

Potential of Quantum Machine Learning for Processing Multispectral Earth Observation Data

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY-SA 4.0

SUBMISSION DATE / POSTED DATE

13-01-2023 / 19-01-2023

CITATION

Gupta, Manish Kumar; Romaszewski, Michał; Le Saux, Bertrand; Gawron, Piotr (2023): Potential of Quantum Machine Learning for Processing Multispectral Earth Observation Data. TechRxiv. Preprint.
<https://doi.org/10.36227/techrxiv.21898902.v1>

DOI

[10.36227/techrxiv.21898902.v1](https://doi.org/10.36227/techrxiv.21898902.v1)

Potential of Quantum Machine Learning for Processing Multispectral Earth Observation Data

Manish K. Gupta ^{*}, Michał Romaszewski[†], Piotr Gawron ^{*}

^{*}AstroCeNT — Particle Astrophysics Science and Technology Centre — International Research Agenda, Nicolaus Copernicus Astronomical Center, Polish Academy of Sciences, Rektorska 4, 00-614 Warsaw, Poland

[†]Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, ul. Bałycka 5, 44-100 Gliwice, Poland

Abstract—Quantum computers with hundreds of noisy qubits are already available for the research community. They have the potential to run complex quantum computations well beyond the computational capacity of any classical device. It is natural to ask the question, what application these devices could be useful for? Land Use and Land Cover classification of multispectral Earth observation data collected from the earth observation satellite mission is one such problem that is hard for classical methods due to its unique characteristics. In this work, we compare the performance of several classical machine learning algorithms on the stilted re-labeled dataset of the Copernicus Sentinel-2 mission, when the algorithm has access to Projected Quantum Kernel (PQK) features. We show that the classification accuracy increases drastically when the model has access to PQK features. We then naively study the performance of these algorithms with and without access to PQK features on the original Copernicus Sentinel-2 mission data set. This study provides key evidence that shows the potential of quantum machine learning methods for Earth Observation data.

Index Terms—Quantum Computing, Quantum Machine Learning, Projected Quantum Kernel, Multispectral Image, Earth Observation, Sentinel-2, Earth Observation, Land Use and Land Cover classification, Remote Sensing.

I. INTRODUCTION

QUANTUM computers with more than 400 noisy qubits will be available to researchers by the year 2023, they will be able to run complex quantum calculations which are well beyond the computational capacity of any classical device. The computational power of these devices is expected to increase in the coming years as the noise in these devices is addressed with error-correcting codes, and producers increase the number of physical qubits available. With such powerful devices at hand, it is natural to look for problems that are generally difficult for classical machines to solve and find algorithms to run on the new device. Earth observation (EO) is one such problem that is difficult due to its unique characteristics. It would be a futuristic step to develop algorithms that can run on these new devices and understand their limitation concerning Earth observation.

Earth observation and specifically Land Use and Land Cover classification (LULC) is an important task for achieving Sustainable Development Goals (SDGs) [1]. Extracting

knowledge from continuous multispectral and hyperspectral data quickly and effectively on-Earth objects and land covers, mapping them, and monitoring their changes on digital twin [2] is the need of the hour. The amount of remote sensing data that is being continuously captured by Earth observation satellites with onboard multispectral, hyperspectral, and radar sensors are in excess of 150 terabytes per day, which is not always processed efficiently [3]. The amount of data generated by EO missions falls in the category of Big Data. The massive amount of data comes with the so-called four challenges of Big Data referred to as “four Vs”: Volume, Variety, Velocity, and Veracity [4], [5]. Extracting meaningful information from a such huge volume of data efficiently requires special tools and methods. Although Machine Learning (ML) algorithms have shown great potential in terms of obtaining a detailed understanding of Earth observation data [6]–[9]. The performance of these ML techniques is limited by the amount of training data and available computational power. With the availability of a quantum computer with a promise to solve complex problems more efficiently when compared to any available classical machine, it has become necessary to explore the new quantum computing methods for multispectral Earth observation data understanding. There has been some progress in this area, where researchers have used some combination of classical and quantum parts for Remote Sensing data classification [10]–[16].

A large-scale, fully error-corrected, quantum computer will not be built for many decades, nevertheless, the recent advancement in their implementation allows us to study their application to real-life computational problems [17]. Quantum computers consisting of hundreds of noisy qubits are already available and can run specific quantum algorithms. One class of such algorithms uses quantum models to generate correlations between variables that are inefficient to represent using classical models of computation. Recent theoretical and experimental evidence suggests that quantum computers can efficiently sample probability distributions that are exponentially hard to sample classically [18], [19]. This is the type of advantage that is exploited by both quantum neural network (QNN) [20] and quantum kernel methods [21]. QNN parameterizes a distribution using a set of adjustable parameters and the quantum kernel method encodes classical data into a quantum state as a feature map which maps the

Supported by IRAP AstroCeNT (MAB/2018/7) funded by FNP from ERDF. Supported by ESA under the contract No. 4000137375/22/NL/GLC/my.

Supported by IRAP AstroCeNT (MAB/2018/7) funded by FNP from ERDF.

data in higher-dimensional quantum Hilbert space and uses a quantum computer to compute the inner products of quantum states [22].

