Feature Selection and Feature Extraction for Satellite Hyperspectal Imagery Data

Abstract

This section explores the utility of quantum algorithms in the tasks of feature selection and feature extraction of hyperspectral satellite data. The feature selection approaches discussed include the quantum and hybrid extensions of the classification-based RFE (Recursive Feature Eliminati on) algorithm, as well as an approach formulated as the QUBO (Quantum Unconstrained Binary Optimization) problem. The application of the quantum variational method is explored in both feature selection and feature extraction. Additionally, the feature extraction techniques of quantum PCA (Principal Component Analysis) and quantum autoencoders are analyzed. The practical implementation and hardware requirements for each of these approaches are discussed.

25 Introduction

Feature selection and feature extraction are common methods for reducing the number of features in large, high-dimensional data sets. A basic distinction between these methods is that the first involves transforming the original features, while the second preserves the features.

The procedures have profound practical consequences, allowing for more effective data storage, transferring, reduction, and analysis. The hyperspectral data satellite data, with even hundreds of narrow spectral bands, provide an example of the area in which utilization of the methods seems virtually unavoidable. The rich spectral information may simply surpass the needs of certain applications.

On the other hand, since the number of possible selections (subsets) grows exponentially with the number of features, the application of the selection methods involves hard optimization tasks. Within the use case, the possibility of applying quantum algorithms to improve the selection methods is examined. The discussion focuses on hyperspectral imagery data as a promising but extremely challenging data type in the field of satellite Earth imagery.

25.1 About hyperspectral imaging

Hyperspectral satellite imaging is a remote sensing technology that captures images of the Earth's surface at high spectral resolution. Unlike traditional satellite imagery, which captures images in several discrete broad spectral bands (usually red, green, blue, and bands in the infrared wavelength spectrum), hyperspectral imaging captures images in hundreds of narrow and contiguous spectral bands, providing a detailed and comprehensive view of the Earth's surface Chang [2003]; Grahn and Geladi [2007]; Lv and Wang [2020].

Each pixel in a hyperspectral image contains information about the reflectance or emission of electromagnetic radiation at each of the spectral bands. This information can be used to identify and map the distribution of different materials and substances on the Earth's surface, such as vegetation, minerals, water, and pollutants.

Hyperspectral satellite imaging has numerous applications in fields such as environmental monitoring Moroni et al. [2013], agriculture Nalepa et al. [2022], mineral exploration Booysen et al. [2022], and urban planning Karoui et al. [2019]. It can provide valuable information for resource management, land-use planning, and environmental protection, among other purposes.

Some examples of hyperspectral satellite missions include:

- Hyperion: Launched in 2000, Hyperion was a hyperspectral imaging instrument carried aboard the Earth Observing-1 (EO-1) satellite. It provided high-resolution hyperspectral data for a variety of applications, including land use and land cover classification, mineral mapping, and more.
- EnMAP: The Environmental Mapping and Analysis Program (EnMAP) is a German hyperspectral satellite mission that was launched in 2021. Its primary goal is to monitor the Earth's environment, including vegetation, water bodies, and urban areas, to support a range of environmental applications.
- PRISMA: The Italian Space Agency's Precursore Iperspettrale della Missione Applicativa (PRISMA) satellite mission was launched in 2019. Its primary objective is to provide high-resolution hyperspectral data for a variety of applications, including agriculture, forestry, and natural resource management.
- HyspIRI: The Hyperspectral Infrared Imager (HyspIRI) is a proposed NASA mission that would carry a hyperspectral sensor suite to monitor the Earth's terrestrial ecosystems, carbon cycle, and natural hazards. The mission is currently in the planning phase and is expected to launch in the mid-2020s.

The rapid growth of hyperspectral imagery product availability translates also to the technical challenges connected with data processing and analysis:

The increased spectral resolution of hyperspectral images allows for more precise identification and analysis of different materials and features within an image, but also results in a much larger amount of data being generated. The size of a hyperspectral image can be several times larger than that of a multispectral image covering the same area, which can present challenges for data storage and processing.

The construction of contiguous spectral coverage between 400 and $2500nm$ with a spectral sampling interval (of several to tens of nm) Rast and Painter [2019], such that the adjacent channels overlap, makes the information carried by the spectral bands strongly correlated and hence partially redundant. This makes pattern recognition tools for hyperspectral data susceptible to the Hughes phenomenon Hughes [1968]; Theodoridis and Koutroumbas [2006]; Ma et al. [2013]. The Hughes phenomenon occurs when the classification accuracy increases gradually in the beginning as the number of spectral bands or dimensions increases but decreases dramatically when the band number reaches some value. It is a manifestation of the wider phenomenon, collectively called the 'curse of dimensionality' Bellman [1957]; Bishop and Nasrabadi [2006].

The above motivates the use of data reduction techniques for hyperspectral data. In the following, we investigate the possibility and feasibility of quantum-assisted feature selection and feature extraction methods.

