACHINE-BASED SEPSIS DIAGNOSIS USING GENE EXPRESSION DATA

Phuong-Thao Vu
Faculty of Interdisciplinary Digital Technologies, Phenikaa University

Abstract: Analyzing differential gene expression in transcriptomic data provides crucial insights into molecular responses to specific biological conditions, potentially revealing valuable biomarkers derived from both established biological knowledge and data-driven approaches. In this study, we present a novel methodology for sepsis diagnosis utilizing immune-related gene expression data to identify optimal biomarker combinations. Our methodological framework incorporates multiple analytical steps: differential gene expression analysis, feature importance evaluation, and an integrated approach combining Recursive Feature Elimination (RFE) with Principal Component Analysis (PCA), and Quantum Support Vector Machine (QSVM)-based classification. The total of 41 immune-related genes is carefully selected to construct a comprehensive diagnostic panel. Implementation of QSVM classification yielded exceptional diagnostic performance, demonstrating superior results across all evaluation metrics. This approach represents a significant advancement in identifying reliable transcriptomic biomarkers for sepsis diagnosis and establishes a robust framework applicable to other complex diseases.

Keywords: Quantum Support Vector Machine, Gene Expression, Machine learning, Principal Component Analysis, Recursive Feature Elimination

I. INTRODUCTION

Sepsis remains one of the most pressing challenges in critical care, being a leading cause of morbidity and mortality worldwide, especially in children. This condition occurs when the body's response to infection triggers widespread inflammation, leading to tissue damage, organ failure, and potentially death. The immune system, attempting to combat infection, releases chemicals into the bloodstream that can provoke an exaggerated inflammatory response. Sepsis progresses through three stages: sepsis, severe sepsis, and septic shock—the latter being the most critical and potentially fatal due to multiorgan dysfunction syndrome [1]. As a time-sensitive medical emergency, early diagnosis and timely therapeutic interventions significantly improve patient outcomes, reduce complications, and decrease mortality. This urgency underscores the necessity for advanced predictive models that can identify at-risk patients before clinical manifestations become severe [2]. Machine learning techniques have emerged as innovative approaches to enhance sepsis prediction accuracy.

Numerous significant studies have explored the application of machine learning algorithms for sepsis prediction, employing various models such as logistic regression, decision trees, and ensemble methods like Gradient Boosting, with promising results. In [3], researchers evaluated predictive factors associated with survival outcomes in sepsis patients by analyzing a dataset of individuals aged 14 years and older admitted to intensive care units (ICUs) during a six-month period in 2018. Using proportional hazards models and data mining algorithms, their survival analysis demonstrated moderate predictive performance, identifying three key variables strongly associated with patient prognosis: time from sepsis onset to outcome (discharge or death), serum lactic acid levels, and body temperature. Similarly, a systematic review [4] highlights the efficacy of machine learning algorithms in predicting sepsis onset in ICU patients. Their analysis reveals that models such as random forests and gradient boosting demonstrate high accuracy and specificity in diagnosing severe sepsis and septic shock, anticipating clinical deterioration several hours before traditional methods can detect it.

Moreover, recent advancements have focused on complex ML architectures like Extreme Gradient Boosting that improve predictive accuracy through dynamic adjustments based on prior models' performance [5]. These approaches highlight a promising trend: utilizing advanced ML techniques not only to enhance prediction accuracy but also to refine the feature space, potentially decreasing the number of required biomarkers. In pursuit of detecting early-onset sepsis across all hospital departments, [6] developed the binary classifier SEPSI Score using gradient boosted trees, which accurately predicted sepsis onset with performance exceeding existing scoring systems.

