

Quantum Transfer Learning for Real-World, Small, and High-Dimensional Remotely Sensed Datasets

Soronzonbold Otgonbaatar , Gottfried Schwarz , Mihai Datcu , *Fellow, IEEE*, and Dieter Kranzlmüller

Abstract—Quantum machine learning (QML) models promise to have some computational (or quantum) advantage for classifying supervised datasets (e.g., satellite images) over some conventional deep learning (DL) techniques due to their expressive power via their local effective dimension. There are, however, two main challenges regardless of the promised quantum advantage. Currently available quantum bits (qubits) are very small in number, whereas real-world datasets are characterized by hundreds of high-dimensional elements (i.e., features). In addition, there is not a single unified approach for embedding real-world high-dimensional datasets in a limited number of qubits. Some real-world datasets are too small for training intricate QML networks. Hence, to tackle these two challenges for benchmarking and validating QML networks on real-world, small, and high-dimensional datasets in one go, we employ quantum transfer learning comprising a classical VGG16 layer and a multiqubit QML layer. We use real-amplitude and strongly-entangling N-layer QML networks with and without data reuploading layers as a multiqubit QML layer, and evaluate their expressive power quantified by using their local effective dimension and the lower local effective dimension of a QML network, the better its performance on unseen data. As datasets, we utilize Eurosat and synthetic datasets (i.e., easy-to-classify datasets), and an UC Merced Land use dataset (i.e., a hard-to-classify dataset). Our numerical results show that the strongly-entangling N-layer QML network has a lower local effective dimension than the real-amplitude QML network and outperforms it on the hard-to-classify datasets. In addition, quantum transfer learning helps tackle the two challenges mentioned above for benchmarking and validating QML networks on real-world, small, and high-dimensional datasets.

Index Terms—Data reuploading, earth observation, image classification, quantum machine learning (QML), quantum transfer learning, remote sensing.

I. INTRODUCTION

UNIVERSAL quantum computers are composed of a collection of quantum bits (qubits) and parameterized quantum gates being arranged according to some given topologies,

Manuscript received 3 July 2023; revised 9 August 2023 and 7 September 2023; accepted 11 September 2023. Date of publication 18 September 2023; date of current version 13 October 2023. (Corresponding author: Soronzonbold Otgonbaatar.)

Soronzonbold Otgonbaatar is with the Remote Sensing Technology Institute, German Aerospace Center DLR, 82234 Weßling, Germany, and also with the Ludwig-Maximilians-Universität München, 80539 München, Germany (e-mail: soronzonbold.otgonbaatar@dlr.de).

Gottfried Schwarz and Mihai Datcu are with the Remote Sensing Technology Institute, German Aerospace Center DLR, 82234 Weßling, Germany (e-mail: gottfried.schwarz@dlr.de; mihai.datcu@dlr.de).

Dieter Kranzlmüller is with the Ludwig-Maximilians-Universität München, 80539 München, Germany (e-mail: kranzlmue@ifi.lmu.de).

Digital Object Identifier 10.1109/JSTARS.2023.3316306

while quantum learning algorithms are algorithms manipulating qubits by using parameterized quantum gates. Based on the learnable parameters of parameterized quantum gates, quantum machine learning (QML) as described by the authors in [1] and [2], in general, contains in the following three different subdirections [3].

- 1) *Quantum-Inspired ML*: Develop novel artificial intelligence (AI/DL) techniques by using concepts from quantum information processing [4] and [5].
- 2) *Quantum ML (sometimes)*: Apply classical DL techniques to quantum data (quantum chemistry) [6].
- 3) *Quantum-Applied ML*: Develop quantum learning circuits for supervised real-world data on quantum computers [7], [8], and [9].

In this work, we focus on quantum-applied ML, and we use QML interchangeably with quantum-applied ML. A QML network encodes input data in qubits and learns the parameters of parameterized quantum gates. Moreover, it promises quantum advantage for some computational problems over conventional learning methods due to either its expressive power measured by the local effective dimension [10] and [11], or its computational time [12].

However, the qubits of currently available universal quantum computers are noisy and small in number. Hence, these types of universal quantum computers are called noisy intermediate-scale quantum (NISQ) computers [13]. Due to the limited number of qubits, there are the following two main challenges.

- 1) *Embedding challenge*: There is not a single unified approach for embedding real-world, high-dimensional data points in a small number of qubits.
- 2) *Small dataset challenge*: QML networks do not capture informative patterns in small datasets in contrast to big datasets, and this challenge even exists with conventional DL methods.

Here, we name a dataset as a low-dimensional dataset if and only if we can represent its data points by at most five elements using a classical dimensionality-reduction technique, and as a high-dimensional dataset, otherwise.

To overcome the embedding challenge, some studies already proposed an embedding strategy for a toy dataset [14], as well as for real-world, big, and low-dimensional datasets [9]. Otgonbaatar and Datcu [9] investigated a binary-label classic-quantum classifier for embedding and classifying a two-label, low-dimensional Eurosat dataset [15], in which they classified this specific binary dataset by measuring directly the output of the quantum layer. On the other hand, Sebastianelli et al. [16]

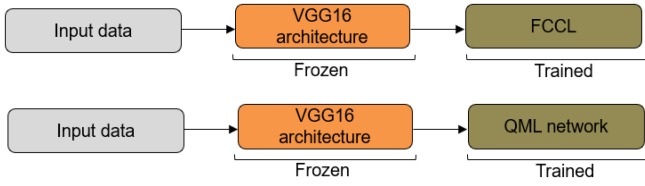


Fig. 1. (Top) Classical transfer learning: Input data, a DL VGG16 architecture, a FCCL, and (Bottom) quantum transfer learning: Input data, a DL VGG16 architecture, a QML network.

focused on a multilabel classic-quantum-classic classifier, in which they classified the same Eurosat dataset by measuring the last classical layer but not a *quantum layer* for a multiclass case. Furthermore, Otgonbaatar and Datcu [17] introduced a single qubit and multiqubit quantum classifier for embedding a selected practical dataset in a parameterized quantum circuit (PQC).