26 Feature selection

Complex datasets, such as the introduced hyperspectral images, may contain much more information than is actually needed for a specific application. Furthermore, due to the spectrum's dense band coverage, significant data redundancy may be present. Consequently, data analysis is laborious and may exhibit undesirable consequences, such as the Hughes phenomenon or low interpretability.

Feature selection is a common strategy in data analysis to overcome these difficulties. The price to pay is, however, the high computational cost of the feature selection procedure.

In what follows, we review four possible paths of applying quantum computing methods which promise to improve the classical feature selection methods. We begin with the possible quantum extensions of the Recursive Feature Elimination (RFE). This will mainly concern the employed quantum and hybrid Support Vector Machines (SVM) algorithms. After that, a relation with the quantum kernel methods will be discussed. Another branch of the methods employs quantum optimization, considered one of the most promising applications of the NISQ-era quantum computers. Finally, the utility of quantum variational methods will be discussed.

26.1 Recursive Feature Elimination (RFE) based on quantum/hybrid SVM

Recursive Feature Elimination (RFE) Guyon et al. [2002] is a systematic feature selection algorithm that enables the identification of the most important variables within a given dataset. This technique iteratively eliminates the least significant features and re-calibrates the model with the remaining ones until the optimal number of features is achieved.

By systematically removing the least relevant features during each iteration, RFE facilitates the identification of the feature subset that is most adequate to the model's predictive performance. Consequently, this results in the development of more effective and precise models, as well as a simplified interpretation of the model's outcomes.

RFE can be employed with a diverse range of machine learning models, including linear and logistic regression, Support Vector Machines (SVM), and random forests, amongst others. This technique can be a powerful tool to improve the efficiency and effectiveness of predictive models in numerous applications. In what follows, we will focus on the SVM-based approach to the RFE. The reason for this is that, already in 2014, a promising quantum extension of the SVM algorithm was proposed Rebentrost et al. [2014] — the QSVM (Quantum Support Vector Machine).

The SVM algorithm consists of two main steps. The first is a calculation of the so-called *Kernel function* $K(\mathbf{x}_i, \mathbf{x}_j)$, which is a generalization of the scalar product between the data vectors, $\mathbf{x}_i \in \mathbb{R}^n$, where n is the dimension of the data under consideration. The second step uses the kernel function to solve a system of equations that determine the position of the hyperplane of the decision (separating the two classes of data). In the classical version of SVM, the computational complexity of the algorithm is polynomial in both the number of vectors of the training data (m) and their dimension (n): $\mathcal{O}(m^2(n+m))$. It is proven that for QSVM, the computational complexity reduces to logarithmic in both m and n: $\mathcal{O}(\log nm)$ Rebentrost et al. [2014], which is an example of the so-called *exponential speedup*.

The first step of the QSVM algorithm can be performed using a quantum register with a number of logical qubits equal to n (which corresponds to the dimension of the data). Therefore, in light of the current and expected in the near future quantum computers, the method is suitable for multispectral and even hyperspectral data with dozens of wavebands. Further improvement can be achieved under the dense embedding of the classical data vector on the quantum register. This may ultimately require $\log_2 n$ qubits for n hyperspectral bands. In this limiting case, only eight logical qubits are sufficient to encode 256 spectral bands. Ultimately, future quantum computers operating on hundreds of noisy physical qubits may allow the direct quantum kernel function determination for hyperspectral data satellites.

However, the second QSVM step, equivalent to finding the inverse of a matrix of dimension $(1+m)\times(1+m)$, requires using a number of qubits that scales linearly with the number of training data vectors. Given that in typical cases, the training of the dataset contains up to several hundred thousand data vectors, performing this step is not possible using NISQ technology. In other words, the NISQ technology will allow one to perform this step only for very small ($m \sim 10^1 - 10^2$) training data sets, which is uninteresting from a practical point of view. Therefore, the application of the full, original QSVM algorithm for the purpose of the RFE procedure is impractical from the perspective of the NISQ technology. The QSVM-based RFE for the satellite hyperspectral data may, however, be implemented in the more distant future, say in more than fifteen years.

Earlier, say in 5 years, hybrid methods, in which the first step of the SVM algorithm is quantum-aided, while the second step utilizes classical computing resources, have a chance to be implemented successfully. Furthermore, besides the expected reduction of the computational complexity, quantum evaluation of the kernel function opens up the possibility of improving the classification accuracy and, in consequence, the selection of the feature. This is because quantum algorithms give rise to a whole new family of kernel functions. Some of them may outperform the considered classical ones in classification accuracy. Furthermore, they may turn out to be hard to emulate on classical machines, justifying the use of quantum computers. We discuss the quantum kernel-based approach in the next subsection.

26.2 Feature selection based on quantum kernel methods

There are various possible maps from the original data space, say $\mathbb{R}^n$ to the Hilbert space $\mathcal{H} = (\mathbb{C}^2)^{\otimes n'}$ in which a state of n'-qubit quantum register is defined. Let the map, which we call a quantum feature map, be defined as

$$\mathbf{x} \to |\Phi(\mathbf{x})\rangle = \hat{\mathcal{U}}_{\Phi(\mathbf{x})}|0\rangle^{\otimes n'}, \tag{14}$$

where $|0\rangle^{\otimes n'}$ is an initial state of the quantum register and $\hat{\mathcal{U}}_{\Phi(\mathbf{x})}$ is the unitary quantum operator that defines the quantum map.

