Applied Soft Computing Journal 162 (2024) 111781

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Evaluation of conventional and quantum computing for predicting mortality based on small early-onset colorectal cancer data

Jae Yong Yu [a,1], Woo Seob Sim [a,1], Jae Yeob Jung [a], Si Heon Park [b], Han Sang Kim [c,*], Yu Rang Park [a,**]

[a] Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, the Republic of Korea
[b] KAIST, School of Electrical Engineering, Daejeon 34141, the Republic of Korea
[c] Yonsei Cancer Center, Division of Medical Oncology, Department of Internal Medicine, Yonsei University College of Medicine, Seoul, the Republic of Korea

HIGHLIGHTS

- Quantum-inspired computing is utilized to identify the empirical quantum advantage for small sample sizes in healthcare.
- Extreme scenarios of variation in the sample size, feature and outcome imbalance are implemented.
- A quantum algorithm outperformed the conventional algorithm in predicting mortality in early-onset colorectal cancer patients.
- A quantum support vector machine showed robustness even in the imbalanced outcomes

ARTICLE INFO

Keywords:
Empirical quantum advantage
Quantum computing
Digital healthcare
Young early-onset colorectal cancer

ABSTRACT

Background: Quantum computing integrated with machine learning (ML) offers novel solutions in various fields, including healthcare. The synergy between quantum computing and ML in classification exploits unique data patterns. Despite theoretical advantages, the empirical application and effectiveness of quantum computing on small medical datasets remains underexplored.
Method: This retrospective study from a tertiary hospital used data on early-onset colorectal cancer with 93 features and 1501 patients from 2008 to 2020 to predict mortality. We compared quantum support vector machine (QSVM) models with classical SVM models in terms of number of features, number of training sets, and outcome ratio. We evaluated the model based on the area under the curve in the receiver operating characteristic curve (AUROC).
Results: We observed a mortality rate of 7.6 % (96 of 1253 subjects). We generated the mortality prediction model using 11 clinical variables, including cancer stage and chemotherapy history. We found that the AUROC difference between the conventional and quantum methods was the maximum for the top 11 variables. We also showed the AUROC in QSVM (mean [standard deviation], 0.863 [0.102]) outperformed all the number of trials in the conventional SVM (0.723 [0.231]). Compared to the conventional SVM, the QSVM showed robust performance, consistent with the AUROC, even in the unbalanced case.
Conclusion: Our study highlights the potential of quantum computing to improve predictive modeling in healthcare, especially for rare diseases with limited available data. The advantages of quantum computing, such as the exploration of Hilbert space, contributed to the superior predictive performance compared to conventional methods.

Abbreviations: AUROC, Area under the receiver operating characteristic; AUPRC, Area under precision-recall curve; EOCRC, early-onset colorectal cancer; EQA, empirical quantum advantage; ML, machine learning; QML, Quantum-Enhanced ML; QSVM, quantum support vector machine; SDs, standard deviations; SVM, support vector machine; VQC, variational quantum circuit.

* Correspondence to: Yonsei Cancer Center, Division of Medical Oncology, Department of Internal Medicine, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, the Republic of Korea.
** Correspondence to: Department of Biomedical Systems Informatics, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, the Republic of Korea.
E-mail address: yurangpark@yuhs.ac (Y.R. Park).
[1] These authors contributed equally.

https://doi.org/10.1016/j.asoc.2024.111781
Received 20 November 2023; Received in revised form 9 May 2024; Accepted 19 May 2024
Available online 4 June 2024

1. Introduction

Quantum computing is a rapidly growing and innovative field based on the fundamental principles of quantum physics [1,2]. Quantum computing, which uses qubits for processing beyond traditional binary bits, offers exponential speed improvements in computation through superposition and quantum parallelism. This technology, which increases efficiency and reduces storage requirements, is advancing fields such as chemistry, finance, and healthcare by enabling rapid problem solving [3–5]. In addition, unique data patterns in quantum mechanics significantly improve machine learning algorithms for data classification, promising novel insights and advances in a new field called quantum machine learning (QML) [6].

Theoretical work has shown quantum advantages in some areas, but few studies have been conducted to empirically determine whether quantum advantages and feasibility apply to real data sets. The impact of data distribution, such as the number of training data and the ratio of target outcomes, has been studied in the conventional space, but is unclear in the quantum Hilbert space.

Colorectal cancer is the third most common cancer and the second leading cause of cancer-related deaths worldwide [7]. In recent decades, a marked increase in early-onset colorectal cancer (EOCRC) diagnosed in individuals under the age of 50 has been observed [8]. EOCRC cases now account for approximately 10 % of newly diagnosed cases of colorectal cancer. In addition, there has been a notable increase in mortality rates associated with EOCRC [9]. Patients with EOCRC often have advanced tumor stages, leading to an increased risk of death. Postoperative chemotherapy and multi-agent regimens are more common in EOCRC compared to colorectal cancer patients overall [10]. Improved prognostication allows patients to be more effectively stratified and assigned to appropriate treatments.

A sufficient sample size is essential to obtain accurate results and generalize the ML model in the conventional ML field [11]. Research in some medical fields faces the challenges with collecting a sufficient dataset size due to low prevalence and high cost, which is a current challenge in rare diseases [12]. As quantum algorithms exploit the potential of quantum properties, complex problems with limited data can be effectively solved by expanding the analysis space. Caro et al. showed that QML can be generalized well with relatively little training data [13]. However, few studies have analyzed health care problems with small sample sizes in quantum space.

1.1. Research gap

Despite much research on quantum-inspired machine learning for healthcare, there is a gap in understanding the effectiveness of extreme experimental settings on real-world datasets. Previous studies have been conducted using public datasets and without real-world data validation, which is essential for reliable use [3,14]. Although some studies used real-world datasets, they focused on limited experimental settings without considering the data distribution or outcome imbalance. To fill this research gap in the small sample size problem in the medical area, we conducted a quantum-inspired support vector machine to predict in-hospital mortality on the EOCRC dataset. The main contributions of this paper follow three aspects:

(1) We used quantum-inspired computing to identify the empirical quantum advantage for small sample sizes in health care.
(2) Extreme experimental effects are explored based on the number of sample sizes and the imbalance of feature and outcome.
(3) Quantum-inspired computing outperformed the conventional algorithm in predicting mortality in early-onset colorectal cancer patients, in particular showing robustness even in an imbalanced environment.