It is postulated that quantum machine learning algorithms can outperform their classical counterparts and the justification that is generally provided for this is that if the quantum circuit is hard to sample classically then there is a potential for quantum advantage. Huang et al. [23] showed this argument to be incomplete and proved that with a sufficient amount of training data classical models can be elevated to rival quantum models, even when the quantum circuit is hard to sample classically. They also proposed “geometric difference” between kernel functions defined by classical and quantum machine learning models and showed that if the geometric is small then the classical machine learning model is guaranteed to provide similar or better performance in prediction on the data set. If the geometric distance is large then there exists a data set where the quantum model exhibits a large prediction advantage. They also found that due to small geometric difference a variety of common quantum models in the literature perform similarly or worse than the classical model. To circumvent it, they proposed the projected quantum kernel (PQK) method that generally enlarges the geometric difference between the kernels.

The mapping of land on Earth is categorized into land use (LU) classification and land cover (LC) classification. Although the two terms are used interchangeably they are different. According to the Food and Agriculture Organization (FAO) of the United Nations, “Land cover is the observed (bio)physical cover on the Earth’s surface”, while “Land use is characterized by the arrangements, activities, and inputs by people to produce, change or maintain a certain land cover type” [24]. In simple terms land cover is what covers the surface of the earth, *e.g.* classes: water, snow, grassland, deciduous forest, and bare soil and land use describe how the land is used, *e.g.* classes: wildlife management area, agricultural land, urban, and recreation area. The two terms land use and land cover are tightly coupled, and they are jointly classified hence “land use and land cover” (LULC) classification is considered a more general concept.

Remote sensing missions capture the imagery using optical, thermal, or Synthetic Aperture Radar (SAR) imaging systems. The optical sensor is sensitive to a spectrum range from visible to mid-infrared radiation from Earth’s surface and captures Panchromatic, multispectral, or hyperspectral images. Commonly, images with more 2 to 13 spectral bands per pixel are called multispectral, while the images with hundreds of spectral bands are called hyperspectral.

The LULC classification problem is mathematically defined as an assignment $f : X \rightarrow Y$ from the set of spectral images to the set of pixel class arrays. The input space $X \subseteq \mathbb{R}^{W \times H \times B}$ represents the set of possible images with W, H, B being respectively the image width, height, and the number of spectral bands. The output space for pixel-level land cover classification is represented as $Y \subseteq C^{W \times H}$, where $C = \{0, 1, \dots, N\} \subseteq \mathbb{Z}^{0+}$ is the set of LULC categories. A variety of approaches is used to perform the Land Use and Land Cover classification task. The approach that is used

depends on the resolution of the image being processed, the exact task to be performed, and the available resources [25].

For this work, we use the Sentinel-2 data that contains multispectral data with pixels of 10m resolution. For such a dataset a common task is to perform semantic segmentation, where labels are assigned to every pixel [26] individually. The traditional Machine Learning (ML) techniques that are used for this task are support vector machines, decision trees, and perceptron-like neural networks. Semantic segmentation is often a two-step process. In the first step, pixel-by-pixel classification is performed where only the spectral information is used, and in the second step, the labels are smoothed out by employing a probabilistic graphical model such as *e.g.* Markov random field or conditional random field. Occasionally spatial and spectral information are used jointly in the semantic segmentation task [27]. In this work, we focus solely on spectral information classification, and we deliberately ignore the spatial relationship between pixels.

In this work, we are interested in answering the question of whether quantum machine learning [28], [29] is suitable for the classification of multispectral data such as, for example, data gathered by the Earth observation satellites. The specific goal is to study a Quantum Machine Learning system for Land Use and Land Cover classification of the Earth’s surface based on Sentinel-2 images. We find evidence that shows that there exists a data set where classical model test accuracy increases drastically when the model has access to Projected Quantum Kernel features. It is a follow-up to the previous work on multi-spectral image classification with QNN, where one of the authors of this work showed that the QNN-based classifier achieved a score of 66% in multi-class classification scenario [10] and extension of the work [16]. We start by providing a brief overview of key concepts used in this work such as quantum kernel methods, geometric distance, and projected quantum kernel (PQK) method in the next section. We then describe the experiment and finally discuss the result and then conclude.

II. METHODOLOGY

A. Quantum kernel methods

In machine learning, kernel function $k : \mathbb{R}^B \times \mathbb{R}^B \rightarrow \mathbb{R}^{0+}$ can be understood as a measure of similarity — a generalized scalar product — between given feature vectors $x_i \in \mathbb{R}^B$ and $x_j \in \mathbb{R}^B$. Given a set of feature vectors $\{x_1, x_2, \dots, x_{N_{\text{samples}}}\}$ and a kernel function k we can calculate a kernel matrix $[K_{i,j}]_{i=1,j=1}^{N_{\text{samples}},N_{\text{samples}}}$ with elements $K_{i,j} = k(x_i, x_j)$ that stores similarity between all N_{samples} feature vectors. Matrix K is positive semi-definite. And therefore we can calculate the geometric difference between two Kernel matrices K^1, K^2 associated with two kernel functions k^1, k^2 . The geometric difference is defined over a dataset as