Therefore, the kernel function is determined as the following square of the amplitude modulus probability:

$$K(\mathbf{x}_i, \mathbf{x}_j) = |\langle \Phi(\mathbf{x}_i)|\Phi(\mathbf{x}_j)\rangle|^2. \tag{15}$$

In the 2019 article, Havlíček et al. [2019], a promising proposal for the kernel function has been studied. The corresponding quantum feature map was given by the operator $\hat{\mathcal{U}}_{\Phi(\mathbf{x})} = (\hat{U}_{\Phi(\mathbf{x})}\hat{H}^{\otimes n})^d$, where $\hat{H}$ is the Hadamard gate and $d \in \mathbb{N}_+$ is the number of repeats. The operator $\hat{U}_{\Phi(\mathbf{x})}$ is given by the following expression:

$$\hat{U}_{\Phi(\mathbf{x})} = \exp\left(i \sum_{S\subseteq[n]} \phi_S(\mathbf{x}) \Pi_{i\in S} \hat{Z}_i\right), \tag{16}$$

where $\hat{Z}_i$ is a Pauli-Z gate, and the functions $\phi_S(\mathbf{x}) \in \mathbb{R}$ introduce the data vectors $\mathbf{x} \in \mathbb{R}^n$ into the register. The map is the so-called ZZ map.

For the data representation given by the operator (16), determining the function of the kernel is a difficult task ($\#P$-hard) for classical computers.

This requires, in general, operating on 2^n-dimensional vectors in the feature space, which, for example, for hyperspectral data containing $n = 100$ bands correspond to composite vectors of dimension $2^{100} \sim 10^{30}$. This is beyond the capabilities of the current and imaginable future classical supercomputers. On the other hand, quantum computations require, in this case, a quantum register composed of $n = 100$ logical qubits and possibly a few thousand noisy physical qubits, which sounds realistic from the 15-year time perspective. This justifies approaching the problem from the side of quantum computing, especially in the case of hyperspectral data (with large n).

However, validating the utility of the map introduced for real-life hyperspectral is currently a challenging task. This is simply because of the high complexity of the map, which makes it difficult to emulate with classical HPC. Therefore, some lower-complexity maps have recently been investigated to make the application to real-world multispectral data possible Miroszewski et al. [2023].

26.3 Feature selection based on quantum optimization

Another approach is to perform feature selection using quantum optimization. Here, one encodes the tasks into a mathematical optimization problem that can be solved using a quantum computer. In Refs. Otgonbaatar and Datcu [2021a]; Mücke et al. [2023], a feature selection approach that utilizes the QUBO (Quantum Unconstrained Binary Optimization) problem and quantum annealing for optimization has been proposed for processing hyperspectral images (HSIs) and practical datasets. This approach involves mapping the feature selection problem to a binary optimization problem, where each feature corresponds to a binary variable, and the objective is to minimize or maximize a particular performance metric, such as precision or accuracy. The resulting optimization problem can then be solved using a quantum optimization algorithm, such as quantum annealing or the Quantum Approximate Optimization Algorithm (QAOA).

Following Ref. Mücke et al. [2023] let us consider labeled dataset $D := \{(\mathbf{x}_i, y_i), i \in [N]\}$, where $[N] = \{1, 2, ..., N\}$ is the set of data point indices and $\mathbf{x}_i \in \mathbb{R}^n$ – the quantum feature selection is even performed on HSIs on a D-Wave quantum annealer Otgonbaatar and Datcu [2021a]. Applying the feature selection procedure reduces the set D to $D_S := \{(\mathbf{x}_{S,i}, y_i), i \in [N]\}$, where the feature-selected data points $\mathbf{x}_{S,i}$ are defined so that they lead to a performance comparable to the original classification problem.

The selected features from the original data vector can be indicated by the binary vector $\mathbf{X} = \{X_1, ..., X_n\} \in \{0, 1\}^n$, so that $X_i = 1$ if a given feature is selected and $X_i = 0$ if it is not. Then the following cost function can be introduced:

$$Q(\mathbf{X}, \alpha) = \mathbf{X}^T \mathbf{Q}(\alpha) \mathbf{X} = -\alpha \sum_{i=1}^{n} I_i X_i + (1 - \alpha) \sum_{i,j=1}^{n} R_{ij} X_i X_j, \tag{17}$$

where the parameter $\alpha \in [0, 1]$ balances the contribution of the two terms. The first term, where $I_i := I(X_i, y) \geq 0$, quantifies the importance of a given measure. Here, $I(X_i, y)$ is the mutual information between the individual features X_i and the class label y. The second is the redundancy term where $R_{ij} := I(X_i; X_j) \geq 0$ is the mutual information among the individual features. In the matrix notation, the elements of $\mathbf{Q}$ can be written as $Q_{ij}(\alpha) = R_{ij} - \alpha(R_{ij} + \delta_{ij} I_i)$.