The rest of this paper is organized as follows. Section 2 summarizes the literature review for QML for healthcare. Section 3 depicts the problem description of quantum computing for small sample sizes. Section 4 gives the experiment details. In Section 5, we analyze the experimental results. Section 6 concludes the work of this paper and gives some directions for future work.

2. Literature review

Many properties of quantum machines are still not well-defined and have been researched using different approaches. Research in QML can be categorized into two mainstream methods; one is a kernel-based, the other is a quantum neural network (QNN)-based.

Schuld et al [15]. presented the basis for using a quantum algorithm for kernel-based method. They focused on the fundamental law of quantum computing in Hilbert space, similar to a high dimensional kernel. They identified mathematical relationships among feature maps, kernel methods, and quantum computing. Furthermore, this work conducted an actual experiment that utilizes a quantum feature map as a support vector machine kernel to build simple benchmark classifiers. Recently, practical studies to investigate and develop QSVM have emerged. Kavitha et al [16]. showed that applying proper feature maps can improve performance by experimenting with varying feature maps. F. Z. Ruskanda [17] introduced QSVM to sentimental classification tasks, which performed better than classical ones; this highlighted that kernel-based QML can be potentially used for natural language process (NLP) tasks. A. Miroszewski et al [18]. used hybrid-QSVM to detect clouds in multispectral satellite images. They observed high accuracy in large datasets with simple quantum kernels, not complex entangled circuits.

The other perspective involves adopting the concepts of conventional deep learning for variational quantum circuit (VQC) structure. VQC comprises several trainable parameterized rotation gates that can calculate gradients or backpropagate it similarly in conventional. Amira Abbas et al. proved that a well-designed QNN has the potential to achieve effective dimensional information and faster training ability than its conventional counterpart [19].

Some application trials have been conducted on quantum-based ML in healthcare [2,20]. Ullah et al [21]. identified quantum random forest (QRF) and QSVM-classified COVID-19 severity and observed the different entanglement and spinning of the gates effects. Krunic et al [22]. attempted to classify the six-month persistence of rheumatoid arthritis by exploiting QSVM. To this end, they designed an experiment of data distribution with consideration for the number of features and sample size. Moreover, this work proposed a metric called the phase space terrain ruggedness index, which is calculated as the numerical ruggedness of the resulting graph representing endurance against reduced sample size and features. Danial et al. used a VQC to predict dementia with various numbers of qubits, ranging from 2 to 5 [23]. In this study, VQC was shown to have robustness concerning a number of features compared to classic SVM. Ghada et al. developed a heart disease prediction model using an ensemble quantum ML approach with Shapley Additive Explanation (SHAP) to make the model explainable [24]. Sünkel et al. conducted quantum transfer learning from a pretrained conventional image classifier to VQC using COVID-19 lung CT images to handle large-scale images [25]. Kanimozhi et al [26]. suggested a hybrid quantum transfer learning method to classify brain MRI images. Despite using small data, they demonstrated that QML may have a quantum advantage in image classification. The summarized table for the literature review for healthcare in QML is shown in Supplementary Table 1.

3. Problem description

3.1. Quantum computing for small sample size

QML, combined with classical ML and quantum computing, has recently been developed and is an attractive field with promising scalability. QML algorithms have been developed to solve complex problems with the potential use of quantum properties such as superposition and entanglement. QML is composed of three parts: encoding, VQC, and measurement. The overall process is illustrated in Fig. 1. Specifically, using kernel method in VQC phase has advantage in two aspects: higher-dimensional data representation in quantum space and utilization of similar mathematic property derived from conventional kernel method. In this paper, we described how to derive kernel function from quantum encoding using simple algebra.

Each data point must be transformed into a quantum state that follows properties of quantum operation. A feature map was used to encode data point $x \in \mathscr{X}$ which originates from the conventional data space $\mathscr{X}$, into feature space $\mathscr{F}$ using functions ϕ. This feature map was defined as $\phi : \mathscr{X} \rightarrow \mathscr{F}$.

Many encoding methods exist, such as amplitude, angle, and higher-order encoding. We experimented with different types of encoders, as listed in Supplementary Table 2, and selected ZZFeatureMap [27] (Fig. 2) based on their accuracy in our case.

When utilizing n-qubits, we employ the Hadamard unitary operation, along with rotation and CNOT gates for entanglement. For the extension to n-qubits, tensor product ($\otimes$)was utilized. We repeated the structure two times to convert classical data into a quantum state. Our feature map denoted as ϕ for conventional input x, is expressed by Eq. (1) as detailed in the original paper [27]:

$$\phi(\lambda(x)) = U(\lambda(x)) \quad \otimes H^n \otimes U(\lambda(x)) \otimes H^n \tag{1}$$

Where x is conventional input, n is number of qubits, and H is Hadamard gate. The unitary operator U and coefficient λ for ith and jth data point are shown in Eqs. (2) and (3). Details of quantum operators are illustrated in Supplementary Table 3.

$$U(\lambda(x)) = \exp\left(i\sum_{S\subseteq[n]}\lambda_S(x)\prod_{i\in S}Z_i\right), \text{where} S\subset\mathbb{R}^d \tag{2}$$

$$\lambda_{\{x_i,x_j\}}(x) = (\pi - x_i)(\pi - x_j) \tag{3}$$

As is broadly known in conventional machine learning concepts, T. Hofmann. et al. [34] introduced a method that maps data points through a kernel to a higher dimension, which can make complex classification problems solvable as linear problem. This can be simply done by calculating the inner product of data points (Details about conventional methods are in the supplementary materials). Schuld et al [15]. proved that the quantum kernel satisfied Mercer's Theorem, a condition for using kernel methods. The quantum kernel for two data points $\{x_i, x_j\}$ is computed by taking the inner product of the quantum feature maps; the quantum kernel $\kappa(x_i, x_j)$ is shown in Eq. (4) given as:

$$\kappa(x_i, x_j) = |\langle\phi(\vec{x_i})|\phi(\vec{x_j})\rangle|^2 \tag{4}$$

Quantum kernel consists of two overlapped feature maps that must be measured. The circuit for the quantum kernel method is shown in Fig. 3.

Finally, trainable measurement operation defines decision boundary based on QSVM circuit for minimizing loss by optimizing the hyperplane parameters. A QSVM with kernels can be used to explore high-dimensional Hilbert spaces efficiently. This kernel function maps our input data to a high-dimensional space in which the data can be precisely separated (Supplementary Figure 1).