$$g(K^1 \| K^2) = \sqrt{\left\| \sqrt{K^1} (K^2)^{-1} \sqrt{K^1} \right\|_\infty}, \quad (1)$$

where $\|\cdot\|_\infty$ is the spectral norm of the resulting matrix. We assume that $\text{Tr}\{(K^1)\} = \text{Tr}\{(K^2)\} = N_{\text{samples}}$ [30]. To evaluate a potential for quantum advantage we must calculate

the geometric difference between quantum kernel and classical kernel. It is a crucial test for comparing classical and quantum machine learning models. The geometric difference for a quantum kernel is defined with respect to the closest efficient classical model. If geometric distance is small, then the classical machine learning model is guaranteed to provide similar or better performance in prediction on the data set, independent of the function values or labels. If geometric distance is large, then there exist a data set where quantum machine learning model exhibits large prediction advantage. These notions provide an important test for finding potentially useful quantum machine learning models.

A number of quantum machine learning models found in literature can be shown to perform similarly or worse than classical machine learning models due to their small geometric difference. The small geometric difference is often due to the fact that encoded features are too far apart because of the exponentially large Hilbert space employed by existing quantum models. To solve this problem the projected quantum kernel method that circumvents this issue and enlarges the geometric difference was proposed in [30].

B. Projected quantum kernel

Projected quantum kernels (PQK) are a family of kernels that work by projecting the quantum states to an approximate classical representation, for example reducing physical observables or classical shadows [30] and then defining the kernel function using the classical representation. The modified quantum kernel is referred to as the projected quantum kernel. The PQK method was first introduced by Huang et al. in [23]. The projection reduces the large training set dimension to a smaller classical space that generalizes better. The projected quantum kernel is defined on the classical feature space to evade the difficulty in learning due to the exponential dimension in quantum Hilbert space. Projecting an exponentially large Hilbert space using a projected quantum kernel is a difficult task on a classical computer. One of the simplest example of projected quantum kernel is to measure the one-particle reduced density matrix (1-RDM) on all qubits for the encoded state, $\rho_k(x_i) = \text{Tr}_{j \neq k} [\rho(x_i)]$, and then define the kernel as

$$k^{\text{PQ}}(x_i, x_j) = \exp \left(-\gamma^{\text{PQ}} \sum_k \|\rho_k(x_i) - \rho_k(x_j)\|_F^2 \right), \quad (2)$$

where γ^{PQ} is a real positive hyperparameter. The partial trace $\text{Tr}_{j \neq k}$ over qubits labeled by j can be defined as follows

$$\text{Tr}_{j \neq k} [\rho(x_i)] = \sum_{j \in \{0,1\}^{k-1}} \sum_{j' \in \{0,1\}^{D-k+1}} \text{Tr}[(P_j \otimes \mathbb{1} \otimes P_{j'}) \rho(x_i)],$$

where $P_j = |j\rangle\langle j|$.

C. Computing PQK features

To compute the PQK features for a given data instance x_i , we encode this data instance into the quantum state

$$|\psi_i\rangle = V(x_i/n_{\text{trotter}})^{n_{\text{trotter}}} U_{\text{qb}} |0\rangle, \quad (3)$$

where $U_{\text{qb}} = \bigotimes_{j=1}^N R_x(\phi_1^j)R_y(\phi_2^j)R_z(\phi_3^j)$ is the tensor product of Pauli rotations operators, $R_x(\phi) = e^{-iX\phi/2}$, angles are randomly selected once as $\phi \sim U(-2\pi, 2\pi)$ and remain equal for all data points. The integer n_{trotter} is a hyperparameter — the number approximation steps for approximating time evolution [31], [32], and $V(\hat{\theta})$ is defined as

$$V(\hat{\theta}) = \exp \left(-i \sum_{j=1}^N \sum_{l \in \{X,Y,Z\}} \hat{\theta}_j \sigma_j^l \sigma_{j+1}^l \right), \quad (4)$$

where σ_j^l acts on j -th qubit and N is number of qubits. We compute the PQK features based on the 1-RDM by measuring the expectation values of $\langle \psi_i | \sigma_j^l | \psi_i \rangle$, where i indexes over data points, j indexes over qubits and l indexes over Pauli operators $\{X, Y, Z\}$. Mathematically it can be represented as $f_{\text{PQK}} : \mathbb{R}^d \rightarrow \mathbb{R}^{3(d+1)}$ where $f_{\text{PQK}}(x_i) = [\langle \psi_i | \sigma_j^l | \psi_i \rangle]_{j \in \{1,2,\dots,d+1\}, l \in \{X,Y,Z\}}$ and $d = N - 1$ is the number of features.

D. Preparing Stilted Dataset

To achieve maximum separation between quantum and classical models we artificially re-label the dataset using the spectral information found in the classical and PQK kernel matrices. To achieve that we perform three following steps.

We first train Radial Basis Function (RBF) kernel support vector machine using the “scikit-learn” library [33] and “scikit-optimize” library to obtain the best gamma $\gamma_{\text{best}}^{\text{RBF}}$ using the original dataset — original feature vectors and labels.