The task is now to find such a configuration of X_i, call it X_i^*, that both the importance and the redundancy terms are minimized. So, we are looking for:

$$\min_{\mathbf{X} \in \{0,1\}^n} \mathbf{X}^T \mathbf{Q}(\alpha) \mathbf{X}, \tag{18}$$

and in consequence, the set of the selected features is given by

$$\mathbf{X}^* = \arg \min_{\mathbf{X} \in \{0,1\}^n} \mathbf{X}^T \mathbf{Q}(\alpha) \mathbf{X}. \tag{19}$$

The task introduced above is the QUBO (Quantum Unconstrained Binary Optimization) problem, which typically is of the NP-hard type. This is because for every X_i, two values of $\{0, 1\}$ are allowed. In consequence, for n features, there are 2^n configurations to be explored. Therefore, finding the global minimum requires performing the number of steps that grow exponentially with n. So for the $n = 100$ hyperspectral bands, the number of configurations to explore is $2^{100} \sim 10^{30}$ and for $n = 100$ we have $2^{200} \sim 10^{60}$. Both numbers are beyond the scope of any current classical computing resources. Nevertheless, in some cases, not all configurations need to be explored, and classical ergodic algorithms, such as those based on the Markov chain, can be applied.

The QUBO is known to be related to the problem of finding the ground state of the Heisenberg-type quantum Hamiltonian:

$$\hat{H} = \sum_{<i,j>} a_{ij} \hat{Z}_i \hat{Z}_j + \sum_i b_i \hat{Z}_i + c, \tag{20}$$

where $\hat{Z}_i$ are the Pauli-Z operators and the summation $< i, j >$ is defined such that it does not repeat over pairs. The coefficients a_{ij}, b_i and c can be directly related to the parameters of the original problem defined by Eq. 17.

There are various approaches to attempting to determine the ground state of the Hamiltonian (20). On the one hand, the task can be formulated as the quantum annealing problem, which may be implemented on a quantum annealer, such as the one provided by the D-Wave company. On the other hand, optimization algorithms that operate on gate-based quantum computers can be utilized. This, in particular, concerns the QAOA algorithm, and the variational algorithms being a subject of the subsequent subsection.

26.4 Feature selection based on variational methods

Quantum variational methods play a significant role in the early stage of applications of quantum computing technologies. It turns out that the approach is NISQ-friendly and may provide an improvement in solving real-life problems. These algorithms use a quantum computer to find the optimal solution to a given problem by minimizing a cost function. The cost function is typically defined in terms of the parameters of a quantum circuit, and the algorithm iteratively updates these parameters until the optimal solution is found. Importantly, variational methods are typical of the hybrid type. Some of the applications of variational quantum methods include: quantum chemistry (determination of the ground stats of molecules), machine learning (optimization of the parameters of machine learningmodels), and quantum error correction (optimization of the parameters of the quantum error correction codes).

In the context of feature selection, the variational methods can be employed both in the quantum/hybrid RFE approach and the QUBO problem discussed in the previous subsection.

Concerning the RFE, a concrete proposal for a quantum variational classifier has been presented in Ref. Havlíček et al. [2019]. Its undoubted advantage is that, unlike the original QSVM method, the number of qubits in the quantum register is independent of the number of training data m. The approach is based on the so-called *Variational Quantum Classifier* (VQC), which uses a variational quantum system to classify a training set in direct analogy to conventional SVM. In this method, the first step is the same as in QSVM, but in the second step, instead of matrix inversion, a parametrized quantum circuit $\hat{W}(\theta)$ is used. Training consists of minimizing a certain cost function with respect to the parameters of the circuit (given by the vector θ).

27 Feature extraction

Feature extraction methods take the original input features and transform them into new, usually lower-dimensional target features while maintaining as much of the informational content of the data as possible Ding et al. [2012].

27.1 Quantum PCA

Principal Component Analysis (PCA) is an over-century-old technique of dimensionality reduction Pearson and Lines [1901]. It is typically performed by diagonalizing the covariance matrix of the given data and keeping the subset of the eigenvectors, which correspond to the largest eigenvalues.

The approach to quantum principal analysis consists of two steps: association of the covariance matrix K with the density matrices ρ and diagonalization of the resulting density matrix.