4. Experiment setting

This section includes three primary aspects: Section 4.1 sets three experimental designs concerning several features, records variation, and outcome ratio trend. Next, Section 4.2 presents the data collection process, including the primary outcome definition of our study. Sections 4.3 and 4.4 offer preprocessing and statistical methods for comparison and evaluation. Finally, Section 4.5 presents the qubit environment utilized for this study.

4.1. Experimental design

For the experimental task, we considered three different settings to identify quantum applications:

1. The number of feature variations and feature ranking lists from the random forest were used from 3 to 20 (original data), as shown in Supplementary Table 4 and 5.
2. The sample size variations were kept with a fixed original outcome ratio.
3. The outcome ratio ranged from unbalanced (1:8) to balanced (1:1), along with corresponding sample size variations.

4.2. Data collection

The dataset we used for implementing our QSVM was a retrospective dataset on young-age colorectal cancer from tertiary hospitals located in South Korea. The cohort was divided into a development cohort for modeling and an evaluation cohort for validation. Patients aged under 50 years with colon cancer diagnosed between January 2008 and October 2020 were included. Patients within 30 days of follow-up and missing overall disease stage values were excluded. We also excluded patients whose surgical T-values were missing from the operative report. We collected 93 candidate predictors from the database of patients with CRC of the Cancer Registry Library Project, a validated database reviewed by data managers and medical oncologists. Among these predictors, three clinically important features (age, sex, and stage) were selected as fixed variables. This database consists of information on patient demographics, including age and sex; clinical history, such as the initial stage of the disease at diagnosis; histological and molecular pathology data; treatment, such as surgery and systemic chemotherapy;

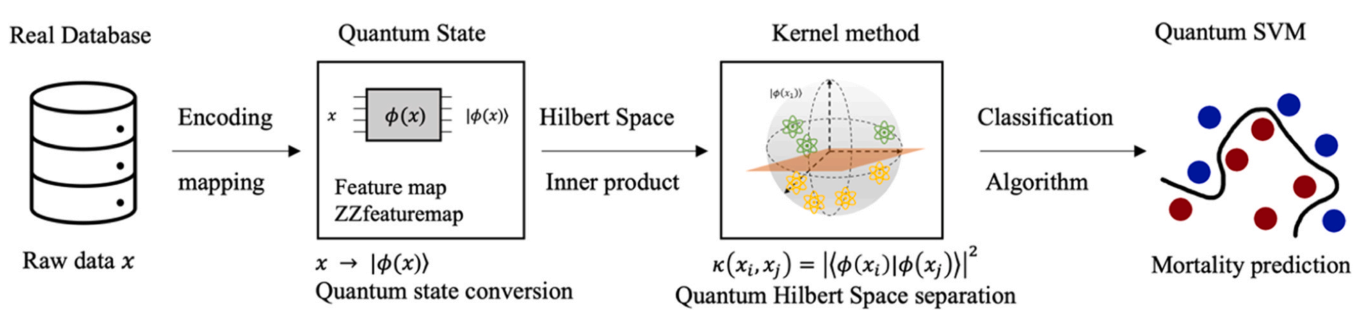

Fig. 1. The overall process of the quantum support vector machine. Early onset colorectal cancer (EOCRC) medical records were prepared. Based on feature map encoding (ZZFeatureMap), data point x can be converted into quantum state $|\phi(x)\rangle$. The kernel method using inner product can map quantum state data to a higher-dimensional Hilbert space. Finally, a hyperplane for classification using support vector machine algorithm is applied to predict mortality in EOCRC.

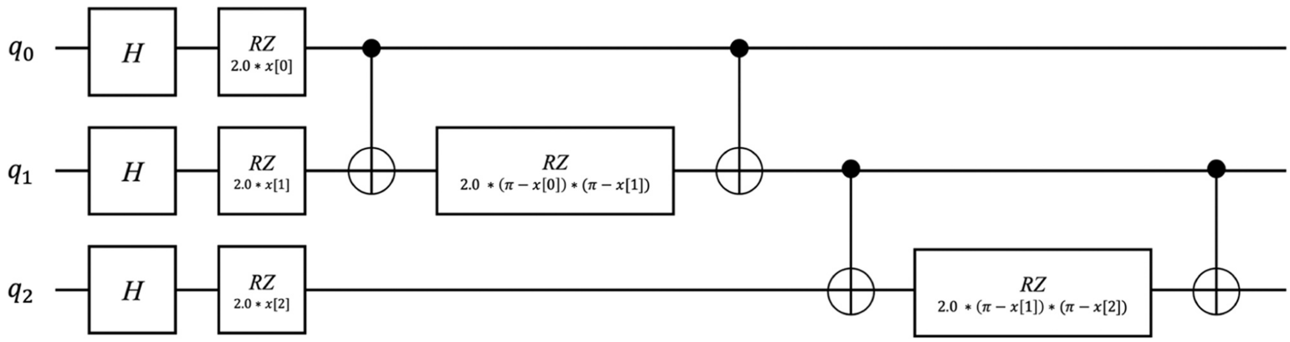

Fig. 2. The quantum circuit of ZZFeatureMap is utilized to convert classical data into a quantum state, utilizing three qubits as examples. For each qubit, we apply the Hadamard gate (*H*) to achieve superposition and the unitary gate operation with Z rotation (*Z*), rotating the Z-axis according to the input data angle. Additionally, the Controlled NOT (CNOT) gate is employed for entanglement.