Building on these advances, recent efforts have focused on enhancing sepsis prediction robustness by refining statistical analysis for differential expression and leveraging data analysis and machine learning techniques

Contact author: Phuong-Thao Vu
Email: thao.vuphuong@phenikaa-uni.edu.vn
Manuscript received: 3/2025, revised: 5/2025, accepted: 6/2025.

to reduce necessary biomarkers. As sepsis is a complex, multifactorial syndrome, developing precise models with streamlined biomarker sets is critical for timely intervention and improved outcomes. In [7], researchers implemented a two-stage gene reduction strategy—identifying differentially expressed genes (DEGs) followed by sequential forward feature selection integrated with a Boosting algorithm—resulting in nine genes that provided remarkably precise prediction of sepsis-related mortality. Similarly, [8] employed a sequential filtering approach incorporating LASSO, Recursive Feature Elimination (RFE), and Random Forest Variable Hunting to identify 18 potential diagnostic marker genes. The study in [9] used LASSO to select eight genes for constructing a Random Forest-based sepsis diagnostic model, while [10] introduced Recurrent Logistic Regression (RLR) to pinpoint five genes strongly linked to immune system function. These findings highlight machine learning algorithms' critical role in genetic data analysis, facilitating effective diagnostic and prognostic models while contributing significantly to understanding sepsis's underlying biological mechanisms.

However, these studies share common limitations: small sample sizes and lack of data diversity reduce finding generalization, complicating model application to broader patient populations. Additionally, overfitting remains concerning when models train on specific datasets without sufficient preventive measures. Quantum Support Vector Machine (QSVM) methods offer potential advantages in addressing these limitations through enhanced processing speed and optimization. Unlike classical models, QSVMs leverage quantum kernel methods to map data into higher-dimensional Hilbert spaces, enabling discovery of complex, non-linear patterns often missed by traditional algorithms. This capability improves generalization, particularly with small or heterogeneous gene expression datasets. Furthermore, QSVMs excel at handling high-dimensional data with correlated features by capturing entangled relationships without overfitting, potentially yielding more robust prediction models. By uncovering consistent patterns across varied datasets, QSVM-based approaches can identify more reliable and biologically meaningful biomarkers, enhancing downstream validation and interpretability. Therefore, incorporating QSVM into sepsis diagnosis models may effectively address current challenges in data adaptability, feature redundancy, and biomarker verification.

The research in [11] proposes a Quantum Variational Support Vector Machine (QV-SVM) as a novel approach for early sepsis prediction, combining quantum computing techniques with traditional machine learning models. The QV-SVM outperformed existing classifiers including Decision Tree, XGBoost, Random Forest, and classical SVM in accuracy, precision, recall, and F1-scores. Despite strong performance, the study acknowledges two key limitations: first, quantum variational circuit complexity may introduce computational challenges in clinical environments lacking quantum infrastructure; second, the datasets used (MIMIC-III and IV), though extensive, may contain imbalanced classes and missing values that affect model generalizability.

In this paper, we propose a novel approach for effective gene selection using data mining and Quantum Support Vector Machine algorithm to identify a compact set of highly informative genes for accurate early sepsis prediction. According to [12], [13], given the significant immune system dysregulation observed in sepsis patients, this research primarily focuses on immune-related gene markers. The key methodology integrates immune-related gene selection across multiple platforms combined with Recursive Feature Elimination and Principal Component Analysis to generate gene combinations. These are then paired with QSVM to select optimal gene combinations, ensuring a highly accurate and generalizable diagnostic model for sepsis. The primary contributions of this study are threefold:

- ✓ To propose a methodology for dimensionality reduction by leveraging immune-related biomarkers through gene selection.
- ✓ To explore the application of Principal Component Analysis (PCA) to uncover underlying patterns and reduce noise within high-dimensional gene expression datasets.
- ✓ To implement Quantum Support Vector Machine (QSVM) to rapidly process high-dimensional and complex medical datasets, identify nonlinear patterns, and maintain robustness against noise, improving both the speed and accuracy of sepsis detection.

The remainder of this paper covers the following content: Section II describes the dataset and preprocessing techniques. Section III outlines the methodology of the study. Section IV presents the experimental results. Lastly, Section V provides the conclusion of the study.