Suppose that we have n-dimensional random vectors $X = \{X_1, \ldots, X_n\}$. The entry $K_{\alpha\beta}$ of the covariance matrix is defined as

$$K_{\alpha\beta} = \mathrm{cov}(X_\alpha, X_\beta) = E\left((X_\alpha - E(X_\alpha))\,(X_\beta - E(X_\beta))\right), \tag{21}$$

where E denotes the expectation value. Now, for the $L2$ normalized dataset $\mathcal{D} = \{x_\alpha^{(1)}, x_\alpha^{(2)}, \ldots, x_\alpha^{(m)}\}$, $\sum_{\alpha=1}^{n}\left(x_\alpha^{(i)}\right)^2 = 1$ one can perform the amplitude encoding Grover [2000]; Grover and Rudolph [2002]; Plesch and Brukner [2011] of each of the m data points $x^{(i)} \mapsto |x^{(i)}\rangle\langle x^{(i)}| = \rho^{(i)}$. The authors of Gordon et al. [2022] have noticed that introducing the ensemble average density matrix $\bar{\rho} = E\left(\rho^{(i)}\right)$ leads to a useful equivalence

$$\bar{\rho}_{\alpha\beta} = K_{\alpha\beta} + \mu_\alpha \mu_\beta, \tag{22}$$

where $\mu_\alpha = \sum_{i=1}^{m} x_\alpha^{(i)}$ is the mean value of the feature α in the data set. Furthermore, in accordance with Weyl's theorem for the eigenvalues of Hermitian matrices, the eigenvalues of K and $\bar{\rho}$ are closely related. This

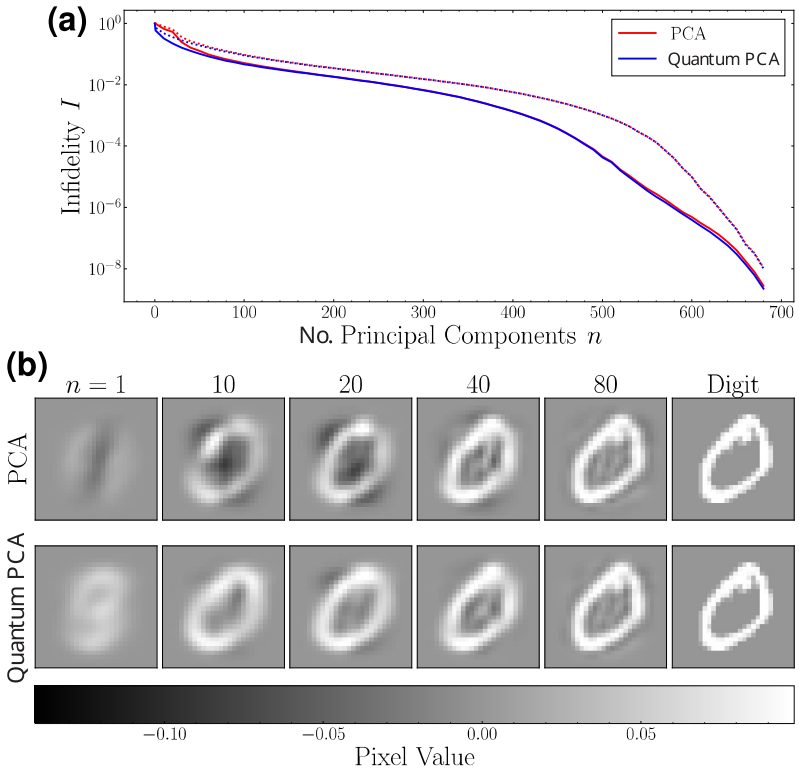

Figure 23: Projection onto n principal components for PCA and quantum PCA of the MNIST dataset. (a) The plot shows the median (solid) and upper 90% interval (dotted) infidelity (I) values between the projected state constructed using n principal components calculated from PCA and quantum PCA. The median and upper interval is calculated over all 50000 digits. (b) We show how the images are reproduced when using $n = 1$ to $n = 80$ principal components from PCA and quantum PCA in the upper and lower rows, respectively. From Gordon et al. [2022] (CC BY 4.0).

gives a concrete implementation of the covariance matrix preparation for quantum computers and guarantees the quality of the quantum principal component analysis (qPCA); see Fig. 23.

For the diagonalization of the covariance matrix on quantum computers, one long-term approach and one near-term approach are presented. In 2014 Lloyd et al. Lloyd et al. [2014] introduced a quantum principal component analysis (qPCA) algorithm which was supposed to provide an exponential speed-up over PCA for low-rank covariance matrices. The author of Tang [2021a] noted that the exponential speedup is possible only with the unjustified assumption of state preparation. With further discussions Cotler et al. [2021] it seems that one can hope for the speedup for qPCA on classical data, but it will not be exponential. The original qPCA algorithm was designed for large, fault-tolerant quantum computers, and hence is not suitable near-term devices. Recently, NISQ-era friendly algorithms, useful for qPCA implementation, were proposed LaRose et al. [2019]; Cerezo et al. [2022]; Verdon et al. [2019]; Ezzell et al. [2022]; Gordon et al. [2022]. Contrary to the original qPCA formulation, the variational quantum state diagonalization (VQSD) LaRose et al. [2019] and variational quantum state eigensolver (VQSE) Cerezo et al. [2022] algorithms require, respectively, only one or two copies of the density matrices ρ. This makes them much better suited for near-term quantum