Next, we compute the kernel matrix for the best classical model using found $\gamma_{\text{best}}^{\text{RBF}}$ and the original feature vectors. We also compute the quantum kernel matrix using feature vectors transformed by f_{PQK} , and $\gamma^{\text{PQ}} = 1$.

Finally, we construct the new stilted dataset that will yield the largest separation between quantum and classical models from a learning-theoretic sense by assigning new labels. The new labels y_{relabel} are obtained from vector $v' = \sqrt{K^Q}v$, where v is the eigenvector of $\sqrt{K^Q}(K^C)^{-1}\sqrt{K^Q}$ corresponding to the eigenvalue of $g^2 = \|\sqrt{K^Q}(K^C)^{-1}\sqrt{K^Q}\|_\infty$, by assigning $y_{\text{relabel},i} = \mathbb{1}_{v'_i > \text{median}(\{v'_i\}_i)}$ and changing 5% labels randomly.

This relabeling of the data maximizes the separation between the quantum and classical models by maximizing the geometric distance between the classical kernel and the PQK kernel. For a detailed description, we refer to Appendix G: “Constructing dataset to separate quantum and classical model” of [23]. The geometric distance [23] between the kernel of classical and quantum models is defined as

$$g(K^C || K^Q) = \sqrt{\|\sqrt{K^Q}(K^C)^{-1}\sqrt{K^Q}\|_\infty} \quad (5)$$

where K^C and K^Q are kernel matrices for the classical and quantum models respectively.

E. Classifiers

The Support Vector Machine (SVM) with the Radial Basis Function (RBF) kernel was used in our classification

experiments as a high performance method with efficient, stable implementation [33]. The classifier was also used to prepare the stilted dataset in Section II-D and is a standard reference classifier for hyperspectral and multispectral classification [34]. As a reference we also used three well-known classifiers combining high performance with explainable decision-making: a k -Nearest Neighbors (k -NN), a Decision Tree (DT) and a Naive Bayes (NB).

III. EXPERIMENTS

In this study, we use Copernicus Sentinel-2 Earth Observation land cover multispectral image data. Sentinel-2 is a European wide-swath, high-resolution, multi-spectral imaging mission. It comprises a constellation of two polar-orbiting satellites that aims to capture land cover changes monitoring and natural disaster management. It carries an onboard Multi-Spectral Imager (MSI) sensor that samples 13 spectral bands: four bands at 10 m, six bands at 20 m, and three bands at 60 m spatial resolution. The orbital swath width is 290 km [35]. The data set has multispectral images sampled in 12 bands and the pixels that are used for classification are categorized into sixteen land use (LU) categories [10], [36]. For this work, we use only four land use categories: “Artificial surfaces and constructions”, “Cultivated areas”, “Broadleaf tree cover” and “Herbaceous vegetation” and perform a binary classification task on pairs classes selected out of the combinations of four classes.

The aim of the experiments is to assess the impact of PQK features on multispectral classification accuracy. To attain this goal we compare the performance of classifiers (SVM, k -NN, DT, NB) on the stilted multispectral dataset, when only PQK feature vectors are provided with case when only original spectral feature vectors are provided. We also naively compare the performance of these classifiers on original Sentinel-2 dataset — using original labels with PQK feature vectors with the original spectral feature vectors. Each experiment consist of the following steps binary datasets preparation, stilted dataset preparation, and classification experiment.

A. Binary datasets preparation

The experiments use binary (two class) datasets created by selecting a subset of labelled multispectral pixels from the multispectral image. The labels correspond to all combinations of four pairs of (LU) classes *i.e.* six pairs in total. For every pair of classes we randomly select a subset of 1000 labelled pixels for each of the classes, and reassign the class labels to zero and one. This procedure is repeated $N_{\text{repetitions}} = 10$ times to create multiple instances of the training dataset \mathcal{T}_i^c where c denotes a pair of classes and $i \in \{0, 1, \dots, N_{\text{repetitions}} - 1\}$ denotes an instance of the experiment.

B. Stilted dataset preparation

A binary dataset $\mathcal{T} = (\mathbf{X}, \mathbf{y})$, where $\mathbf{X} \in \mathbb{R}^{l \times m}$ denotes an array of pixels with multispectral features and \mathbf{y} denotes vector of class labels. The aim of this step it to engineer a stilted dataset $\mathcal{T}_{\text{stilted}} = (\mathbf{X}^{\text{pkq}}, \mathbf{y}_{\text{relabel}})$ where \mathbf{X}^{pkq} denotes

an array of PQK features and $\mathbf{y}_{\text{relabel}}$ denotes a vector of new labels, resulting from the relabeling of class labels such as to maximize the geometric distance between classical and quantum models. The procedures are described in Section II-B, Section II-C and Section II-D and can be summarized in the following steps:

- 1) *Train/test split*: Binary dataset $\mathcal{T} = (\mathbf{X}, \mathbf{y})$ is split into train set $\mathcal{T}_{\text{train}} = (\mathbf{X}_{\text{train}}, \mathbf{y}_{\text{train}})$ and test set $\mathcal{T}_{\text{test}} = (\mathbf{X}_{\text{test}}, \mathbf{y}_{\text{test}})$.
- 2) *Standardization (calculating z-score)*: Arrays $\mathbf{X}_{\text{train}}, \mathbf{X}_{\text{test}}$ are standardized into $\mathbf{X}_{\text{train}}^{\text{standardised}}, \mathbf{X}_{\text{test}}^{\text{standardised}}$.
- 3) *PCA (dimension reduction)*: Arrays $\mathbf{X}_{\text{train}}^{\text{standardised}}, \mathbf{X}_{\text{test}}^{\text{standardised}}$ are reduced to 4 features $\mathbf{X}_{\text{train}}^{\text{reduced}}, \mathbf{X}_{\text{test}}^{\text{reduced}}$. We use only $\mathbf{X}_{\text{train}}^{\text{reduced}}$ to fit the PCA model.
- 4) *SVM kernel optimization*: Grid-search is performed on $(\mathbf{X}_{\text{train}}^{\text{reduced}}, \mathbf{y}_{\text{train}})$ to find the best value of the parameter γ .
- 5) *PQK features extraction*: PQK procedure as described in Section II-C is used on the arrays $\mathbf{X}_{\text{train}}^{\text{reduced}}, \mathbf{X}_{\text{test}}^{\text{reduced}}$ to compute $\mathbf{X}_{\text{train}}^{\text{pkq}}, \mathbf{X}_{\text{test}}^{\text{pkq}}$.
- 6) *Stilted dataset creation*: The training set $\mathbf{X}_{\text{train}}^{\text{reduced}}, \mathbf{X}_{\text{test}}^{\text{reduced}}$, PQK features $\mathbf{X}_{\text{train}}^{\text{pkq}}, \mathbf{X}_{\text{test}}^{\text{pkq}}$, and best gamma value γ are used to assign new labels $\mathbf{y}_{\text{relabel}}$ as described in Section II-D.

C. Classification experiment

To compare the classification accuracy of classifiers (SVM, k -NN, DT, NB) over the original and the stilted datasets, we use a standard two-stage cross-validation experiment. For every binary dataset $\mathcal{T} = (\mathbf{X}, \mathbf{y})$, its labelled examples are divided into training sets $\mathcal{T}_{\text{train}}^k = (\mathbf{X}_{\text{train}}^k, \mathbf{y}_{\text{train}}^k)$ and test sets $\mathcal{T}_{\text{test}}^k = (\mathbf{X}_{\text{test}}^k, \mathbf{y}_{\text{test}}^k)$ using k -fold stratified cross-validation with a number of folds $k = 5$. For every fold the following steps are performed:

- 1) *Standardization*: Training and tests are standardized using mean and standard deviation computed on the training set.
- 2) *Parameter selection*: Classifier parameters are selected using grid-search and internal stratified 3-fold cross-validation on the training set.
- 3) *Classifier testing*: The accuracy of the classifier that is trained on the training set is computed using a test set.

The final accuracy of a classification experiment is computed by averaging accuracies for every fold.

D. Parameter selection

The range of selected parameters for each classifier is as follows:

- SVM: Parameters $\gamma^{\text{RBF}} \in \{10^{-2}, \dots, 10^2\}$, $C \in \{10^{-2}, \dots, 10^2\}$.
- k -NN: Number of neighbors $k \in \{1, \dots, 15\}$ with two different neighbors weighting strategies — *uniform*: where all neighbors have the same weight, *distance*: where neighbors are weighted by the inverse of their distance.

- DT: The minimum number of examples required to split a node $s \in \{2, 3, 4\}$, maximum depth of the tree $d \in \{2, 3, 5, 10, \alpha\}$, where α denotes expansion of nodes until all leaves contain less than s examples.
- NB: The classifier is non-parametric.

E. Experiment list

In order to fully assess the influence of PQK features on the classification accuracy, the classification experiment described in Section III-C was repeated for four combinations of original and PQK features on original and stilted datasets:

- Classification with original spectral features on stilted dataset: binary datasets had the form: $\mathcal{T} = (\mathbf{X}^{\text{reduced}}, \mathbf{y}_{\text{relabeled}})$.
- Classification with PQK features on stilted dataset: binary datasets had the form: $\mathcal{T} = (\mathbf{X}^{\text{pqr}}, \mathbf{y}_{\text{relabeled}})$.
- Classification with original spectral features on original dataset: binary datasets had the form: $\mathcal{T} = (\mathbf{X}^{\text{reduced}}, \mathbf{y})$.
- Classification with PQK features on original dataset: binary datasets had the form: $\mathcal{T} = (\mathbf{X}^{\text{pqr}}, \mathbf{y})$.

IV. RESULTS AND DISCUSSION

Results of the experiments are gathered in two tables that contain averages over all 6 combinations of labels, 10 experiment repetitions, and 5 cross-validation rounds.

TABLE I
MEAN TRAINING AND TEST ACCURACY OF CLASSIFICATION FOR
RELABELED DATASETS.