To overcome the *small dataset* challenge, Mari et al. [18] proposed a novel method named quantum transfer learning, the quantum version of classical transfer learning. Classical transfer learning is a procedure for training fully connected classical layers (FCCL) residing on the top level of conventional DL architectures (for small datasets when the parameters of frozen conventional DL architectures are initialized a priori on big similar datasets). In contrast, quantum transfer learning trains a QML network placed on the top level of frozen conventional DL architectures instead of FCCL (see Fig. 1). In particular, for quantum transfer learning, one replaces the fully connected top layers of frozen conventional DL architectures by a multiqubit QML network. One also profits from the advantage of quantum transfer learning since it simultaneously helps tackle the two challenges mentioned above.

For practically important datasets, the authors in [9] and [16] implicitly used a quantum transfer learning method, but they employed a real-world, *big*, and *low-dimensional* dataset (i.e., an Eurosat dataset). In this work, we explicitly propose and employ a quantum transfer learning method for real-world, *small*, and *high-dimensional* datasets. Our proposed quantum transfer learning method consists of a multiqubit QML network (with or without *data reuploading*), and a very deep convolutional network (in our case, a VGG16 architecture), playing the role of a feature extractor from datasets, as shown in Fig. 1 [Bottom], [19], [20], and [21]. In particular, we employ real-amplitude and strongly-entangling N-layer QML networks with and without *data reuploading* layers [22] as a multiqubit QML network, and quantify their expressive power by using their *local effective dimension* (the lower, the better), which gives us a portion of the active parameters in the trained QML network [10], [11], [23], [24], and [25] and currently available quantum resources. The expressive power is referred to as the capacity of any learning model, and its capability to capture intricate patterns in any dataset. Moreover, the lower the local effective dimension of a given QML network, the better its performance on still unseen data points.

Our practical datasets are synthetic, Eurosat, and UC Merced Land Use images [26]. First, we compute the local effective

dimension of our QML networks, that is, real-amplitude and strongly-entangling N-layer QML networks without *data reuploading* layers, on two-class low-dimensional synthetic and Eurosat datasets since we can compress and represent these low-dimensional datasets by three and four elements using a classical DL network as proposed in [9], respectively, and here, we generated and used two-class synthetic data including 100 data points, where each data point is a 2-D vector drawn from circular data with the error according to a normal distribution $\mathcal{N}(\mu = 0, \sigma = 1)$. For low-dimensional Eurosat data, we utilized its *annual crop* and *residential* area classes, and this two-class set consists of 4000 data points, each of which is characterized by $64 \times 64 \times 3$ low-dimensional elements. To validate the relationship between local effective dimension and performance (i.e., the classification accuracy) of a given QML network, we trained, subsequently, our QML networks via quantum transfer learning on the small, high-dimensional, and hard-to-classify three-class images, i.e., *dense residential*, *medium residential*, and *sparse residential* area classes of the high-dimensional UC Merced Land Use dataset. In addition, a *dense residential*, *medium residential*, and *sparse residential* area classification problem meets our two challenges mentioned, because these three-class images consist of only 288 data points, and each image is characterized by $256 \times 256 \times 3$ high-dimensional elements [26], [27]. As a quantum simulator, we used a PennyLane Python library for training our QML networks [28].

Our experimental results demonstrate that the strongly-entangling N-layer QML network without *data reuploading* layers has a lower local effective dimension than the real-amplitude N-layer QML network without *data reuploading* layers. It also has a higher test accuracy on real-world datasets than the real-amplitude QML network. Furthermore, quantum transfer learning helps tackle the two abovementioned challenges for benchmarking and validating different QML networks on real-world, small, high-dimensional, and hard-to-classify datasets.

The rest of this article is organized as follows. In Sections II and III, we provide some background for quantum transfer learning and multiqubit QML networks, respectively. In Section IV, we present the expressive power of QML networks via their local effective dimension. Subsequently, in Section V, we introduce practical datasets being used in this article. In Section VI, we present our experiments and some of our findings. Finally, Section VII concludes this article.

II. QUANTUM TRANSFER LEARNING

Quantum transfer learning is referred to as training a QML network with and without *data reuploading* on real-world small datasets when the weights of the VGG16 architecture are initialized on the *ImageNet* dataset. Moreover, we froze the weights of the VGG16 such that none of its weights were updated during the training of a QML network. The QML network we propose in this study is a multiqubit N-layer quantum classifier (with and without *data reuploading*) and these classifiers are extremely simple as well as very powerful learning networks for nonlinear datasets [19],[20].

A. Single-Qubit QML Network With and Without Data Reuploading

We characterize a single-qubit QML network by a tensorial feature map and universal parameterized quantum gates as

$$U(\phi_1, \phi_2, \phi_3) V_i(\theta_{i,1}, \gamma_{i,2}, \phi_{i,3}) \quad (1)$$

where ϕ_1, ϕ_2 , and ϕ_3 embed input data points in qubits, and $\theta_{i,1}, \gamma_{i,2}$, and $\phi_{i,3}$ are learning weights at the i th layer. Namely, we encode an input data point having three features in qubits using a tensorial feature map, $U(\phi_1, \phi_2, \phi_3) = U(\phi_1) \otimes U(\phi_2) \otimes U(\phi_3)$, and subsequently, we train a parameterized quantum gate $V_i(\theta_{i,1}, \gamma_{i,2}, \phi_{i,3})$ on the embedded data points. For simplicity, we express $U = U(\phi_1, \phi_2, \phi_3)$ and $V_i = V_i(\theta_{i,1}, \gamma_{i,2}, \phi_{i,3}) = V_i(\theta_{i,1}) V_i(\gamma_{i,2}) V_i(\phi_{i,3})$ since they are unitary quantum gates such that $V_i = V_i^\dagger, i = 1, \dots, N$, where N is the number of layers or the depth of a given QML network [17]. In matrix form, we express U and the V_i 's as

$$U(\phi_j) = e^{i\phi_j} \begin{pmatrix} \cos(\phi_j) & -\sin(\phi_j) \\ \sin(\phi_j) & \cos(\phi_j) \end{pmatrix}, \quad j = 1, 2, 3 \quad (2)$$

and

$$V_i(\theta_{i,1}) = \begin{pmatrix} e^{i\theta_{i,1}} & 0 \\ 0 & e^{-i\theta_{i,1}} \end{pmatrix}, \quad V_i(\gamma_{i,2}) = \begin{pmatrix} e^{i\gamma_{i,2}} & 0 \\ 0 & e^{-i\gamma_{i,2}} \end{pmatrix} \quad (3)$$

and

$$V_i(\phi_{i,3}) = \begin{pmatrix} \cos(\phi_{i,3}) & -\sin(\phi_{i,3}) \\ \sin(\phi_{i,3}) & \cos(\phi_{i,3}) \end{pmatrix}, \quad i = 1, \dots, N. \quad (4)$$

For a QML network with *data reuploading*, the quantum circuit UV_i expressed by (1) is repeated N times

$$\prod_{i=1}^N UV_i = \underbrace{UV_1}_{L_1} \underbrace{UV_2}_{L_2} \dots \underbrace{UV_{N-1}}_{L_{N-1}} \underbrace{UV_N}_{L_N} \quad (5)$$

where an input data point is reuploaded N times in the quantum gate U , and the parameterized quantum gate V_N is also repeated N times, as shown in Fig. 2. The N th layer is denoted as L_N .