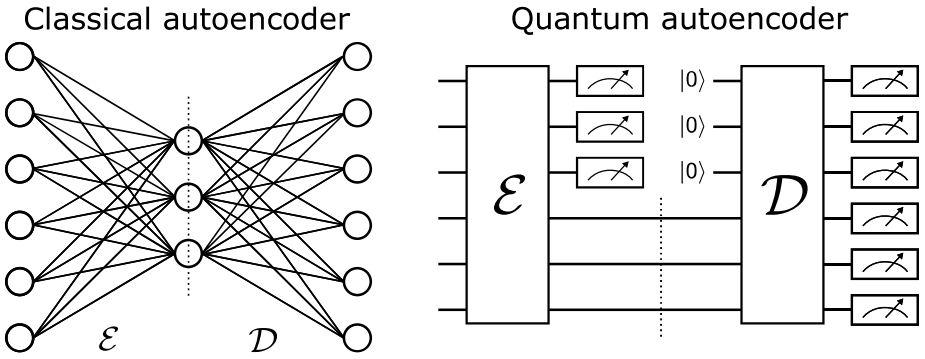

Figure 24: *Left:* Classical graphical representation of a 6-bit autoencoder with a 3-bit latent space. The map $\mathcal{E}$ encodes a 6-bit input into a 3-bit intermediate state, after which the decoder $\mathcal{D}$ attempts to reconstruct the input bits at the output. The bottleneck is represented by a dashed line. *Right:* Circuit implementation of a quantum algorithm inspired by the classical autoencoder on the left.

devices. In Figure 23 the application of qPCA for simulated ten-qubit system in Gordon et al. [2022] is presented.

27.2 Quantum autoencoders

The classical autoencoder Kramer [1991] is a type of neural network used to learn an efficient representation of unlabeled data. Its distinct feature is an encoder-decoder architecture. An encoder function $\mathcal{E}$ takes the $(l+k)$ bit input data x and transforms it into a lower-dimensional l bit representation $\mathcal{E} : x \to \tilde{x}$, effectively compressing the information contained in x. A decoder function $\mathcal{D}$ recreates the input data from the encoded representation $\tilde{x} \to \hat{x}$. The typical graphical representation (Figure 24a) of the autoencoder consists of a convergent neural network with a decreasing number of neurons with consecutive layers, followed by a set of layers of neural networks growing in width. The narrowest layer in the network, the bottleneck, includes the representation of data in latent space $\tilde{x}$. During training, an autoencoder tries to minimize the difference between input data x and output $\hat{x}$. The trained autoencoder, with the use of an encoding function, allows for efficient dimensionality reduction or feature extraction of the data. Unlike PCA, autoencoders can learn non-linear relationships and therefore perform better at compressing data.

Inspired by the erasure of k bits of information in the classical encoder function, the idea of quantum autoencoders was introduced in Romero et al. [2017]. Now, the input data is stored in $l+k$ qubits and processed by the variational layers in $\mathcal{E}$. k-qubits are then traced out (indicated by measurement in Figure 24b). The newly prepared k qubits with latent l qubits are then processed with the decoder function $\mathcal{D}$. The natural task for training the model would be to use the classical optimization of fidelity between input and output states $F(|\psi_i\rangle, \rho_i^{out}) = \langle\psi_i|\rho_i^{out}|\psi_i\rangle$ to train the model. The authors of Romero et al. [2017] argue that invoking a reference state and performing a swap test on it and k trash qubits leads (for high fidelities) to the equivalent optimization task, while training on only the trash state. In the concrete realization (see Figure 25) of the quantum autoencoder circuit, one initializes l qubits in the latent space, k qubits in the trash space and for reference state and one auxilliary qubit needed for the swap test.

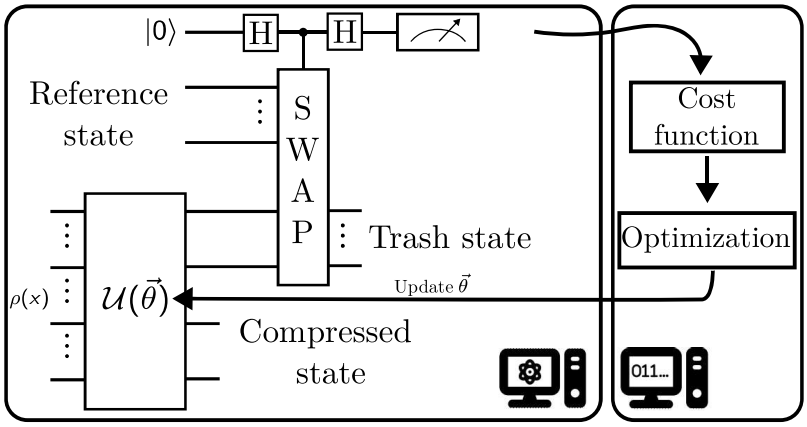

Figure 25: Schematic representation of the variational circuit for training a quantum autoencoder proposed by Romero et al. [2017]. The input state, carrying classical data x is transformed by a parameterized unitary $\mathcal{U}(\vec{\theta})$. For the computation of the cost function, SWAP test is employed. The quality of the compression of the initial state is optimized on a classical computer.