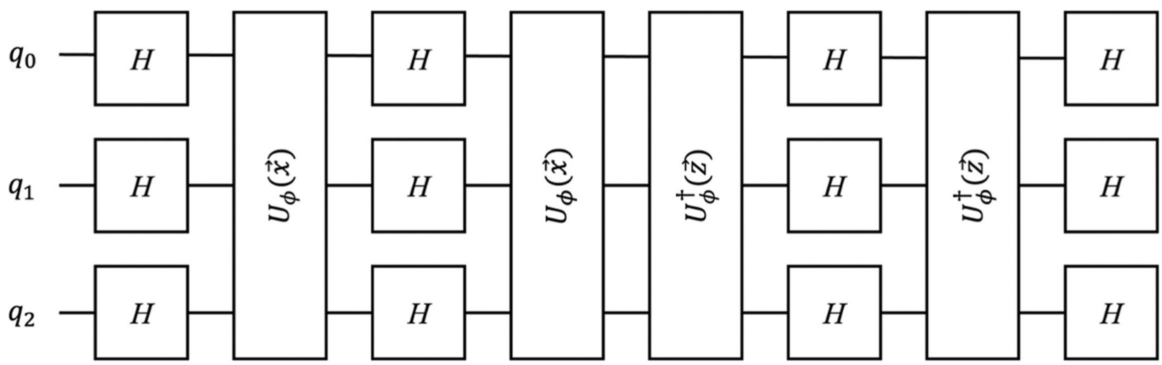

Fig. 3. Quantum circuit for the quantum kernel. The Hadamard gate *H* is considered for optimizing the angles of superposition and unitary operation *U*, where $U^\dagger$ represents the conjugate transpose of *U*. The inner product of the feature map with a depth of 2 is utilized for kernel mapping.

and disease outcomes, such as disease recurrence, progression, and survival. Supplementary Table 4 provides detailed information on the collected variables. The primary outcome was in-hospital mortality, which is vital for the evaluation of young patients with cancer. We used the target features to build the QSVM model.

4.3. Preprocessing

Preprocessing is an essential step in QSVM applications. First, we ensure that our input into the model is appropriately scaled to provide uniformity across all variables in the classification task. To achieve this normalization, we standardize each variable. Feature selection was conducted using random forest and expert opinion to focus on relevant information in a limited qubit environment. The feature ranking list is presented in Supplementary Table 4.

4.4. Statistical analysis

Categorical features are expressed as frequencies with percentages, and continuous features are expressed as means with standard deviations (SDs). Comparison tests between in-hospital mortality cases and others were performed using analysis of variance and chi-square tests at a 5 % significance level. Performance was evaluated using the area under the receiver operating characteristic (AUROC) and the area under the precision-recall curve (AUPRC). A heatmap was used to compare AUROCs based on the two experimental tasks. The quantum-based results were compared with those of conventional ML. To obtain the confidence intervals (CIs), we implemented a 5-fold cross-validation for each metric.

4.5. Qubit environment

Our study was conducted in a simulation setting. Firstly, we used simulator-based experiments with a state vector backend. This backend, sourced from IBM, enables the creation of a qubit environment for research purposes.

5. Results and discussion

Between 2008 and 2020, an initial cohort of 1501 patients diagnosed with EOCRC was identified (Fig. 4). Of these, 22 patients were excluded due to a fewer than 30 observation days, 62 were excluded due to missing whole-stage data, and 164 were excluded due to missing surgical T-stage data. Ultimately, a total of 1253 patients were included in the final analysis, among 96 (8.0 %) died.

Demographic data for patients with EOCRC are shown in Table 1. Mortality cases were significantly associated with advanced tumor stage and younger age of the patients, with rectal cancer being the most frequent location compared to survivors (Table 1). There were 13 (13 %) mortality cases under 30 years old, compared to 40 (3 %) in the survival group. There is no difference in terms of the male portion and the portion of primary tumor surgery between mortality cases and survivors. Regarding the stage at diagnosis, stages III and IV were more common, accounting for approximately 85 (88 %) mortality cases. Most patients in the mortality (n=72 [75 %]) experienced relapse, compared to only 10 % (n=117) of patients in the survival group. There were 43 (45 %) patients who received more than 3rd line of palliative chemotherapy, compared to 50 (4 %) in the survival group.

To predict prognosis in patients with EOCRC, we applied QSVM. Among the three encoding types, ZFeatureMap, Angle encoding, and ZZFeatureMap had AUROC of 0.640, 0.727 and 0.776, respectively. "ZZFeatureMap" achieved the highest AUROC (0.776) under the same

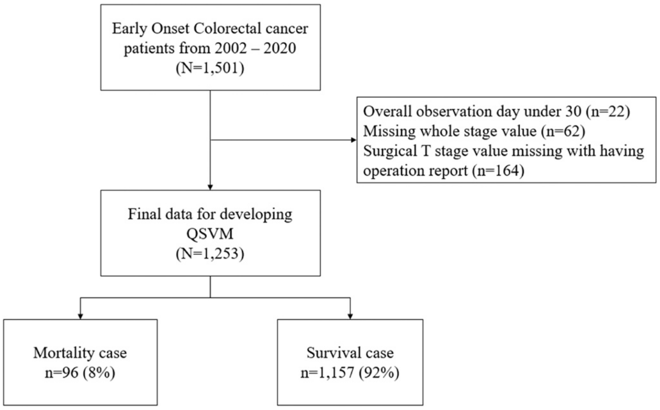

Fig. 4. Overall flowchart of patients with early-onset colorectal cancer for quantum support vector machine.

Table 1
Baseline characteristics of patients with early-onset colorectal cancer.

Variables	Survival case (n = 1157)	Mortality case (n = 96)	P-value
Age (years)			<0.001
40–49	828 (72 %)	50 (52 %)	
30–39	289 (25 %)	33 (34 %)	
20–29	38 (3 %)	11 (11 %)	
<20	2 (<1 %)	2 (2 %)	
Sex			1.000
Male	593 (51 %)	49 (51 %)	
Female	564 (49 %)	47 (49 %)	
CRC site			<0.001
Colon	806 (70 %)	21 (22 %)	
Rectum	351 (30 %)	75 (78 %)	
Stage at diagnosis			<0.001
Stage 0	22 (2 %)	1 (1 %)	
Stage I	345 (30 %)	3 (3 %)	
Stage II	284 (25 %)	7 (7 %)	
Stage III	414 (36 %)	28 (29 %)	
Stage IV	92 (8 %)	57 (59 %)	
Surgery on primary tumor			0.926
Yes	832 (72 %)	70 (73 %)	
No	325 (28 %)	26 (27 %)	
Adjuvant chemotherapy			
Relapse	117 (10 %)	72 (75 %)	<0.001
Line of palliative chemotherapy			<0.001
1st line	422 (36 %)	7 (7 %)	
2nd line	47 (4 %)	23 (24 %)	
3rd line	28 (2 %)	13 (14 %)	
more than 4th line	22 (2 %)	30 (31 %)	

CRC, colorectal cancer; CI, confidence interval.; p-value was calculated with t-test for continuous variable and chi-square test for categorical variables.

conditions and was selected as the encoder for subsequent experiments (Supplementary Table 2). Based on the number of features as shown in Fig. 5 and Supplementary Table 5, the saturation of AUROC was found at the top 11 variables. These variables included the cancer stage, the number of lymph nodes examined after surgery, and the history of chemotherapy, with maximum difference between conventional performances (0.863 of AUROC in quantum and 0.691 of AUROC in conventional SVM).