Alternatively, for a QML network without *data reuploading*, the quantum circuit UV_i expressed in (1) is repeated N times

$$U \prod_{i=1}^N V_i = \underbrace{UV_1}_{L_1} \underbrace{V_2}_{L_2} \dots \underbrace{V_{N-1}}_{L_{N-1}} \underbrace{V_N}_{L_N} \quad (6)$$

where an input data point is uploaded in U only once at layer L_1 , as shown in Fig. 3.

For single-qubit QML networks, a qubit is represented as

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (7)$$

and as a bra vector $\langle 0| = |0\rangle^\dagger$ and $\langle 1| = |1\rangle^\dagger$, where † represents both transpose and conjugate. In general, qubits can exist in a superposition as

$$|\psi\rangle = c_1|0\rangle + c_2|1\rangle \quad \text{such that} \quad |c_1|^2 + |c_2|^2 = 1 \quad (8)$$

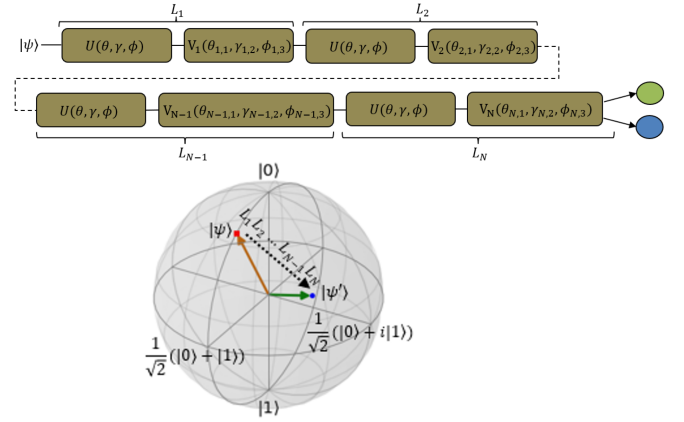


Fig. 2. (Top) Single-qubit N -layer QML network with *data reuploading*. (Bottom) A QML network operation on a single-qubit represented on a Bloch sphere, in which $|\psi\rangle$ is an input single-qubit, and $|\psi'\rangle$ is the single-qubit after $L_1, L_2, \dots, L_{N-1}, L_N$ quantum layers.

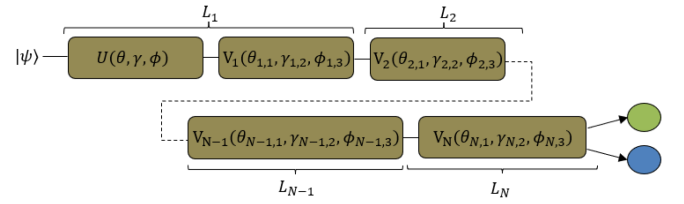


Fig. 3. Single-qubit N -layer QML network without *data reuploading*.

where c_1 and c_2 are complex numbers, as shown in Fig. 2 [Bottom].

Let us consider a simple example for a single-qubit *data reuploading* QML network: We assume $|\psi\rangle = |1\rangle$ and a QML network with one layer L_1 , $(0, 0, \phi_3) = (0, 0, \pi/2)$ and $(\theta_{1,1}, \gamma_{1,2}, \phi_{1,3}) = (0, 0, 0)$. A qubit after layer L_1 becomes

$$|\psi'\rangle = V_1(0, 0, 0)U(\pi/2)|1\rangle. \quad (9)$$

If we measure a single-qubit by using a projective measurement $\hat{z} = |0\rangle\langle 0| - |1\rangle\langle 1|$, (i.e., the outer product of $|0\rangle$ and $|1\rangle$), we obtain an expected value of measuring the state $|\psi'\rangle$ in the basis \hat{z}

$$\langle \hat{z} \rangle_{\bar{\theta}} = \langle \psi' | \hat{z} | \psi' \rangle = e^{-i\pi} e^{i\pi} \langle 0 | \hat{z} | 0 \rangle = 1 \quad (10)$$

where $\bar{\theta} = (\theta_{i,1}, \gamma_{i,2}, \phi_{i,3})^T \in \Theta$ and T denotes the transpose operation of a vector. In the end, this expected value is connected to two classical neurons shown in Fig. 2, and these neurons denoted as $l = 1, 2$ are defined as

$$\tilde{y}_l = A(a_l + w_l \langle \hat{z} \rangle_{\bar{\theta}}) \quad (11)$$

where $A(\cdot)$ is a nonlinear activation function, that is, a sigmoid function, \tilde{y}_l is their predicted output, a_l is their bias, and w_l is their edge parameter [17].

In general, we express the expected value of a single-qubit N -layer QML network with and without *data reuploading* into the basis $\hat{z} = |0\rangle\langle 0| - |1\rangle\langle 1|$ as

$$\langle \hat{z} \rangle_{\bar{\theta}} = \langle \psi | L_1^\dagger L_2^\dagger \dots L_{N-1}^\dagger L_N^\dagger \hat{z} L_N L_{N-1} \dots L_2 L_1 | \psi \rangle \quad (12)$$

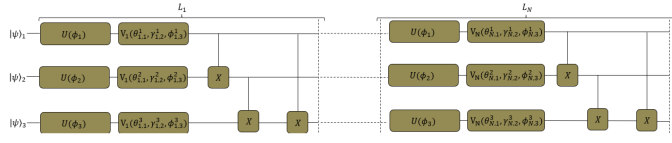


Fig. 4. Multiqubit N-layer QML network with *data reuploading*: L_1, \dots, L_N quantum layers.