28 Sizing quantum machines for feature selection and feature extraction of hyperspectral imagery data

One of the most promising quantum hardware architectures is the circuit-based model, where quantum algorithms are implemented as sequences of quantum gates. This architecture allows for the efficient implementation of quantum feature selection and quantum feature extraction algorithms, such as quantum SVM-based RFE or quantum PCA. Various hardware implementations of gate-based quantum computers are currently under development, including superconducting quantum computers, trapped ions quantum computers, photonic quantum computers, spin qubit quantum computers, or topological quantum computers.

While the most advanced approach is currently the one based on superconducting qubits, it faces numerous difficulties, such as short coherence times and low interconnectivity of qubits. Consequently, only relatively shallow quantum circuits can be evaluated using the technology. The allowed scalability of the approach with the number of qubits partially compensates for this weakness. Nevertheless, because the quantum algorithms investigated in this Section are expected to utilize deep quantum circuits for real-world hyperspectral data, applying gate-based superconducting quantum computers may pose a difficulty.

Both the interconnectivity and decoherence properties of qubits can be substantially improved by applying other techniques, such as those based on trapped ions and photons. While promising, these technologies face challenges concerning their scalability. Therefore, except in the cases of multispectral rather than hyperspectral data or low-complexity feature maps, it is rather unlikely that gate-based quantum computers will be ready to implement successfully the discussed algorithms for real-world hyperspectral data earlier than in the 15 years time horizon.

Another important architecture is adiabatic quantum computing (quantum annealing) model, which is well-suited for solving optimization problems. As discussed in Sec. 26.3, adiabatic quantum computers can be utilized for quantum feature selection by formulating the problem as an optimization task and mapping it onto the Ising-type model.

The current and expected to develop in 3-5 years time horizon architectures of quantum annealers have a chance to meet the requirements of real-world feature selection problems. However, adiabatic quantum

computers, despite their potential, face several limitations. In particular, the requirement of adiabatic evolution imposes constraints on the execution time and the scalability of the algorithms. The adiabatic process needs to be slow enough to ensure the system remains in its ground state, which can result in longer computation times compared to classical methods or other quantum computing models. Furthermore, because the currently most advanced quantum annealers utilize the superconducting qubits technology, the approach faces challenges with maintaining coherence and providing a high level of interconnectivity. In consequence, even if potentially possible to apply to real-world hyperspectral data, superconducting quantum annealers remain difficult to assess with respect to their accuracy and time performance.

Regarding the feature extraction on universal quantum computers, for a hyperspectral dataset with approximately 200-300 spectral bands, the amplitude encoding of the ensemble average density matrix $\bar{\rho}$ requires the use of 8-qubit registers. In the original implementation of qPCA by Lloyd et al. Lloyd et al. [2014], one needs to prepare $n \in \mathcal{O}(1/\epsilon^3)$ copies of $\bar{\rho}$ to solve the eigenproblem with an accuracy of ϵ. On the other hand, the VQSD algorithm by Larose et al. LaRose et al. [2019] requires two copies of $\bar{\rho}$, resulting in a quantum architecture consisting of 16 qubits. In the case of VQSE proposed by Cerezo et al., Cerezo et al. [2022], only one copy of $\bar{\rho}$ is needed, hence requiring 8 qubits.

In the current implementation of autoencoders proposed by Romero et al. Romero et al. [2017], a single set of quantum registers is used for the latent space, along with two sets of quantum registers for the trash space and an additional qubit register for the swap test. This results in a total of $l + 2k + 1$ qubits. Considering the amplitude encoding of 8 input qubits, one would require quantum architectures consisting of 10 to 16 qubits.

28.1 Currently

Currently, the most sophisticated opportunities for implementing quantum algorithms in feature selection and extraction lie within the capabilities of classical emulators of quantum computers. Specifically, these emulators facilitate the investigation of hybrid feature selection methods for multispectral datasets comprising several dozens of bands. While the methodology applies amplitude encoding, it also possesses the potential to accommodate hyperspectral data with hundreds of bands. However, amplitude encoding offers limited noise control within genuine quantum implementations. Consequently, this approach may not seamlessly translate to future quantum computing technology.

The computational limitations of emulators pose substantial challenges for predicting the feasibility of hybrid quantum algorithms, especially those based on variational methods. Indeed, drawing definitive conclusions about the effectiveness of these hybrid techniques in managing real-world hyperspectral data is an intricate task, bordering on impossible.

In addition, the existing quantum computers' opportunities are considerably restricted due to the absence of practical quantum correction code implementations. Thus, only relatively uncomplicated tasks with limited practical value can be achieved. An example of this would be hybrid feature selection for low-dimensional (a few bands) data vectors.

At present, some opportunities are accessible through quantum annealers. This particularly refers to the feature selection method outlined in Section 26.3. However, there is no evidence yet that currently available quantum annealers offer a computational advantage.

28.2 3-5 years

Given the relatively short forecast horizon of 3-5 years, significant advancements in quantum computing capabilities are not expected. Nonetheless, this period may witness the deployment of the initial practical implementations equipped with quantum error correction codes. This development would pave the way for the usage of noisy superconducting quantum processors to carry out more complex quantum algorithms.