Features Classifier	Mean Training Accuracy		Mean Test Accuracy	
	Original	PQK	Original	PQK
dtree	0.87 ± 0.12	0.93 ± 0.06	0.64 ± 0.04	0.73 ± 0.03
knn	0.97 ± 0.08	0.97 ± 0.06	0.63 ± 0.04	0.78 ± 0.03
nb	0.56 ± 0.03	0.83 ± 0.02	0.54 ± 0.04	0.82 ± 0.03
svc	0.92 ± 0.09	0.93 ± 0.02	0.63 ± 0.04	0.88 ± 0.02

Table I contains mean training and test accuracies of classification for relabeled stilted datasets. We can observe that the use of PQK features significantly raises the accuracy of the classification.

TABLE II
MEAN TRAINING AND TEST ACCURACY OF CLASSIFICATION FOR
DATASETS WITH ORIGINAL LABELS.

Features Classifier	Mean Training Accuracy		Mean Test Accuracy	
	Original	PQK	Original	PQK
dtree	0.95 ± 0.04	0.79 ± 0.07	0.93 ± 0.04	0.73 ± 0.08
knn	0.98 ± 0.03	0.91 ± 0.10	0.93 ± 0.04	0.75 ± 0.06
nb	0.92 ± 0.04	0.73 ± 0.06	0.91 ± 0.04	0.72 ± 0.07
svc	0.95 ± 0.04	0.84 ± 0.07	0.94 ± 0.04	0.77 ± 0.05

Table II contains mean training and test accuracy of classification for datasets with original labels. Unfortunately in this — more realistic — scenario we can not observe classification accuracy increase. Contrarily, we can observe a significant drop in the classification accuracy when original features are transformed using PQK.

The results in Table I indicate that two of the classifiers: SVM and NB achieve high accuracy for PQK features and their result for test data is similar to the result for training data. This is not the case with k -NN and DT classifiers, for which low accuracy on the test set may indicate overtraining. This effect can also be observed in Table II for the k -NN classifier.

For the purpose of the results' visualization we provide box plots of classification accuracies in Fig. 1 and Fig. 2 that represent the accuracy of classifiers (SVM, k -NN, DT, NB) over the stilted and original datasets created using the spectral data for two classes: "Broadleaf tree cover" and "Herbaceous vegetation". The classification accuracies for the original features are shown in Fig. 1a. The classification accuracies for the PQK features are shown in Fig. 1b. Visibly, the accuracy increases by a significant amount over the same stilted datasets if the classifiers have access to the PQK features. This difference is clearly significant as it is orders of magnitude greater than the value typically considered when comparing classifier performance, which is about 1% [37]. We also naively repeat the experiment over original datasets to assess the effectiveness of PQK features and the result is shown in Fig. 2a and Fig. 2b.

V. CONCLUSIONS

The main conclusion we draw from this empirical investigation is that there exists a dataset that is easy for the quantum model to learn and hard for the classical model to learn. We showed that all the important classical ML models generally used for image segmentation underperform on the stilted data set using standard spectral features. We provide a key evidence that suggest that quantum procedures in addition to classical methods could give an advantage in learning tasks for Earth observation datasets. There are at least two open question that needs further investigation:

- Are there any other classical ML model that perform better using only the standard spectral features over the stilted dataset?,
- Does there exist a natural EO dataset that matches the characteristic of stilted dataset?

If we can show the answer to the first question as false and find a natural EO dataset where we see an advantage with PQK features then we can confidently claim that quantum computers would be useful in processing EO data.

Our investigation, presented in this paper, is limited by many factors: we use only a particular data encoding and quantum kernel out of many, we consider a small ideal simulated quantum computer, and we also assume an infinite number of samples sampled from the quantum circuits that encode the data — among others. Yet we can claim that this empirical study shows the potential for quantum machine learning methods to Earth Observation data analysis and encourages us to perform further investigation.

VI. ACKNOWLEDGMENTS

The data used in this work was prepared by project team: S. Lewiński, R. Malinowski, M. Rybicki, E. Gromny, M. Jenerowicz, Marcin Krupiński, C. Wojtkowski, Michał Krupiński

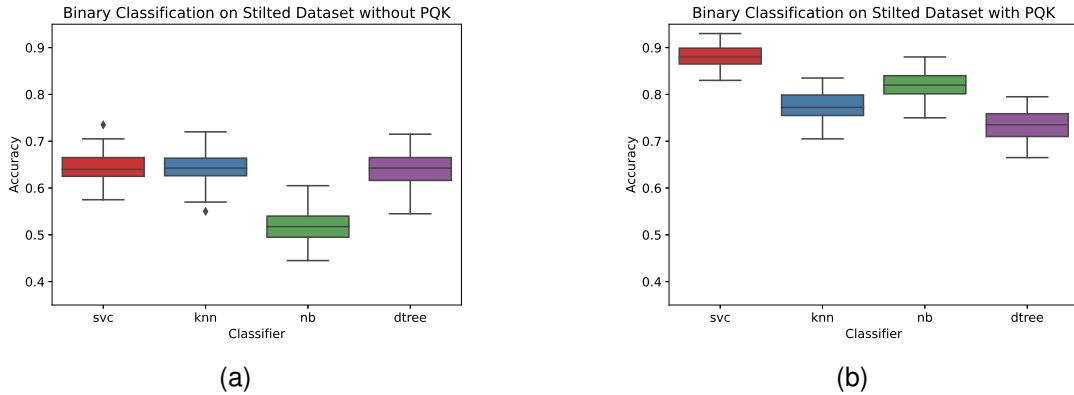


Fig. 1. 5-fold Cross Validation accuracy on relabeled data for classes “Broadleaf tree cover — 82”, and “Herbaceous vegetation — 102”. (a) Without PQK features. (b) With PQK features.