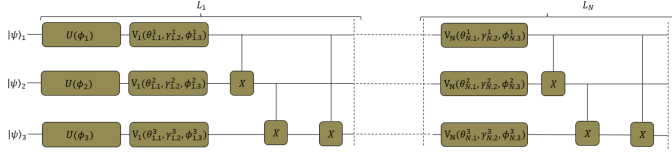


Fig. 5. Multiqubit N-layer QML network without *data reuploading*: L_1, \dots, L_N quantum layers.

yielding a continuous value from -1 to $+1$. We then, connect two classical neurons to the expected value $\langle \hat{z} \rangle_{\vec{\theta}}$ as shown in (11).

III. OUR MULTIQUBIT QML NETWORKS WITH AND WITHOUT DATA REUPLOADING

A multiqubit QML network is a PQC with some input qubits and parameterized quantum gates exploiting (2) and (3). We encoded our data in its input qubits by utilizing a tensorial feature map expressed by (2) as proposed by the authors in [14], [19], and [20], and we trained its parameterized quantum gates following (3). Moreover, we name a multiqubit QML network with N layers as a multiqubit N -layer QML network similar to the single-qubit N -layer QML network mentioned in the previous section.

In this work, we use three-qubit real-amplitude and strongly-entangling N -layer QML networks with and without *data reuploading* layers. We express the parameterized quantum gates of these QML networks with *data reuploading* layers as

$$\prod_{i=1}^N \underbrace{U(\phi_1, \phi_2, \phi_3) V_i(\theta_{i,1}^q, \gamma_{i,2}^q, \phi_{i,3}^q) W_i}_{L_i}, \quad q = 1, 2, 3 \quad (13)$$

where q represents the qubit number $|\psi\rangle_q$ and each W_i denotes controlled-X quantum gates (see Fig. 4). A controlled-X quantum gate is a two-qubit quantum gate acting on a target qubit if and only if a control qubit is in the state $|1\rangle$. In contrast, these QML networks without *data reuploading* layers are characterized as

$$U(\phi_1, \phi_2, \phi_3) \prod_{i=1}^N \underbrace{V_i(\theta_{i,1}^q, \gamma_{i,2}^q, \phi_{i,3}^q) W_i}_{L_i} \quad (14)$$

where the classical data is uploaded once at the layer L_1 as illustrated in Fig. 5.

A. Real-Amplitude N -Layer QML Networks With and Without Data Reuploading Layers

Our real-amplitude N -layer QML networks with and without *data reuploading* layers are composed of a quantum gate U encoding a data point in qubits, parameterized quantum gates V_i , and nonparameterized quantum gates W_i , respectively, as

$$\prod_{i=1}^N \underbrace{U(\phi_1, \phi_2, \phi_3) V_i(0, 0, \phi_{i,3}^q) W_i}_{L_i} \quad (15)$$

and

$$U(\phi_1, \phi_2, \phi_3) \prod_{i=1}^N \underbrace{V_i(0, 0, \phi_{i,3}^q) W_i}_{L_i} \quad (16)$$

where W_i represents all-to-all entanglement [16].

B. Strongly-Entangling N -Layer QML Networks With and Without Data Reuploading Layers

Our strongly-entangling N -layer QML networks with and without *data reuploading* layers are defined as

$$\prod_{i=1}^N \underbrace{U(\phi_1, \phi_2, \phi_3) V_i(\theta_{i,1}^q, \gamma_{i,2}^q, \phi_{i,3}^q) W_i}_{L_i} \quad (17)$$

and

$$U(\phi_1, \phi_2, \phi_3) \prod_{i=1}^N \underbrace{V_i(\theta_{i,1}^q, \gamma_{i,2}^q, \phi_{i,3}^q) W_i}_{L_i} \quad (18)$$

where W_i represents strong entanglement [22].

IV. POWER OF QUANTUM MACHINE LEARNING NETWORKS

Deep neural networks are powerful learning models for identifying intricate patterns in big datasets. Their power is quantified by using the so-called local effective dimension, which yields a portion of active parameters in the trained network [11], [24]. Quantum neural networks, i.e., QML networks, are novel learning models based on PQCs exploiting quantum superposition, entanglement, and the interference of qubits. The authors in [10] and [11] demonstrated that some QML networks have a lower local effective dimension—a lower local effective dimension and being more powerful—for analyzing some data-driven tasks than their classical counterparts. In particular, the local effective dimension gives us the active parameters θ^* in the trained model and the lower effective dimension of a learning model (classical or quantum), the better it generalizes on unseen data points.

According to [10], [11], and [29], the local effective dimension of a QML network $\langle \hat{z} \rangle_{\vec{\theta}}$ around the active parameters $\vec{\theta}^* \in \Theta \subset \mathbb{R}^d$ with n training data points, where Θ weight space, is defined as

$$d_{n,\lambda}(\langle \hat{z} \rangle_{\vec{\theta}}) = \frac{2 \log \left(\frac{1}{V_\epsilon} \int_{B_\epsilon(\vec{\theta}^*)} \sqrt{\det(I_d + k_{n,\lambda} \bar{F}(\vec{\theta}))} d\vec{\theta} \right)}{\log k_{n,\lambda}} \quad (19)$$

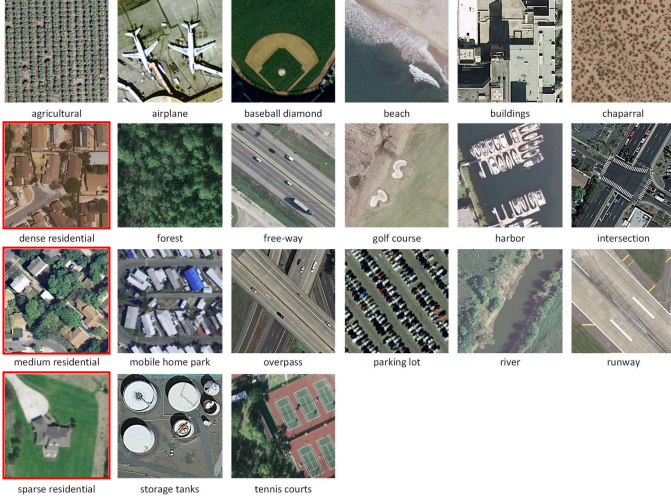


Fig. 6. Some examples of the high-dimensional UC Merced Land Use dataset taken from [26]. Here, we contoured its hard-to-classify three-class examples in red.