Among the variables, time to recurrence from diagnosis exhibited the highest importance, indicating its significant role in mortality prediction. A shorter time from initial diagnosis to recurrence is associated with poorer survival outcomes. We observed enhanced model performance upon incorporating the 10th variable (time to 4th line chemotherapy from diagnosis, which yielded an increase of 0.108 in AUROC) and the 11th variable (the number of tumor-positive lymph nodes, resulting in an increase of 0.085 in AUROC). In conventional SVM, age at initial cancer diagnosis was the features that contributed most (0.18 in AUROC). The included variables are listed in Supplementary Table 5. These findings indicate that QSVM may enhance prognostic accuracy for patients with EOCRC, a condition that is rare but requires enhanced postoperative monitoring to prolong life expectancy, in contrast to conventional SVM methods.

Next, we evaluated whether QSVM can predict the prognosis with a smaller number of subjects in EOCRC. Interestingly, QSVM showed better performance in predicting the prognosis with a higher AUROC with a smaller training set, compared to conventional AUROC (mean [SD], 0.852 [0.028] vs. 0.716 [0.029] with conventional SVM), suggesting that QSVM can improve prediction with a smaller number of training set in rare disease. Both the conventional and quantum AUROCs were robust based on the amount of data (Fig. 6 and Supplementary Table 6). AUPRC for conventional (0.408 [0.031]) and quantum SVM (0.406 [0.054]) was similar across variations in sample size.

The frequency of mortality events in a disease can impact the performance of a model in predicting prognosis. Generally, the performance of the model tends to decrease when predicting rarer events. Therefore, we assessed whether QSVM could effectively predict prognosis even with fewer mortality events in the training set. As expected, the model performance decreased using the conventional SVM, and the AUROC for the conventional SVM decreased by 10 % from 0.895 [0.013] to 0.802 [0.005], as the imbalance ratio of the morality cases over the total number of subjects increased from 1 to 8. In contrast, QSVM can predict the prognosis with fewer mortality events in the training set (Fig. 7). In the range of 1:2–1:4, both the quantum and the conventional algorithm had similar performance; however, from the range of 1:5 and above, quantum performed significantly better. Overall, the values of AUPRC were higher for the conventional SVM than for QSVM.

Compared to conventional SVM, QSVM showed robust AUROC performance from 0.839 [0.053] to 0.869 [0.009], which is consistent even in the unbalanced case (Fig. 7, Supplementary Table 7 and Supplementary Figure 2). Taken together, QSVM can enhance the mortality prediction in patients with EOCRC, with fewer training set and rare mortality events, compared to the conventional SVM.

In this study, a quantum-based SVM was developed to predict EOCRC mortality. The data space in a quantum state can be mapped to a higher Hilbert space, which deals with complex numbers, instead of a conventional space. This could be advantageous in terms of solving complicated problems such as clinical decision problems, with the kernel method in QSVM. We calculated the hyperplane in the Hilbert space for the discriminative task. The implemented QSVM was validated well in the EOCRC domain, which is empirically advantageous for quantum-based SVM. The quantum-based results were more accurate than those of the conventional SVM, which had an AUROC of 0.863 based on 11 clinical features. This is the first quantum computing study to infer mortality in patients with EOCRC. We found saturation of AUROC performance at the top 11 importance features in QSVM. Those features include days from diagnosis to recurrence, initial cancer stage, second line chemotherapy related information, lymph node count, age, relapse which is critical in predicting mortality in colon cancer which showed consistent results from a previous stud [28]. We also found total cycles of line chemotherapy had positive performance gain in quantum which can be the potential risk factor in higher dimensional space. Age and relapse status had positive performance in conventional algorithm, which highlights a key distinction between quantum and conventional performance gain.

Improved prediction of mortality in the EOCRC addresses significant unmet medical needs. Firstly, EOCRC survivors typically have a longer life expectancy and thus require extended post-treatment surveillance compared to older colorectal cancer patients. Secondly, the rising incidence of EOCRC necessitates a heightened focus on treatment strategies for this group. Thirdly, EOCRC is often more aggressive than late-onset colorectal cancer. Enhancing the prognosis for these rare events could significantly extend the life expectancy of EOCRC patients. Furthermore, in healthcare settings, QSVM have demonstrated superior performance