II. DATASET AND PREPROCESSING

A. Dataset Acquisition

The seven publicly available datasets described in Table I originated from the Gene Expression Omnibus (GEO) database. The dataset included around 1,164 samples encompassing samples from both sepsis patients and healthy controls. All datasets underwent uniform preprocessing and normalization using the Robust Multi-array Average (RMA) algorithm. RMA algorithm addresses issues like background correction, normalization, and summarization of probe-level data to improve the reliability and interpretability of gene expression studies [20]. For genes corresponding to multiple probe sets, expression values were computed as the average of all associated probes, based on the probe-to-gene mapping provided in the SOFT and notation files accompanying each GEO dataset. To identify discriminative gene features and train the diagnostic model for sepsis, the training set includes four gene expression datasets — GSE26440 [14], GSE26378 [15], GSE95233 [16], and GSE57065 [17], while the independent validation set includes three datasets — GSE4607 [18], GSE65682 [19], and E MTAB-1548 [11]. The use of independent training and validation datasets helps to reduce the risk of overfitting and ensures that the models built can generalize new and unseen data. These datasets

exhibited diversity in age groups, sample sizes, and were distributed across three distinct microarray platforms, enhancing the biological variability and ensuring generalizability in downstream machine learning models. This stratified selection of datasets is crucial for ensuring the generalizability of the findings, as it reflects the heterogeneity of sepsis cases across different populations and experimental conditions. This structured selection supports a comprehensive and stratified approach to analyzing gene expression patterns in sepsis across varied patient demographics.

Table I. Dataset description

	Gene set	No of			Age
		samples	sepsis	non sepsis	
Training set	GSE26440	130	32	98	Children
	GSE26378	103	21	82	Children
	GSE95233	124	22	102	Adult
	GSE57065	107	25	82	Adult
Validation set	GSE4607	84	15	69	Children
	GSE65682	521	42	479	Adult
	E-MTAB-1548	95	15	80	Adult

B. Immune-Related Gene Selection

A total of approximately 770 immune-related genes were acquired from the NanoString platform, a popular data resource commonly utilized in research examining host-pathogen interactions and immune system responses. Gene mapping was performed across three primary microarray platforms utilized in this study, namely Affymetrix Human Genome U219 (AffyU219), Affymetrix Human Genome U133 Plus 2.0 (AffyU133P2), and Agilent Human Gene Expression 4x44K v2 Microarray (AgilentV2). As a result, it gained 740, 737 and 627numbers of immune-related genes respectively. To facilitate the development of robust, platform-independent biomarkers, we focused our analyses on 608 immune-related genes shared across all three platforms. These common genes served as the foundation for subsequent computational modeling and diagnostic signature development.

C. Differential expression analysis of Immune-Related Genes

To identify key immune-related genes associated with sepsis, we performed differential expression analysis as a preprocessing step on the gene expression datasets. This analysis was conducted using the limma package in R, a widely used tool for microarray and RNA-seq data analysis. To control the false discovery rate, we applied the Benjamini-Hochberg (BH) correction to adjust the p-values resulting from the statistical tests. Differential gene expression was quantified using log-fold change, representing the relative expression of genes in sepsis samples compared to healthy controls. The two main thresholds enabled the selection of immune-related biomarkers that exhibit robust and statistically meaningful differences in expression between sepsis patients and healthy individuals.

III. METHODOLOGY

A structured, multi-phase analytical pipeline was designed to identify and validate immune-related gene signatures with high diagnostic relevance, using a combination of statistical analyses and advanced technique called quantum support vector machine (QSVM). As shown in Figure 1, we employed four main steps: (1) Gene acquisition and preprocessing: Filtering and reducing the number of genes as the input. Raw gene expression data is processed through statistical methods, like differential expression analysis, to highlight genes that show significant changes between sepsis and non-sepsis samples. (2) Gene selection: The differential IRG expression analysis is performed for the construction of the most regulated biomarkers between normal and sepsis patients. Moreover, based on the importance values computed by the XGB model, the RFE feature selection algorithm is used to create various gene subsets. These subsets are then transformed into transformed gene combinations (TGCs) by applying the Principal Component Analysis (PCA) approach for figuring out the most effective gene combination. PCA plays a crucial role in this step by reducing the dimensionality of the gene expression data while retaining the most significant variations. PCA allows for the identification of key patterns within the gene subsets generated by RFE and XGB. This reduction in dimensionality enhances the efficiency of subsequent analyses by eliminating irrelevant features, focusing on the TGCs that explain the highest variance in the data. This step not only simplifies the dataset but also captures the most prominent features for disease classification. (3) Gene combination selection: Selecting the best combinations of genes that will be used for the classification task. The gene combinations, now refined and reduced in dimensionality, are then passed to the QSVM for classification. The input to QSVM consists of these transformed gene combinations. QSVM uses a quantum kernel method, which allows it to map the data into a higher-dimensional space, enabling the algorithm to detect complex, non-linear patterns within the data that might be missed by classical methods. The quantum nature of QSVM provides the advantage of improved generalization, especially in situations where the dataset is small or heterogeneous, as is often the case with medical datasets. This enhanced capability makes QSVM a powerful tool for sepsis diagnosis, as it can effectively handle the challenges of high-dimensional gene expression data without overfitting. The primary purpose of incorporating QSVM is to improve classification accuracy. In this study, it provides a more reliable way to classify sepsis cases based on immune-related gene expression, offering high accuracy and robustness. (4) Gene combination validation: Evaluating the effectiveness of using the optimal gene subset within the validation cohorts. We validate the identified biomarkers using three separate gene expression datasets derived from different microarray platforms, ensuring robustness and generalizability of our findings.