Fig. 7. Annual crop and residential area classes of the low-dimensional Eurosat dataset.

where $\mathcal{B}_\epsilon(\vec{\theta}^*) := \{\vec{\theta} \in \Theta : \|\vec{\theta} - \vec{\theta}^*\| \leq \epsilon\}$ is an ϵ -ball with a volume V_ϵ , $\epsilon > 1/\sqrt{n}$, $k_{n,\lambda} = \frac{\lambda n}{2\pi \log n}$, $\lambda \in (\frac{2\pi \log n}{n}, 1]$, I_d is a unit diagonal matrix, and $\bar{F}(\vec{\theta}) \in \mathbb{R}^{d \times d}$ is the normalized Fisher information matrix of $F(\vec{\theta})$ [30] with an element as

$$\bar{F}_{ij}(\vec{\theta}) = \frac{d \cdot V_\epsilon}{\int_{\mathcal{B}_\epsilon(\vec{\theta}^*)} \text{Tr}(F(\vec{\theta})) d\vec{\theta}} F_{ij}(\vec{\theta}). \quad (20)$$

V. OUR DATASETS

We generated and used a synthetic dataset using

$$x_m = r_m \begin{pmatrix} \cos \phi_m \\ \sin \phi_m \end{pmatrix} + \begin{pmatrix} \epsilon_m^x \\ \epsilon_m^y \end{pmatrix} \quad (21)$$

where $r_m = 1$ if $y_m = 1$, and $r_m = 0.15$ if $y_m = 0$, ϵ_m 's are a normal distribution $\mathcal{N}(\mu = 0, \sigma = 1)$, and $\phi_m \in (0, 2\pi]$ linearly spaced [31]. This synthetic dataset is a two-class dataset composed of $m = 100$ data points.

A Eurosat dataset is a Sentinel-2 image dataset with 27 000 labeled and georeferenced images. In addition, this dataset is a patch-based dataset with 64×64 pixels comprising ten classes, where each image is characterized by 13 spectral bands ranging from 443 to 2190 nm, and having a spatial resolution of 60 m/pixel (see Fig. 7). We used selected two-class image scenes, namely, the *annual crop* and *residential* area classes, consisting of 4000 images each of which is characterized by $64 \times 64 \times 3$ low-dimensional elements.

UC Merced Land Use data contain image scenes of 21 classes with three RGB spectral bands. Its smallest class, *tennis court*, includes 42 data points, whereas its largest classes (e.g., *agricultural*), include 100 data points. In total, this dataset comprises 2100 image scenes [26], [27]. By visual inspection and experiment, we learned that the hard-to-classify three-class examples of this dataset are the *dense residential*, *medium residential*, and *sparse residential* area classes, as shown in Fig. 6. For our tests, we employed these hard-to-classify three-class examples comprising 288 image scenes, where each image is a $256 \times 256 \times 3$ high-dimensional elements.

VI. OUR EXPERIMENTS

In this section, we define the local effective dimension and the classification performance of our QML networks, that is, real-amplitude and strongly-entangling N-layer QML networks, while utilizing synthetic and Eurosat datasets. In particular, we derive the relationship between the local effective dimension and the classification accuracy of our QML networks on synthetic and real-world datasets. In order to validate the relationship between the local effective dimension and the classification power of our QML networks, we train them via quantum transfer learning on the real-world, small, and high-dimensional three-class dataset (the hard-to-classify dataset), i.e., *dense residential*, *medium residential*, and *sparse residential* area classes, of the UC Merced Land Use dataset, because this three-class dataset meets the following two challenges mentioned before.

- 1) *Embedding challenge*: The three-class images are characterized by $256 \times 256 \times 3$ high-dimensional elements, while currently available quantum computers already have around 50 qubits.
- 2) *Small dataset challenge*: The three-class dataset consists of only 288 image scenes, and we split it into a training dataset comprising 201 image scenes, and a test dataset consisting of 87 image scenes.

In particular, we benchmarked and validated our QML networks placed on the top level of the frozen VGG16 network on the hard-to-classify three-class dataset, because these datasets play a very important role in QML [32].

A. Local Effective Dimension and Performance of Our QML Networks Without Data Reuploading

In our numerical study, we employed N-layer QML networks with only three and four qubits in order to keep the quantum resources as low as possible when $N = 3, 6$, or 9 . For the synthetic dataset, we computed the local effective dimension of real-amplitude and strongly-entangling N-layer QML networks without *data reuploading* while increasing the size of the dataset. We plotted the local effective dimension of our QML networks with respect to the dataset size shown in Fig. 8 and present their classification accuracy in Table I. This result leads to the conclusion that the local effective dimension of QML networks seems to correlate with their classification accuracy. Furthermore, we discovered that strongly-entangling N-layer QML networks have a lower effective dimension and, at the

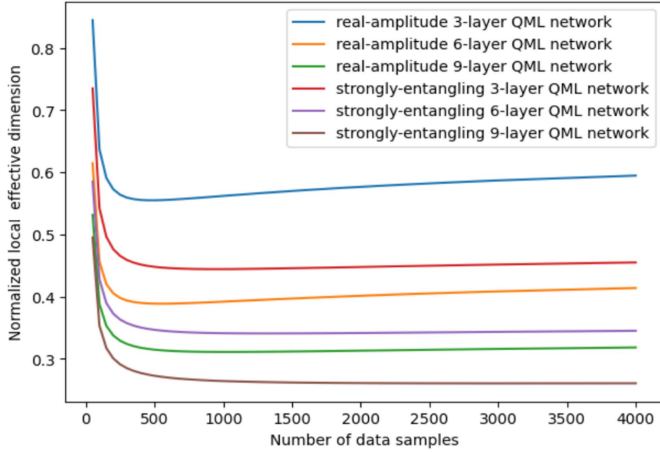


Fig. 8. Normalized local effective dimension of real-amplitude and strongly-entangling N -layer QML networks without *data reuploading* with respect to the number of data samples.