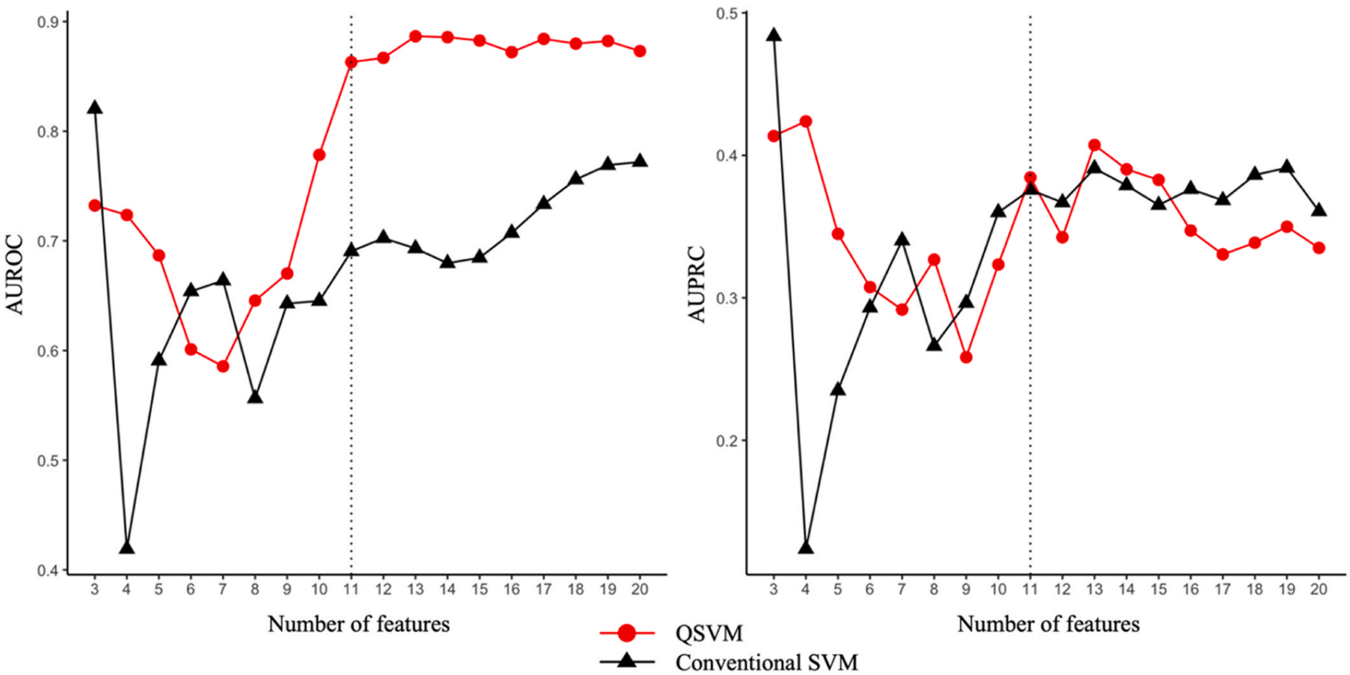

Fig. 5. The area under receiver operating characteristics curve for conventional (black) and quantum (red) support vector machines is plotted based on the number of feature variations. The dotted line at the top 11 variable utilizations, indicating where the performance difference between conventional and quantum support vector machines was maximal and saturated for the quantum support vector machine.

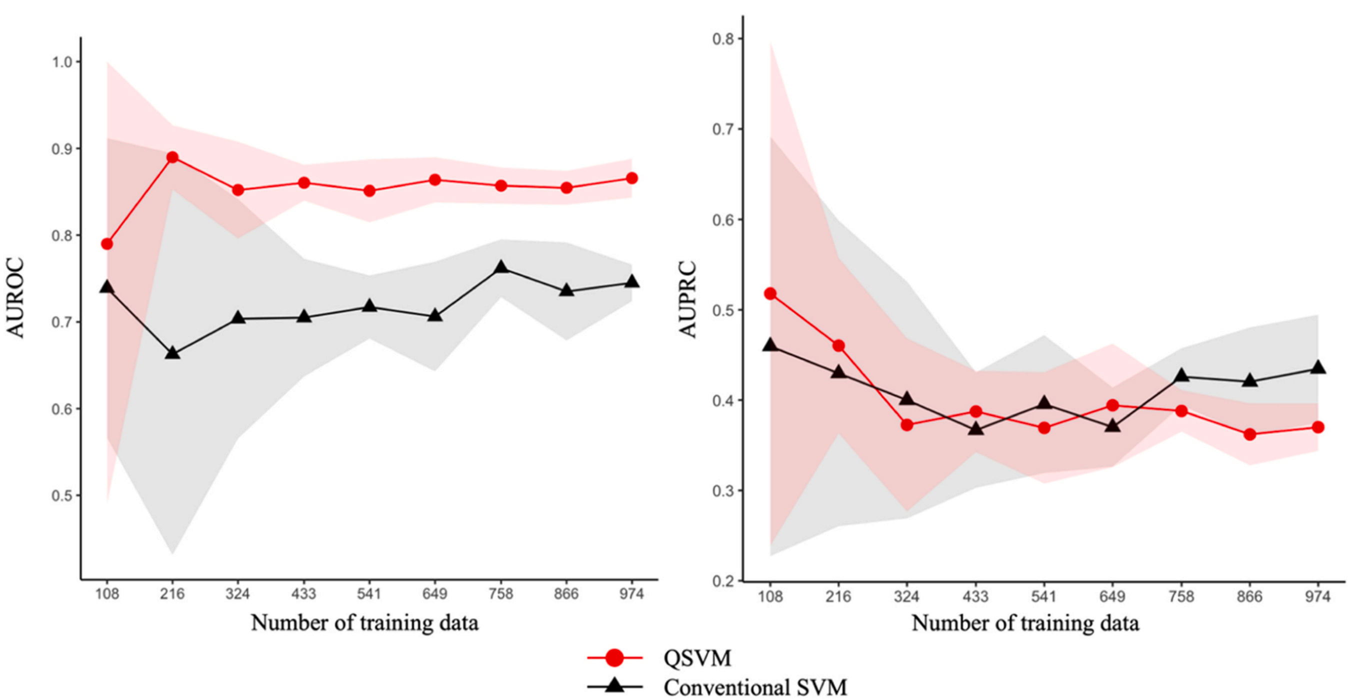

Fig. 6. Area under receiver operating characteristics curve performance of conventional (black) and quantum (red) support vector machines based on the number of sample size variations. The quantum support vector machine outperformed all sample size experiments and showed robustness regardless of sample size variation.

in modeling rare events within rare diseases. In healthcare, only a small number of patients and a vast number of features can be collected in rare diseases such as EOCRC. Quantum-enhanced analysis can propose the potential utilization of different points of view. Data is transformed from the current dimension to a higher dimension to overcome the problem of nonlinearity in SVM. For this conversion, kernel functions, such as polynomial, Gaussian, and radius basis function kernels, are utilized to map real data numbers to higher-dimensional feature spaces and calculate nonlinear hyperplanes to classify complicated datasets. The analysis of high-dimensional space can effectively predict the progression of the disease in terms of sensitive and specific classification results [29]. Similar to classical kernel-based ML, in any task in quantum computing, each data point can be mapped to a complex number dimension Hilbert space using a feature map. The converted data in the Hilbert space can be separated well from conventional higher dimensions. Consistent with previous studies [30,31], we also experienced