In light of our current knowledge, it is anticipated that quantum or hybrid feature selection and extraction for multispectral data encompassing approximately ten bands could be realized on actual quantum computers. However, this advancement is not expected to surpass or even match the capabilities of classical computers.

28.3 15 years

The next 15 years hold immense potential for significant advancements compared to the current state of the art. Notably, the successful implementation of quantum correction codes on superconducting quantum computers comprising thousands of qubits may pave the way for hybrid and quantum feature selection and extraction, enabling the utilization of approximately 100 bands. However, it is crucial to emphasize that this does not guarantee a definitive advantage in terms of time performance or accuracy. There is still limited knowledge regarding the algorithms employed and the future capabilities of the hardware, making it challenging to draw conclusive statements in this regard. Nonetheless, certain advantages can be anticipated, e.g., in terms of the energy consumption of quantum computers when compared to classical computing units. The implementation of Recursive Feature Elimination (RFE) based on Quantum Support Vector Machine (QSVM), which is expected to offer exponential speed-up, will possibly exceed the given 15-year time frame.

29 SWOT Analysis

29.1 Strengths

- High level of motivation resulting from the computational intricacy of the underlying problem.
- The availability of real-world datasets for training and testing.
- Documented evidence of exponential speed-up for Recursive Feature Elimination (RFE) based on Quantum Support Vector Machine (QSVM).

29.2 Weaknesses

- Implementation challenges of the RFE using QSVM due to the required number of logical qubits being of the order of the size of the training dataset (typically, tens of thousands). This poses significant difficulties in practical applications.
- Absence of definitive proof showcasing a clear edge for quantum kernel-based methods.

29.3 Opportunities

- Potential to utilize hybrid approaches that require a relatively small number of qubits (of the order on 10^2 logical qubits), thereby increasing feasibility.
- Possible advancements pertaining to accuracy levels of selection and feature extraction, alongside the prospect of reducing computational energy consumption.

29.4 Threats

- Typically, the deep quantum circuits under consideration require long coherence times for execution. This imposes implementation difficulties on superconducting quantum computers.
- The highly dimensional feature spaces offered in the quantum approach may not be needed in many relevant scenarios.

Table 3: Summary of the identified feature selection methods for hyperspectral imagery data.

Method	QSVM RFE Rebentrost et al. [2014]	VQC RFE Havlíček et al. [2019]	Quantum optimization Otgonbaatar and Datcu [2021a]
Resources	high $\gtrsim 10^5$ logical qubits	moderate $\sim 10^2$ logical qubits	low/moderate $\sim 10^2$ logical qubits
Time horizon	>15 years	3-5 years	now/3-5 years
Architecture	gate-based quantum	gate-based hybrid	annealing/gate-based hybrid
Speedup	exponential	polynomial	polynomial/exponential

Table 4: Summary of the identified feature extraction methods for hyperspectral imagery data.

Method	qPCA Lloyd et al. [2014]	variational qPCA Gordon et al. [2022]	qAutoencoders Romero et al. [2017]
Resources	high $\gtrsim 10^3$ logical qubits	moderate $\sim 10^2$ logical qubits	low/moderate $\sim 10^1 - 10^2$ logical qubits
Time horizon	15 years	3-5 years	3-5 years
Architecture	gate-based quantum	gate-based hybrid	gate-based hybrid
Speedup	polynomial	polynomial	polynomial

30 Conclusions

In this Section, a set of quantum and hybrid (classical-quantum) algorithms for feature selection and feature extraction was presented. The selected set includes long-term and near-term algorithms. The greatest opportunity for quantum advantage lies in the long-term routines for fault-tolerant quantum computers. These include the following: Recursive Feature Elimination based on the full Quantum Support Vector Machines, feature selection based on quantum kernel methods, and original Quantum Principal Component Analysis. To harness the advantage of these algorithms, long-term technological development of quantum computers is needed. A key issue is to be able to access a large ($\sim 10^2$) number of logical qubits, increase their coherence time, and improve the precision of the unitary gates used in quantum circuits. Therefore, we estimate at least a 15-year time horizon for such methods to become available. On the shorter time scale (5-year horizon), promising candidates for quantum advantage include hybrid algorithms based on variational architectures. These include feature selection methods based on Variational Quantum Classifiers, quantum Principal Component Analysis using Variational Quantum State Diagonalization or Variational Quantum State Eigensolver, and quantum autoencoders. The identified quantum and hybrid approaches to feature selection have been summarised in Tab. 3, while the identified approaches to feature extraction have been collected in Tab. 4

The crucial issue in the context of hyperspectral data analysis is the state preparation for the above algorithms. On the one hand, it is possible to employ an amplitude encoding scheme, which allows one to operate some of the selected algorithms on the order of 10^1 qubits. However, this requires operating with high precision on the quantum states. On the other hand, other encoding schemes require access to at least hundreds of good-quality qubits. No quantum advantage in practical tasks has been confirmed so far, but we can expect it with the technical development of quantum computers. Answering whether such an advantage will realistically provide a net gain for Earth observation is an issue that requires further in-depth studies.