TABLE I

CLASSIFICATION ACCURACY OF REAL-AMPLITUDE AND STRONGLY-ENTANGLING N -LAYER QML NETWORKS WITHOUT DATA REUPLOADING ON THE TWO-CLASS SYNTHETIC DATASET

{Class}/ N -layer	Real-amplitude			Strongly-entangling		
	3	6	9	3	6	9
{1, 2}	0.35	0.45	0.40	0.90	0.95	1.00

TABLE II

CLASSIFICATION ACCURACY OF N -DEPTH REAL-AMPLITUDE AND STRONGLY-ENTANGLING NETWORKS WITHOUT A DATA REUPLOADING LAYER ON THE ANNUAL CROP AND RESIDENTIAL AREA CLASSES OF THE EUROSAT DATASET

{Class}/ N -depth	Real-amplitude			Strongly-entangling		
	3	6	9	3	6	9
{1, 2}	0.70	0.73	0.75	0.74	0.72	0.70

Here, {1, 2} represents the annual crop and residential area classes.

same time, a higher accuracy for generating two-class labels than their counterpart real-amplitude QML networks.

To validate this conclusion for a real-world, big, and low-dimensional dataset comprising the *annual crop* and *residential* area classes of the Eurosat dataset, we first applied a so-called two-level encoding, which maps each image scene characterized by $64 \times 64 \times 3$ low-dimensional elements to 2×2 informative features using the VGG16 architecture, and encoded these informative features in four qubits by employing a tensorial feature map [9]. Then, we calculated numerically the local effective dimension of the real-amplitude and strongly-entangling N -layer QML networks with a 3200-element training set and an 800-element testing set randomly sampled from the *annual crop* and *residential* area classes. We found that the strongly-entangling N -layer QML network generates two-class labels similar (and even better in some instances) to the ones generated by the real-amplitude N -layer QML network (see Table II), because the *annual crop* and *residential* area classes have less overlap, which is proven by generating two-class labels and by visual inspection. In our case, this two-class classification problem is an easy-to-classify two-class labelling problem, although the strongly-entangling N -layer QML network has a lower local

Algorithm 1: Quantum Transfer Learning for the Hard-To-Classify Three-Class Images on the PennyLane Simulator.

- 1: INPUT: The hard-to-classify three-class examples
- 2: OUTPUT: The hard-to-classify three-class labels
- 3: QUANTUM TRANSFER LEARNING: a sequential model having the frozen VGG16 layer for extracting informative features from remotely-sensed datasets, and a three-qubit QML layer for training on the output of the frozen VGG16 layer (see Fig. 1 [Bottom])
- 4: TRAINING PARAMETERS: epochs=20, batch=64, and the Adam optimizer having the learning rate of 10^{-4} .
- 5: STOP ALGORITHM

TABLE III

CLASSIFICATION ACCURACY OF N -DEPTH REAL-AMPLITUDE AND STRONGLY-ENTANGLING NETWORKS WITHOUT A DATA REUPLOADING LAYER ON THE DENSE RESIDENTIAL, MEDIUM RESIDENTIAL, AND SPARSE RESIDENTIAL AREA CLASSES OF THE UC MERCED LAND USE DATASET

{Class}/ N -depth	Real-amplitude			Strongly-entangling		
	3	6	9	3	6	9
{1, 2, 3}	0.64	0.67	0.72	0.81	0.83	0.83

Here, {1, 2, 3} represents the dense residential, medium residential, and sparse residential area classes.

TABLE IV

CLASSIFICATION ACCURACY OF N -DEPTH REAL-AMPLITUDE AND STRONGLY-ENTANGLING NETWORKS WITH A DATA REUPLOADING LAYER ON THE DENSE RESIDENTIAL, MEDIUM RESIDENTIAL, AND SPARSE RESIDENTIAL AREA CLASSES OF THE UC MERCED LAND USE DATASET

{Class}/ N -depth	Real-amplitude			Strongly-entangling		
	3	6	9	3	6	9
{1, 2, 3}	0.41	0.52	0.48	0.74	0.66	0.74

Here, {1, 2, 3} represents the dense residential, medium residential, and sparse residential area classes.

effective dimension and is more powerful than the real-amplitude N -layer QML network.

B. Quantum Transfer Learning for Real-World, Small, High-Dimensional, and Hard-to-Classify Three-Class Images

We validated the relationship between the local effective dimension and the classification accuracy of our QML networks with and without *data reuploading* on the real-world, small, and high-dimensional three-class images (i.e., hard-to-classify images) of the UC Merced Land Use dataset, that is, *dense residential*, *medium residential*, and *sparse residential* area classes. To validate this relationship illustrated in Fig. 8, we placed our QML networks with three input qubits on the top layer of the VGG16 network, since we applied our hard-to-classify three-class images split into a 201-element training set and an 87-element testing set, where each data point is characterized by $256 \times 256 \times 3$ high-dimensional elements compared with the big, low-dimensional Eurosat dataset. We summarized our results in Tables III and IV. Here, we could prove that strongly-entangling N -layer networks without *data reuploading* layers outperform real-amplitude N -layer networks without *data reuploading* layers in most instances, except for the $N = 9$ case due to their lower local effective dimension (see Table III and Fig. 8). The poor performance of the strongly-entangling

nine-layer QML network without *data reuploading* layers is caused by their vanishing gradient [33], that is beyond the main scope of this article but an important future research direction. Moreover, real-amplitude and strongly-entangling N-layer networks without *data reuploading* layers outperform (on practical high-dimensional datasets) ones with *data reuploading* layers on the same dataset (see Tables III and IV).

VII. CONCLUSION

We employed and benchmarked real-amplitude and strongly-entangling N-layer QML networks with and without *data reuploading* layers. As practical datasets, we used a two-class synthetic dataset, easy-to-classify two-class images of the Eurosat dataset and hard-to-classify three-class images of the UC Merced Land Use dataset. The hard-to-classify three-class dataset consists of 288 image scenes, where each image scene has $256 \times 256 \times 3$ high-dimensional elements, whereas the easy-to-classify two-class images are composed of 4000 image scenes each of which is characterized by $64 \times 64 \times 3$ low-dimensional elements. In particular, our hard-to-classify three-class images meet the abovementioned two challenges for training multiqubit QML networks: the embedding and the small dataset challenge.