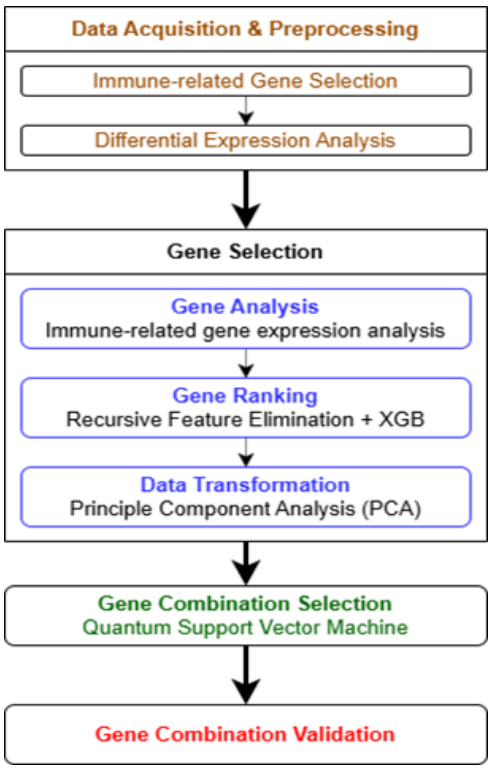

Figure 1. Diagram of the methodology

A. Gene Ranking

In this phase, we evaluate and order the immune-related genes showing differential expression according to the magnitude of their log fold-change values. Genes exhibiting a large expression gap between the two sample types are considered the most informative for identifying sepsis. Subsequently, the Recursive Feature Elimination (RFE) method is employed to construct multiple gene subsets, which are then projected into a component space using Principal Component Analysis (PCA). Initially, genes are ranked according to their importance scores, calculated through the XGBoost (XGB) model combined with a cross- validation approach. Gene subsets are iteratively refined by removing the gene with the lowest average importance score. This process, applying the XGB model and cross-validation, is repeated for each subset until no further genes remain for elimination.

B. Gene Selection

The feature subsets created through the RFE algorithm and cross-validation are transformed into a component space using PCA, with each subset containing a unique combination of top genes. We implemented PCA to resolve the high correlation issues present in the original immune-related gene dataset. This transformation produces a component subset with substantially lower gene correlation, enhancing the selection of key informative genes and improving the diagnostic model's effectiveness. Throughout this process, the total information remains constant, as the number of genes in each subset equals the number of components.

All gene subsets are processed to identify optimal learning and structural parameters for each gene combination using the training dataset. Hyper-parameter tuning is essential for developing the best models and preventing overfitting problems. The models associated with these top gene combinations (TGCs) are then assessed for classification performance using a separate validation dataset. Our methodology combines grid search with five-fold cross-validation to determine the ideal parameter values across all TGCs.

C. Gene Combination Selection and Validation

The optimal subsets of genes, identified during the gene selection phase, are evaluated based on their diagnostic performance on the validation dataset. Subsequently, the Quantum Support Vector Machine (QSVM) algorithm is retrained with the training data and assessed on the validation cohorts using these selected gene combinations to diagnose sepsis.