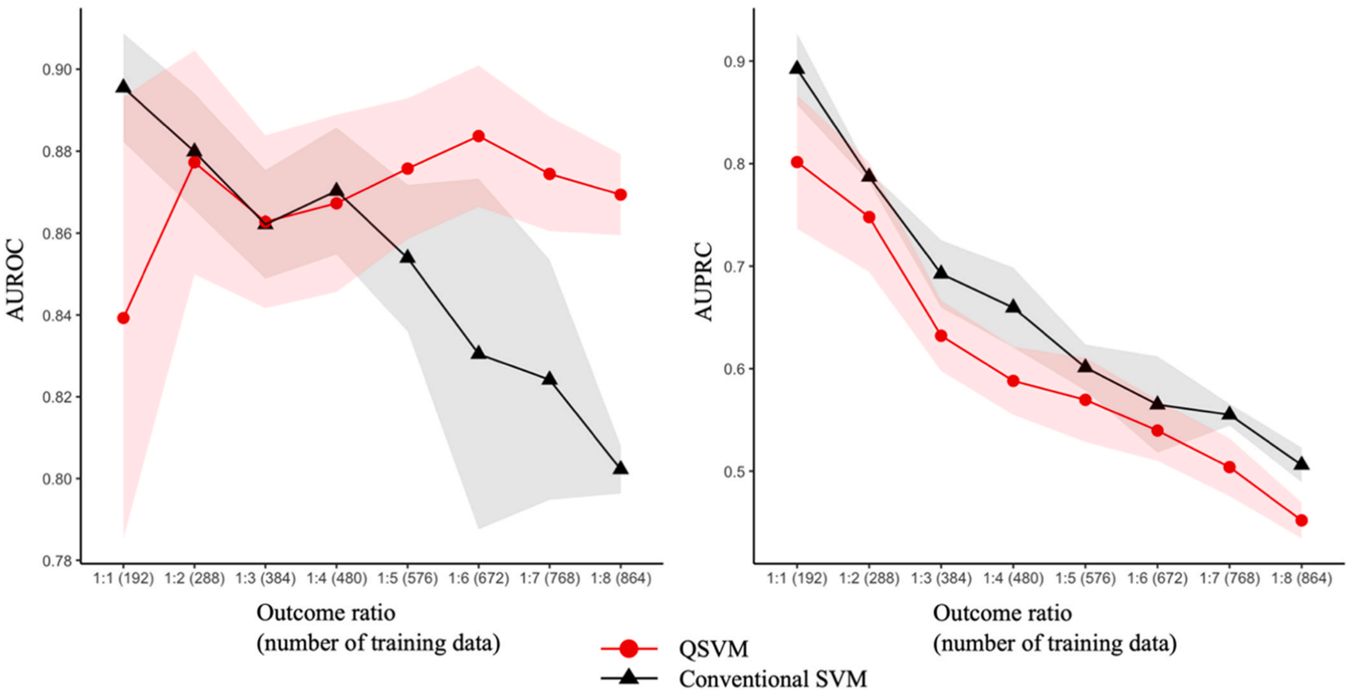

Fig. 7. Area under the receiver operating characteristics curve performance of conventional (black) and quantum (red) support vector machines based on the number of outcome ratios. The performance of the conventional algorithm decreased with larger differences in outcome ratio, whereas the quantum support vector machine showed robustness even in an unbalanced environment.

advantages from accessing Hilbert spaces with higher dimensions. Information from high-level views can result in better performance than low-level ones.

In QML, the applications that benefit from the power of quantum tasks are unclear. Huang et al. suggested a flowchart to understand and prescreen potential quantum advantages [6]. Which circuit is suitable for each task remains an open question in the quantum field. Our study showed that EOCRC tabular data can be predicted well in quantum compared to conventional ML using ZZFeatureMap and a quantum circuit for the kernel. Many current ML models can be transformed into QML schemes to use the advantages of quantum space and are evaluated for future quantum advantages. After the quantization, effectiveness of quantum model can be measured based on fidelity and trace distance. However, data-level comparison with conventional space is still challenging.

Contrary to the results of conventional ML, our novel method exhibited robust performance, even on an unbalanced dataset. Conventional SVMs have minimal performance when applied to unbalanced data, where negative cases severely outnumber positive ones [32]. Amiri et al. also suggested the consistent impact of data imbalance, determining that independently and nonidentically distributed data have a negative impact on performance [33]. Although conventional SVM is more accurate regarding balance, most real-world data have a skewed outcome distribution. QSVM can be useful for real-world data validation.

This study had some limitations. First, only a small amount of data was considered environmental resources. Second, this study was conducted in a simulation setting. Real devices can operate on extremely small sample sizes with quantum noise. In the future, we plan to conduct experiments using real devices.

6. Conclusion and future recommendation

This paper identified the power of quantum computing in healthcare problems with small sample sizes. Unlike conventional research on quantum computing for healthcare, our study conducted several experiments, including sample size, feature size, and outcome ratio, for small sample size healthcare problems. The advantages of Hilbert space were employed in the quantum computing algorithm. The algorithm outperformed conventional algorithms regarding AUROC values in various experimental settings on the EOCRC dataset. Different numbers of qubits and training sample sizes were also assessed in various aspects. In particular, QSVM showed robust performance even in the unbalanced case, whereas conventional methods showed a poorer performance. The results were better than those achieved by SVM method with less training data. According to the results, implementing quantum computing for the small sample size problem could provide sufficient evidence. Moreover, the quantum computing algorithm can be expanded to various healthcare problems with small sample sizes using VQC-based quantum algorithms. Real quantum device applications are another potential next step.

Funding

This research was supported by the National Research Foundation funded by the Korean government (MSIT) (No. RS-2023-00261820) and Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE) (No. P0023675, HRD Program for Industrial Innovation).

CRediT authorship contribution statement

Yu Rang Park: Writing – review & editing, Supervision, Methodology, Investigation, Data curation, Conceptualization. **Han Sang Kim:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Si Heon Park:** Investigation. **Woo Seob Sim:** Writing – original draft, Methodology, Formal analysis, Data curation. **JaeYong Yu:** Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jae Yeob Jung:** Investigation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgment

Nothing declared.

Conflict of Interest

The authors have no conflict of interests.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.asoc.2024.111781.