We analyzed the expressive power of real-amplitude and strongly-entangling N-layer QML networks without *data reuploading* layers via their so-called local effective dimension, while utilizing the synthetic and Eurosat datasets. Our numerical experiments proved the statement that the lower the local effective dimension of a multiqubit QML network, the better its classification accuracy on unseen data points. More importantly, we discovered that our strongly-entangling N-layer QML networks have a lower local effective dimension and a higher test accuracy than real-amplitude QML networks (see Fig. 8 and Table I). Here, the local effective dimension of QML networks seems to correlate with their classification performance. We note, however, that for easy-to-classify datasets (in our case, two-class labels of the low-dimensional Eurosat dataset), multiqubit QML networks perform equally well on unseen data points even though one of them is more powerful than other ones (see Table II). Thus, the hard-to-classify datasets are very important datasets for benchmarking and validating QML networks [32].

To validate the relationship between the local effective dimension and the classification performance of real-amplitude and strongly-entangling N-layer QML networks, we trained them via quantum transfer learning on the real-world, small, and hard-to-classify three-class images of the high-dimensional UC Merced Land Use dataset. Our experimental results demonstrate that strongly-entangling N-layer QML networks perform better than real-amplitude N-layer QML networks in most instances (see Table III). Furthermore, real-amplitude and strongly-entangling N-layer QML networks without *data reuploading* outperform ones with *data reuploading*. More importantly, quantum transfer learning even helps tackle the two main challenges in one-go encountered for benchmarking and validating multiqubit QML networks on real-world, small, and high-dimensional datasets of practical importance.

We must note, however, that our message is that we did not attempt to demonstrate the computational advantage of QML models over their conventional counterparts, which is already demonstrated by the authors in [9], [10], [11], and [23] but design and select a powerful model among existing QML models for real-world problems of practical importance. Our contribution is also twofold: 1) We designed QML models and analyzed their expressive power via their local effective dimension since their local effective dimension correlates with their classification capability. 2) We proposed and utilized quantum transfer learning for benchmarking and analyzing QML models on real-world, hard-to-classify datasets because weaker models generate similar performance metrics (e.g., classification accuracy or loss) on real-world, easy-to-classify datasets as more powerful models.

For ongoing and future work, we will integrate faster and simpler QML models with artificial intelligence methodologies based on their quantum resource required [34]. We will invent and design quantum-inspired networks for practical and significant problems to obtain quantum advantage as early and efficiently as possible. More importantly, quantum and quantum-inspired models help boost conventional probabilistic models for remotely sensed datasets. In addition to the computational advantage of a quantum computer, another advantage of quantum computing to remote sensing is that quantum machine learning algorithms are operate inherently on complex vector space [35], and some remote sensing datasets are complex-numbered images. Therefore, we design inherently complex quantum machine learning algorithms for complex-numbered remote sensed images, such as synthetic aperture radar images without the need of modification in the conventional machine and deep learning techniques operating on real number space.

ACKNOWLEDGMENT

We would like to thanks Begüm Demir (Technical University of Berlin) for her valuable comments and suggestions on the very first draft of this article.

REFERENCES

- [1] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning," *Nature*, vol. 549, no. 7671, pp. 195–202, Sep. 2017. [Online]. Available: <https://doi.org/10.1038/nature23474>
- [2] S. Lloyd, M. Mohseni, and P. Rebentrost, "Quantum principal component analysis," *Nature Phys.*, vol. 10, no. 9, pp. 631–633, Sep. 2014. [Online]. Available: <https://doi.org/10.1038/nphys3029>
- [3] V. Dunjko and H. J. Briegel, "Machine learning & artificial intelligence in the quantum domain: A review of recent progress," *Rep. Prog. Phys.*, vol. 81, no. 7, Jun. 2018, Art. no. 074001. [Online]. Available: <https://doi.org/10.1088/1361-6633/aab406>
- [4] E. M. Stoudenmire and D. J. Schwab, "Supervised learning with quantum-inspired tensor networks," 2016. Accessed: 2023. [Online]. Available: <https://arxiv.org/abs/1605.05775>
- [5] T. Felsen et al., "Quantum-inspired machine learning on high-energy physics data," *NPJ Quantum Inf.*, vol. 7, no. 1, Jul. 2021, Art. no. 111. [Online]. Available: <https://doi.org/10.1038/s41534-021-00443-w>
- [6] P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, and P. O. Dral, "Artificial intelligence-enhanced quantum chemical method with broad applicability," *Nature Commun.*, vol. 12, no. 1, Dec. 2021, Art. no. 7022. [Online]. Available: <https://doi.org/10.1038/s41467-021-27340-2>
- [7] D. Willsch, M. Willsch, H. De Raedt, and K. Michielsen, "Support vector machines on the D-Wave quantum annealer," *Comput. Phys. Commun.*, vol. 248, 2020, Art. no. 107006. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S001046551930342X>