QSVM uses a quantum kernel method where classical data is encoded into quantum bits (qubits). The data points, which is gene expression levels in this case, are represented as quantum states, manipulated using quantum gates on a quantum computer. This process is central to QSVM's ability to map the data into higher-dimensional spaces efficiently, capturing non-linear relationships that classical SVM might miss. QSVM leverages quantum computing to compute the quantum kernel, mapping data to a higher-dimensional feature space. Since current quantum hardware is limited, QSVM typically uses a hybrid quantum-classical approach, where the quantum part calculates the kernel, and a classical SVM handles classification. This setup combines quantum parallelism with classical optimization. The study applied Qulacs, a high-performance, open-source, quantum computing library. It provides tools for designing and simulating quantum circuits, making it suitable for running quantum machine learning algorithms like QSVM [21]. This library can be used for both testing on quantum simulators and executing on real quantum processors, depending on the available hardware. QSVM has the potential to enhance diagnostic accuracy, particularly in high-dimensional medical datasets. As quantum computing technology matures and hybrid quantum-classical models become more refined, QSVM may play a key role in clinical diagnostics.

IV. EXPERIMENTS AND RESULTS

A. Evaluation Metrics

A set of different metrics is used to evaluate the performance of immune-related gene combinations and machine learning models. Accuracy (Acc) is used to evaluate the performance of machine learning models, particularly in classification tasks. Sensitivity and Specificity metrics are to measure how well the model identifies the positive and negative instances. And Precision metric is to measure the accuracy of positive predictions made by a classification model, specifically focusing on how many of the predicted positive instances are true positives.

B. Differentially Expressed Immune-related Gene Identification

Our primary research focus was on immune-

related genes (IRGs), acknowledging their crucial role in the molecular mechanisms behind sepsis. By analyzing data from multiple platforms, 608 IRGs emerged as potential biomarkers for future studies. A well-established sequential analysis was used to identify differentially expressed genes, including 127 genes with an absolute log-fold change ≥ 1.75 and the p-value ≤ 0.05. A log-fold change of 1.75 represents genes with a significant change in expression levels between sepsis and normal samples. This threshold helps to minimize the inclusion of genes with small, potentially inconsequential fluctuations that might be due to technical variability rather than meaningful biological differences. Meanwhile, the p-value threshold (≤ 0.05) was set to assess the statistical significance of the differential gene expression results. A p-value of less than 0.05 indicates that the likelihood of the observed expression differences occurring by chance is less than 5%, which is a widely accepted standard in genomic research to ensure robust and reproducible results. Notably, the up-regulated genes are those with higher expression in sepsis samples compared to normal samples. The fold change reflects the extent of expression variation between the two groups.

C. Gene Ranking

The gene expression analysis yields 127 genes which are then fed into our selection process. We utilize the RFE feature selection approach to prioritize genes based on their significance. Within this framework, the XGB algorithm coupled with 5-fold cross-validation is executed 10 times to calculate average importance scores for each gene. Following this assessment, we eliminate the gene demonstrating the lowest importance. It should be clarified that in our study, the important metric represents how frequently a gene is chosen to participate in classifier construction via the RFE methodology. We then establish the final gene hierarchy by arranging them in descending order according to their selection frequency.

D. Gene Validation

Using the gene ranking outcomes, we identify the optimal quantity of m genes by testing different classifiers with subsets of ranked genes, beginning with the top-ranked gene and systematically incorporating lower-ranked ones, ultimately retaining the m genes that produce superior classification performance. We evaluate all 127 possible gene combinations to discover the most effective grouping. Three distinct machine learning algorithms are utilized for this assessment, and we generate 127 gene subsets, which correspond to 127 transformed gene combinations (TGCs) that undergo PCA transformation to project them into an alternative data space. It is evident that applying the PCA transformation leads to improved clustering performance in the new data space, which enhances the development of the diagnostic model, as demonstrated in subsequent experiments.

In total, 381 models are created using the TGCs, and their diagnostic efficiency is assessed on the training cohort through five-fold cross-validation. Each model is selected based on its comparison to other classification models among the 127 TGC-based models. Notably, the mean values for the optimal algorithms align with the most effective gene combination identified by the RF classifier, which uses only 41 immune- related genes. Therefore, we select the 41-gene combination as the most informative subset related to sepsis.