References

[1] D. Solenov, J. Brieler, J.F. Scherrer, The potential of quantum computing and machine learning to advance clinical research and change the practice of medicine, Mo Med. 115 (2018) 463–467.
[2] R. Ur Rasool, H.F. Ahmad, W. Rafique, A. Qayyum, J. Qadir, Z. Anwar, Quantum computing for healthcare: a review, Future Internet 15 (2023) 94.
[3] Y. Kumar, A. Koul, P.S. Sisodia, J. Shafi, V. Kavita, M. Gheisari, M.B. Davoodi, Heart failure detection using quantum-enhanced machine learning and traditional machine learning techniques for internet of artificially intelligent medical things, Wirel. Commun. Mob. Comput. 2021 (2021) 1–16.
[4] N. Heidari, S. Olgiati, D. Meloni, F. Pirovano, A. Noorani, M. Slevin, L. Azamfirei, A quantum-enhanced precision medicine application to support data-driven clinical decisions for the personalized treatment of advanced knee osteoarthritis: development and preliminary validation of precisionKNEE_QNN, medRxiv (2021), 2021.2012. 2013.21267704.
[5] K. Bharti, T. Haug, V. Vedral, L.-C. Kwek, Machine learning meets quantum foundations: a brief survey, AVS Quantum Sci. 2 (2020).
[6] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, S. Lloyd, Quantum machine learning, Nature 549 (2017) 195–202.
[7] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J. Clin. 71 (2021) 209–249.
[8] F.A. Sinicrope, Increasing incidence of early-onset colorectal cancer, N. Engl. J. Med. 386 (2022) 1547–1558.
[9] R.L. Siegel, K.D. Miller, A. Goding Sauer, S.A. Fedewa, L.F. Butterly, J.C. Anderson, A. Cercek, R.A. Smith, A. Jemal, Colorectal cancer statistics, 2020, CA Cancer J. Clin. 70 (2020) 145–164.
[10] P.J. Kneuertz, G.J. Chang, C.Y. Hu, M.A. Rodriguez-Bigas, C. Eng, E. Vilar, J. M. Skibber, B.W. Feig, J.N. Cormier, Y.N. You, Overtreatment of young adults with colon cancer: more intense treatments with unmatched survival gains, JAMA Surg. 150 (2015) 402–409.
[11] D. Rajput, W.-J. Wang, C.-C. Chen, Evaluation of a decided sample size in machine learning applications, BMC Bioinforma. 24 (2023) 48.
[12] A. Vabalas, E. Gowen, E. Poliakoff, A.J. Casson, Machine learning algorithm validation with a limited sample size, PLoS One 14 (2019) e0224365.
[13] M.C. Caro, H.-Y. Huang, M. Cerezo, K. Sharma, A. Sornborger, L. Cincio, P.J. Coles, Generalization in quantum machine learning from few training data, Nat. Commun. 13 (2022) 4919.
[14] H. Heidari, G. Hellstern, Early heart disease prediction using hybrid quantum classification, arXiv preprint arXiv:2208.08882, (2022).
[15] M. Schuld, N. Killoran, Quantum machine learning in feature hilbert spaces, Phys. Rev. Lett. 122 (2019) 040504.
[16] S. Kavitha, N. Kaulgud, Quantum machine learning for support vector machine classification, Evolut. Intell. (2022) 1–10.
[17] F.Z. Ruskanda, M.R. Abiwardani, R. Mulyawan, I. Syafalni, H.T. Larasati, Quantum-enhanced support vector machine for sentiment classification, IEEE Access (2023).
[18] A. Miroszewski, J. Mielczarek, G. Czelusta, F. Szczepanek, B. Grabowski, B.Le Saux, J. Nalepa, Detecting clouds in multispectral satellite images using quantum-kernel support vector machines, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. (2023).
[19] A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, S. Woerner, The power of quantum neural networks, Nat. Comput. Sci. 1 (2021) 403–409.
[20] D. Maheshwari, B. Garcia-Zapirain, D. Sierra-Sosa, Quantum Machine Learning Applications in the Biomedical Domain: A Systematic Review, Ieee Access, 2022.
[21] U. Ullah, D. Maheshwari, H.H. Gloyna, B. Garcia-Zapirain, Severity classification of COVID-19 patients data using quantum machine learning approaches. in: 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), IEEE, 2022, pp. 1–6.
[22] Z. Krunic, F.F. Flöther, G. Seegan, N.D. Earnest-Noble, O. Shehab, Quantum kernels for real-world predictions based on electronic health records, IEEE Trans. Quantum Eng. 3 (2022) 1–11.
[23] D. Sierra-Sosa, J. Arcila-Moreno, B. Garcia-Zapirain, C. Castillo-Olea, A. Elmaghraby, Dementia prediction applying variational quantum classifier, arXiv preprint arXiv:2007.08653, (2020).
[24] G. Abdulsalam, S. Meshoul, H. Shaiba, Explainable heart disease prediction using ensemble-quantum machine learning approach, Intell. Autom. \ Soft Comput. 36 (2023) 761–779.
[25] L. Sünkel, D. Martyniuk, J.J. Reichwald, A. Morariu, R.H. Seggoju, P. Altmann, C. Roch, A. Paschke, Hybrid quantum machine learning assisted classification of COVID-19 from computed tomography scans. 2023 IEEE International Conference on Quantum Computing and Engineering (QCE), IEEE, 2023, pp. 356–366.
[26] T. Kanimozhi, S. Sridevi, T.S. Manikumar, T. Dheeraj, A. Sumanth, Brain tumor recognition based on classical to quantum transfer learning. 2022 International Conference on Innovative Trends in Information Technology (ICITIIT), IEEE, 2022, pp. 1–5.
[27] N. Thumwanit, C. Lortaraprasert, R. Raymond, Invited: trainable discrete feature embeddings for quantum machine learning, 2021 58th ACM/IEEE Des. Autom. Conf. (DAC) (2021) 1352–1355.
[28] S. Alinia, M. Asghari-Jafarabadi, L. Mahmoudi, G. Roshanaei, M. Safari, Predicting mortality and recurrence in colorectal cancer: Comparative assessment of predictive models, Heliyon 10 (2024).
[29] Y. Wang, Y. Fan, P. Bhatt, C. Davatzikos, High-dimensional pattern regression using machine learning: from medical images to continuous clinical variables, Neuroimage 50 (2010) 1519–1535.
[30] D. Cozzolino, B. Da Lio, D. Bacco, L.K. Oxenløwe, High-dimensional quantum communication: benefits, progress, and future challenges, Adv. Quantum Technol. 2 (2019) 1900038.
[31] J.Y. Chan, A.P. Leung, Y. Xie, Efficient high-dimensional Kernel k-Means++ with random projection, Appl. Sci. 11 (2021) 6963.
[32] R. Akbani, S. Kwek, N. Japkowicz, Applying support vector machines to imbalanced datasets, in: Machine Learning: ECML 2004: 15th European Conference on Machine Learning, Pisa, Italy, September 20-24, 2004. Proceedings 15, Springer, 2004, pp. 39-50.
[33] S. Amiri, A. Belloum, E. Nalisnick, S. Klous, L. Gommans, On the impact of non-IID data on the performance and fairness of differentially private federated learning, in: 2022 52nd Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W), 2022, pp. 52-58.