- [8] C. Tüysüz et al., "Hybrid quantum classical graph neural networks for particle track reconstruction," *Quantum Mach. Intell.*, vol. 3, no. 2, Nov. 2021, Art. no. 29. [Online]. Available: <https://doi.org/10.1007/s42484-021-00055-9>
- [9] S. Otgonbaatar and M. Datcu, "Classification of remote sensing images with parameterized quantum gates," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Sep. 9, 2021, Art. no. 8020105.
- [10] A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, and S. Woerner, "The power of quantum neural networks," *Nature Comput. Sci.*, vol. 1, no. 6, pp. 403–409, Jun. 2021. [Online]. Available: <https://doi.org/10.1038/s43588-021-00084-1>
- [11] A. Abbas, D. Sutter, A. Figalli, and S. Woerner, "Effective dimension of machine learning models," 2021. Accessed: 2023. [Online]. Available: <https://arxiv.org/abs/2112.04807>
- [12] P. Rebentrost, M. Mohseni, and S. Lloyd, "Quantum support vector machine for big data classification," *Phys. Rev. Lett.*, vol. 113, Sep. 2014, Art. no. 130503. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.113.130503>
- [13] J. Preskill, "Quantum computing in the NISQ era and beyond," *Quantum*, vol. 2, Aug. 2018, Art. no. 79. [Online]. Available: <https://doi.org/10.22331/q-2018-08-06-79>
- [14] M. Schuld and N. Killoran, "Quantum machine learning in feature Hilbert spaces," *Phys. Rev. Lett.*, vol. 122, Feb. 2019, Art. no. 040504. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.122.040504>
- [15] P. Helber, B. Bischke, A. Dengel, and D. Borth, "EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 7, pp. 2217–2226, Jul. 2019.
- [16] A. Sebastianelli, D. A. Zaidenberg, D. Spiller, B. L. Saux, and S. L. Ullo, "On circuit-based hybrid quantum neural networks for remote sensing imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 565–580, Dec. 13, 2021.
- [17] S. Otgonbaatar and M. Datcu, "Natural embedding of the stokes parameters of polarimetric synthetic aperture radar images in a gate-based quantum computer," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Sep. 15, 2021, Art. no. 4704008.
- [18] A. Mari, T. R. Bromley, J. Izaac, M. Schuld, and N. Killoran, "Transfer learning in hybrid classical-quantum neural networks," *Quantum*, vol. 4, Oct. 2020, Art. no. 340. [Online]. Available: <https://doi.org/10.22331/q-2020-10-09-340>
- [19] A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, and J. I. Latorre, "Data re-uploading for a universal quantum classifier," *Quantum*, vol. 4, Feb. 2020, Art. no. 226. [Online]. Available: <https://doi.org/10.22331/q-2020-02-06-226>
- [20] S. Lloyd, M. Schuld, A. Ijaz, J. Izaac, and N. Killoran, "Quantum embeddings for machine learning," May 2020. Accessed: 2023. [Online]. Available: <https://arxiv.org/abs/2001.03622>
- [21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Representations*, 2014. [Online]. Available: <https://api.semanticscholar.org/CorpusID:14124313>
- [22] M. Schuld, A. Bocharov, K. M. Svore, and N. Wiebe, "Circuit-centric quantum classifiers," *Phys. Rev. A*, vol. 101, Mar. 2020, Art. no. 032308. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.101.032308>
- [23] T. Haug, K. Bharti, and M. Kim, "Capacity and quantum geometry of parametrized quantum circuits," *PRX Quantum*, vol. 2, Oct. 2021, Art. no. 040309. [Online]. Available: <https://link.aps.org/doi/10.1103/PRXQuantum.2.040309>
- [24] W. J. Maddox, G. Benton, and A. G. Wilson, "Rethinking parameter counting in deep models: Effective dimensionality revisited," 2020. Accessed: 2023. [Online]. Available: <https://arxiv.org/abs/2003.02139>
- [25] Y. Du, M.-H. Hsieh, T. Liu, and D. Tao, "Expressive power of parametrized quantum circuits," *Phys. Rev. Res.*, vol. 2, Jul. 2020, Art. no. 033125. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevResearch.2.033125>
- [26] 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 J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3735–3756, Jun. 29, 2020.
- [27] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.
- [28] V. Bergholm et al., "PennyLane: Automatic differentiation of hybrid quantum-classical computations," 2018. Accessed: 2023. [Online]. Available: <https://arxiv.org/abs/1811.04968>
- [29] J. J. Meyer, "Fisher information in noisy intermediate-scale quantum applications," *Quantum*, vol. 5, Sep. 2021, Art. no. 539. [Online]. Available: <https://doi.org/10.22331/q-2021-09-09-539>
- [30] A. Ly, M. Marsman, J. Verhagen, R. Grasman, and E.-J. Wagenmakers, "A tutorial on Fisher information," 2017. Accessed: 2023. [Online]. Available: <https://arxiv.org/abs/1705.01064>
- [31] S. Otgonbaatar and M. Datcu, "Assembly of a coreset of earth observation images on a small quantum computer," *Electron.*, vol. 10, no. 20, 2021, Art. no. 2482. [Online]. Available: <https://www.mdpi.com/2079-9292/10/20/2482>
- [32] H.-Y. Huang et al., "Power of data in quantum machine learning," *Nature Commun.*, vol. 12, no. 1, May 2021, Art. no. 2631. [Online]. Available: <https://doi.org/10.1038/s41467-021-22539-9>
- [33] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, "Barren plateaus in quantum neural network training landscapes," *Nature Commun.*, vol. 9, no. 1, Nov. 2018, Art. no. 4812. [Online]. Available: <https://doi.org/10.1038/s41467-018-07090-4>
- [34] S. Otgonbaatar and D. Kranzlmüller, "Exploiting the quantum advantage for satellite image processing: Quantum resource estimation," Sep. 2023, doi: [10.36227/techrxiv.24085122.v1](https://doi.org/10.36227/techrxiv.24085122.v1).
- [35] I. Glasser, R. Sweke, N. Pancotti, J. Eisert, and J. I. Cirac, "Expressive power of tensor-network factorizations for probabilistic modeling, with applications from hidden Markov models to quantum machine learning," Jul. 2019, *arXiv:1907.03741*.

Soronzonbold Otgonbaatar received the bachelor's degree in theoretical physics from the National University of Mongolia, Ulan Bator, Mongolia, and the master's degree in theoretical physics from the University of Siegen, Siegen, Germany. He is currently working toward the doctoral degree with the German Aerospace Center, Oberpfaffenhofen, Germany, and Ludwig-Maximilians-Universität München, Munich, Germany.

His research interests include artificial intelligence, machine learning, quantum computing and algorithms, and quantum machine learning.

Gottfried Schwarz received the graduate degree in electrical engineering from the Technical University of Munich, Munich, Germany.

He has been involved in a number of National and International space projects with the German Aerospace Center, Oberpfaffenhofen, Germany, among them were deep-space missions as well as Earth observation missions.

Mihai Datcu (Fellow, IEEE) received the M.S. and Ph.D. degrees in electronics and telecommunications from the University Politehnica of Bucharest (UPB), Bucharest, Romania, respectively, and the Habilitation from the University Louis Pasteur, Strasbourg, France.

He is currently a Full Professor with the Faculty of Electronics, UPB, Bucharest.

Dieter Kranzlmüller received the doctoral degree and Habilitation from the Johannes Kepler Universität, Linz, Austria.

He is currently the Chairman of the Board of Directors, Leibniz Supercomputing Centre with the Bavarian Academy of Sciences and Humanities, Garching, Germany, and a Full Professor for communication systems and system programming with the Institute of Informatics, Ludwig-Maximilians-Universität München, Munich, Germany.