D. Performance Evaluation

To evaluate the effectiveness of the minimized gene groups, multiple machine learning models are trained on the training datasets, and its performance is subsequently estimated using a combination of three distinct validation gene sets.

Table II. Performance evaluations

Model	No of genes	Acc (%)	Sn (%)	Sp (%)	Precision (%)
QSVM	41	99.57	99.72	99.00	99.72
LR	19	99.57	99.72	99.00	99.73
NB	13	99.78	100.0	99.00	99.73
SVM-RBF	13	99.78	100.0	99.00	99.73

Table II presents the performance evaluation results of various classical machine learning models and QSVM. The QSVM model demonstrates exceptional performance across all evaluation metrics. With an accuracy of 99.57%, the model reliably predicts sepsis and non-sepsis cases. It achieved a sensitivity rate of 99.72%, ensuring excellent identification of positive cases (sepsis). Additionally, the specificity of 99.00% highlights its capability to correctly identify negative cases (non-sepsis). The precision is 99.72%, which indicates that the model's positive predictions are highly reliable with minimal chances of false positives. These results suggest that QSVM is a robust and effective model for sepsis classification, performing well in terms of accuracy, balanced classification, and reliability. The QSVM model is demonstrated to be a highly reliable tool for achieving optimal diagnostic performance where accurate and balanced classification is essential.

V. CONCLUSION

This paper studies the diagnostic complexities of sepsis, a major global health concern, by leveraging traditional and quantum machine learning methodologies. The proposed strategy features a streamlined gene selection process that emphasizes immune-related genes, integrating data-driven machine learning techniques. A panel of 41 immune-associated marker genes was identified to improve diagnostic precision for sepsis. These genes were selected as the optimal subset from 127 Differentially Expressed Immune Gene Response (DEIGR) combinations, derived from statistical analysis of seven publicly accessible sepsis-related genomic datasets. When applied in conjunction with some machine learning models, including Quantum Support Vector Machine (QSVM) algorithm, this gene panel achieved high diagnostic accuracy and demonstrated strong performance across multiple validation cohorts. Additionally, the robustness and applicability of the sequential gene selection framework and predictive model

were confirmed across three independent gene expression platforms. The numerical validation results—99.57% accuracy, 99.72% sensitivity, 99.00% specificity, and 99.72% precision, highlighting the promising potential of our approach for real-world clinical applications.

REFERENCES

[1] S. L. Weiss, J. C. Fitzgerald, J. Pappachan, D. Wheeler, J. C. Jaramillo- Bustamante, A. Salloo, S. C. Singhi, S. Erickson, J. A. Roy, J. L. Bush, V. M. Nadkarni, and N. J. Thomas, "Global epidemiology of pediatric severe sepsis: The sepsis prevalence, outcomes, and therapies study," American Journal of Respiratory and Critical Care Medicine, vol. 191,pp. 1147–1157, 5 2015.

[2] S. Nemati, A. Holder, F. Razmi, M. D. Stanley, G. D. Clifford, and T. G. Buchman, "An interpretable machine learning model for accurate prediction of sepsis in the icu," Critical Care Medicine, vol. 46, pp. 547–533, 4 2018.

[3] A. Alanazi, L. Aldakhil, M. Aldhoayan, and B. Aldosari, "Machine learning for early prediction of sepsis in intensive care unit (ICU) patients," Medicina (Kaunas), vol. 59, no. 7, Jul. 2023.

[4] Z. Yang, X. Cui, and Z. Song, "Predicting sepsis onset in ICU using machine learning models: a systematic review and meta-analysis," BMC Infect. Dis., vol. 23, no. 1, p. 635, Sep. 2023.

[5] R.-Q. Yao, X. Jin, G.-W. Wang, Y. Yu, G.-S. Wu, Y.-B. Zhu, L. Li, Y.-X. Li, P.-Y. Zhao, S.-Y. Zhu, Z.-F. Xia, C. Ren, and Y.-M. Yao, "A machine learning-based prediction of hospital mortality in patients with postoperative sepsis," Front. Med. (Lausanne), vol. 7, p. 445, Aug. 2020.