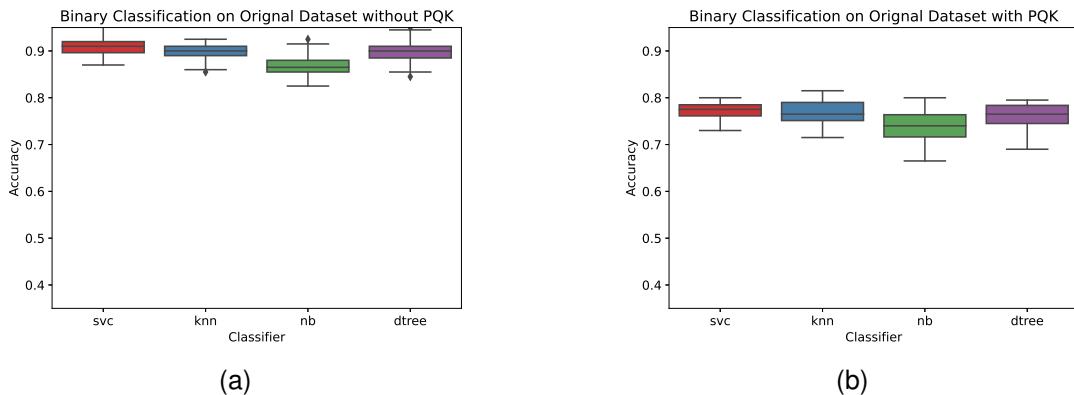


Fig. 2. 5-fold Cross Validation accuracy on original data for classes “Broadleaf tree cover — 82”, and “Herbaceous vegetation — 102”. (a) Without PQK features. (b) With PQK features.

from Space Research Centre of the Polish Academy of Sciences, E. Krätzschmar and S. Guüther from IAB GmbH.

REFERENCES

- [1] B. Ferreira, M. Iten, and R. G. Silva, “Monitoring sustainable development by means of earth observation data and machine learning: a review,” *Environmental Sciences Europe*, vol. 32, no. 1, p. 120, 2020. [Online]. Available: <https://doi.org/10.1186/s12302-020-00397-4>
- [2] P. Bauer, B. Stevens, and W. Hazeleger, “A digital twin of earth for the green transition,” *Nature Climate Change*, vol. 11, no. 2, pp. 80–83, 2021. [Online]. Available: <https://doi.org/10.1038/s41558-021-00986-y>
- [3] E. E. Online, “Working towards AI and Earth observation.” March 2019.
- [4] P. Zikopoulos and C. Eaton, *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. McGraw-Hill Osborne Media: New York, NY, USA, 2011.
- [5] A. Vail, S. Comai, and M. Matteucci, “Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review,” *Remote Sensing*, vol. 12, no. 15, 2020.
- [6] L. Zhang and B. Du, “Deep learning for remote sensing data: A technical tutorial on the state of the art,” *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, pp. 22–40, 2016.
- [7] N. Audebert, A. Boulch, H. Randrianarivo, B. Le Saux, M. Ferecatu, S. Lefèvre, and R. Marlet, “Deep learning for urban remote sensing,” in *2017 Joint Urban Remote Sensing Event (JURSE)*, 2017, pp. 1–4.
- [8] X. X. Zhu, D. Tuia, L. Mou, G. Xia, L. Zhang, F. Xu, and F. Fraundorfer, “Deep learning in remote sensing: A comprehensive review and list of resources,” *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 8–36, 2017.
- [9] G. Cheng, X. Xie, J. Han, L. Guo, and G. S. Xia, “Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 3735–3756, 2020.
- [10] P. Gawron and S. Lewiński, “Multi-spectral image classification with quantum neural network,” in *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 2020, pp. 3513–3516.
- [11] D. A. Zaidenberg, A. Sebastianelli, D. Spiller, B. Le Saux, and S. L. Ullo, “Advantages and bottlenecks of quantum machine learning for remote sensing,” in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 2021, pp. 5680–5683.
- [12] S. Otgonbaatar and M. Datcu, “Assembly of a coresnet of earth observation images on a small quantum computer,” *Electronics*, vol. 10, no. 20, 2021. [Online]. Available: <https://www.mdpi.com/2079-9292/10/20/2482>
- [13] A. Sebastianelli, D. A. Zaidenberg, D. Spiller, B. Le Saux, and S. L. Ullo, “On circuit-based hybrid quantum neural networks for remote sensing imagery classification,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 565–580, 2022.
- [14] S. Otgonbaatar and M. Datcu, “Classification of remote sensing images with parameterized quantum gates,” *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.
- [15] S. Y. Chang, B. L. Saux, S. Vallecorsa, and M. Grossi, “Quantum convolutional circuits for earth observation image classification,” in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022, pp. 4907–4910.
- [16] M. K. Gupta, M. Beseda, and P. Gawron, “How quantum computing-friendly multispectral data can be?” in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022, pp. 4153–4156.