[6] P.-E. Thiboud, Q. Franc¸ois, C. Faure, G. Chaufferin, B. Arribe, and N. Ettahar, "Development and validation of a machine learning model for early prediction of sepsis onset in hospital inpatients from all departments," Diagnostics (Basel), vol. 15, no. 3, Jan. 2025.

[7] L. D. Vu, V. S. Pham, M. T. Nguyen, and H.-C. Le, "Pediatric sepsis diagnosis based on differential gene expression and machine learning method," 2022 14th International Conference on Knowledge and Systems Engineering (KSE), pp. 1–6, 10 2022.

[8] J. Ying, Q. Wang, T. Xu, and Z. Lu, "Diagnostic potential of a gradient boosting-based model for detecting pediatric sepsis," Genomics, vol. 113, pp. 874–883, 1 2021.

[9] J. Lu, R. Chen, Y. Ou, Q. Jiang, L. Wang, G. Liu, Y. Liu, B. Yang, Z. Zhou, L. Zuo, and Z. Chen, "Characterization of immune-related genes and immune infiltration features for early diagnosis, prognosis and recognition of immunosuppression in sepsis," International Immunopharmacology, vol. 107, p. 108650, 6 2022.

[10] Y. Yang, Y. Zhang, S. Li, X. Zheng, M. H. Wong, K. S. Leung, and L. Cheng, "A robust and generalizable immune-related signature for sepsis diagnostics," IEEE/ACM Transactions on Computational Biology and Bioinformatics, pp. 1–1, 2021.

[11] R. R. Al-Fatlawy, S. M. Sundaram, E. A. Rathnakumari, M. Sudha et al., "Quantum variational based support vector machine for early detection of sepsis," in 2024 4th International Conference on Mobile Networks and Wireless Communications (ICMNWC). IEEE, 2024, pp. 1–5.

[12] L. McHugh, T. A. Seldon, R. A. Brandon, J. T. Kirk, A. Rapisarda, A. J. Sutherland, J. J. Presneill, D. J. Venter, J. Lipman, M. R. Thomas, P. M. C. K. Klouwenberg, L. van Vught, B. Scicluna, M. Bonten, O. L. Cremer, M. J. Schultz, T. van der Poll, T. D. Yager, and R. B. Brandon, "A molecular host response assay to discriminate between sepsis and infection-negative systemic inflammation in critically ill patients: Dis- covery and validation in independent cohorts," PLOS Medicine, vol. 12, p. e1001916, 12 2015.

[13] B. P. Scicluna, M. A. Wiewel, L. A. van Vught, A. J. Hoogendijk, A. M. Klarenbeek, M. Franitza, M. R. Toliat, P. Nu¨rnberg, J. Horn, M. J. Bonten, M. J. Schultz, O. L. Cremer, and T. van der Poll, "Molecular biomarker to assist in diagnosing abdominal sepsis upon icu admission," American Journal of Respiratory and Critical Care Medicine, vol. 197, pp. 1070–1073, 4 2018.

[14] H. R. Wong, N. Cvijanovich, R. Lin, G. L. Allen, N. J. Thomas, D. F. Willson, R. J. Freishtat, N. Anas, K. Meyer, P. A. Checchia, M. Monaco, K. Odom, and T. P. Shanley, "Identification of pediatric septic shock subclasses based on genome-wide expression profiling," BMC Med., vol. 7, no. 1, p. 34, Jul. 2009.

[15] H. R. Wong, N. Z. Cvijanovich, G. L. Allen, N. J. Thomas, R. J. Freishtat, N. Anas, K. Meyer, P. A. Checchia, S. L. Weiss, T. P. Shanley, M. T. Bigham, S. Banschbach, E. Beckman, K. Harmon, and J. J. Zimmerman, "Corticosteroids are associated with repression of adaptive immunity gene programs in pediatric septic shock," Am. J. Respir. Crit. Care Med., vol. 189, no. 8, pp. 940–946, Apr. 2014.

[16] O. Tabone, M. Mommert, C. Jourdan, E. Cerrato, M. Legrand, A. Lep- ape, B. Allaouchiche, T. Rimmele´, A. Pachot, G. Monneret, F. Venet, A. Mallet, and J. Textoris, "Endogenous retroviruses transcriptional modulation after severe infection, trauma and burn," Front. Immunol., vol. 9, p. 3091, 2018.

[17] M.-A. Cazalis, A. Lepape, F. Venet, F. Frager, B. Mougin, H. Vallin, M. Paye, A. Pachot, and G. Monneret, "Early and dynamic changes in gene expression in septic shock patients: a genome-wide approach," Intensive Care Med. Exp., vol. 2, no. 1, p. 20, Dec. 2014.

[18] H. R. Wong, T. P. Shanley, B. Sakthivel, N. Cvijanovich, R. Lin, G. L. Allen, N. J. Thomas, A. Doctor, M. Kalyanaraman, N. M. Tofil, S. Penfil, M. Monaco, M. A. Tagavilla, K. Odoms, K. Dunsmore, M. Barnes, B. J. Aronow, and Genomics of Pediatric SIRS/Septic Shock Investigators, "Genome-level expression profiles in pediatric septic shock indicate a role for altered zinc homeostasis in poor outcome," Physiol. Genomics, vol. 30, no. 2, pp. 146–155, Jul. 2007.

[19] B. P. Scicluna, P. M. C. Klein Klouwenberg, L. A. van Vught, M. A. Wiewel, D. S. Y. Ong, A. H. Zwinderman, M. Franitza, M. R. Toliat,P. Nu¨rnberg, A. J. Hoogendijk, J. Horn, O. L. Cremer, M. J. Schultz, M. J. Bonten, and T. van der Poll, "A molecular biomarker to diagnose community-acquired pneumonia on intensive care unit admission," Am.J. Respir. Crit. Care Med., vol. 192, no. 7, pp. 826–835, Oct. 2015.

[20] Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., & Scherf, U. "Exploration, normalization, and summaries of high density oligonucleotide array probe level data." Biostatistics, vol. 4, no. 2, pp. 249-264, 2003.

[21] Suzuki, Y., Kawase, Y., Masumura, Y., Hiraga, Y., Nakadai, M., Chen, J., ... & Fujii, K. (2021). "Qulacs: a fast and versatile quantum circuit simulator for research purpose". Quantum, 5, 559.

CHẨN ĐOÁN NHIỄM TRÙNG HUYẾT DỰA TRÊN MÁY VECTƠ HỖ TRỢ LƯỢNG TỬ SỬ DỤNG DỮ LIỆU BIỂU HIỆN GEN

Tóm tắt: Phân tích biểu hiện gen phân biệt trong dữ liệu transcriptome cung cấp cái nhìn sâu sắc về cách thức biểu hiện gen thay đổi để phản ứng với các điều kiện sinh học cụ thể. Một số gen có thể được xác định như các chỉ thị sinh học, được chọn dựa trên những hiểu biết sinh học hiện có hoặc chiến lược dựa trên dữ liệu. Trong nghiên cứu này, chúng tôi đề xuất một phương pháp mới để chẩn đoán nhiễm trùng huyết bằng cách sử dụng dữ liệu gen liên quan đến miễn dịch để xác định sự kết hợp gen hiệu quả nhất làm chỉ thị chẩn đoán. Tổng cộng có 41 gen liên quan đến miễn dịch đã được lựa chọn kỹ lưỡng để hình thành một bảng gen chẩn đoán vững chắc. Phương pháp này bao gồm nhiều bước, bao gồm phân tích biểu hiện gen phân biệt, đánh giá tầm quan trọng của đặc trưng và phương pháp Loại bỏ Đặc trưng Đệ quy (RFE) kết hợp với phương pháp PCA. Áp dụng Quantum Support Vector Machine, kết quả cuối cùng thể hiện hiệu năng giải pháp với kết quả cao ở tất cả các chỉ số đánh giá.

Từ khóa: Gen liên quan đến miễn dịch, Quantum Support Vector Machine, Loại bỏ đặc trưng đệ quy (RFE).

Vu Phuong Thao, received B.Sc. and M.Sc. degree in Economics from Financial University under the Government of the Russian Federation. Now, she is a lecturer at Faculty Interdisciplinary Digital Technologies of Phenikaa University.