- [17] J. Preskill, "Quantum Computing in the NISQ era and beyond," *Quantum*, vol. 2, p. 79, Aug. 2018. [Online]. Available: <https://doi.org/10.22331/q-2018-08-06-79>
- [18] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, "Characterizing quantum supremacy in near-term devices," *Nature Physics*, vol. 14, no. 6, pp. 595–600, 2018. [Online]. Available: <https://doi.org/10.1038/s41567-018-0124-x>
- [19] F. A. et al., "Quantum supremacy using a programmable superconducting processor," *Nature*, vol. 574, no. 7779, pp. 505–510, 2019. [Online]. Available: <https://doi.org/10.1038/s41586-019-1666-5>
- [20] E. Farhi and H. Neven, "Classification with quantum neural networks on near term processors," 2018.
- [21] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, "Supervised learning with quantum-enhanced feature spaces," *Nature*, vol. 567, no. 7747, pp. 209–212, 2019. [Online]. Available: <https://doi.org/10.1038/s41586-019-0980-2>
- [22] M. Schuld and N. Killoran, "Quantum machine learning in feature hilbert spaces," *Phys. Rev. Lett.*, vol. 122, p. 040504, Feb 2019. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.122.040504>
- [23] H. Y. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven, and J. R. McClean, "Power of data in quantum machine learning," *Nature Communications*, vol. 12, no. 1, p. 2631, 2021. [Online]. Available: <https://doi.org/10.1038/s41467-021-22539-9>
- [24] A. Vali, S. Comai, and M. Matteucci, "Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review," *Remote Sensing*, vol. 12, no. 15, 2020. [Online]. Available: <https://www.mdpi.com/2072-4292/12/15/2495>
- [25] S. Talukdar, P. Singha, S. Mahato, S. Pal, Y.-A. Liou, and A. Rahman, "Land-use land-cover classification by machine learning classifiers for satellite observations—a review," *Remote Sensing*, vol. 12, no. 7, p. 1135, 2020.
- [26] P. Ghamisi, N. Yokoya, J. Li, W. Liao, S. Liu, J. Plaza, B. Rasti, and A. Plaza, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 37–78, 2017.
- [27] M. Romaszewski, P. Głomb, and M. Cholewa, "Semi-supervised hyperspectral classification from a small number of training samples using a co-training approach," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 121, pp. 60–76, 2016.
- [28] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning," *Nature*, vol. 549, no. 7671, pp. 195–202, 2017. [Online]. Available: <https://doi.org/10.1038/nature23474>
- [29] V. Dunjko and H. J. Briegel, "Machine learning & artificial intelligence in the quantum domain: a review of recent progress," vol. 81, no. 7, p. 074001, 2018. [Online]. Available: <http://dx.doi.org/10.1088/1361-6633/aab406>
- [30] H. Y. Huang, R. Kueng, and J. Preskill, "Predicting many properties of a quantum system from very few measurements," *Nature Physics*, vol. 16, no. 10, pp. 1050–1057, 2020. [Online]. Available: <https://doi.org/10.1038/s41567-020-0932-7>
- [31] H. F. Trotter, "On the product of semi-groups of operators," *Proceedings of the American Mathematical Society*, vol. 10, no. 4, pp. 545–551, 1959.
- [32] M. Suzuki, "Generalized trotter's formula and systematic approximants of exponential operators and inner derivations with applications to many-body problems," *Communications in Mathematical Physics*, vol. 51, no. 2, pp. 183–190, 1976. [Online]. Available: <https://doi.org/10.1007/BF01609348>
- [33] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [34] P. Ghamisi, J. Plaza, Y. Chen, J. Li, and A. J. Plaza, "Advanced spectral classifiers for hyperspectral images: A review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 1, pp. 8–32, 2017.
- [35] M. Berger, J. Moreno, J. A. Johannessen, P. F. Levelt, and R. F. Hanssen, "Esa's sentinel missions in support of earth system science," *Remote Sensing of Environment*, vol. 120, pp. 84–90, 2012, the Sentinel Missions - New Opportunities for Science. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S003442571200065X>
- [36] M. Campos-Taberner, F. J. García-Haro, B. Martínez, E. Izquierdo-Verdiguier, C. Atzberger, G. Camps-Valls, and M. A. Gilabert, "Understanding deep learning in land use classification based on sentinel-2 time series," *Scientific Reports*, vol. 10, no. 1, p. 17188, 2020. [Online]. Available: <https://doi.org/10.1038/s41598-020-74215-5>
- [37] A. Benavoli, G. Corani, J. Demšar, and M. Zaffalon, "Time for a change: a tutorial for comparing multiple classifiers through Bayesian analysis," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 2653–2688, 2017.