SEQUENT: Towards Traceable Quantum Machine Learning Using Sequential Quantum Enhanced Training

Philipp Altmann[1], Leo Sünkel[1], Jonas Stein[1], Tobias Müller[2],
Christoph Roch[1] and Claudia Linnhoff-Popien[1]
[1]*LMU Munich, Germany*
[2]*SAP SE, Walldorf, Germany*

Keywords: Quantum Machine Learning, Transfer Learning, Supervised Learning, Hybrid Quantum Computing.

Abstract: Applying new computing paradigms like quantum computing to the field of machine learning has recently gained attention. However, as high-dimensional real-world applications are not yet feasible to be solved using purely quantum hardware, hybrid methods using both classical and quantum machine learning paradigms have been proposed. For instance, transfer learning methods have been shown to be successfully applicable to hybrid image classification tasks. Nevertheless, beneficial circuit architectures still need to be explored. Therefore, tracing the impact of the chosen circuit architecture and parameterization is crucial for the development of beneficially applicable hybrid methods. However, current methods include processes where both parts are trained concurrently, therefore not allowing for a strict separability of classical and quantum impact. Thus, those architectures might produce models that yield a superior prediction accuracy whilst employing the least possible quantum impact. To tackle this issue, we propose *Sequential Quantum Enhanced Training* (SEQUENT) an improved architecture and training process for the traceable application of quantum computing methods to hybrid machine learning. Furthermore, we provide formal evidence for the disadvantage of current methods and preliminary experimental results as a proof-of-concept for the applicability of SEQUENT.

1 INTRODUCTION

With classical computation evolving towards performance saturation, new computing paradigms like quantum computing arise, promising superior performance in complex problem domains. However, current architectures merely reach numbers of 100 quantum bits (qubits), prone to noise, and classical computers run out of resources simulating similar sized systems (Preskill, 2018). Thus, most real-world applications are not yet feasible solely relying on quantum compute. Especially in the field of machine learning, where parameter spaces sized upwards of 50 million are required for tasks like image classification, the resources of current quantum hardware or simulators are not yet sufficient for pure quantum approaches (He et al., 2016). Therefore, hybrid approaches have been proposed, where the power of both classical and quantum computation are united for improved results (Bergholm et al., 2018). By this, it is possible to leverage the advantages of quantum computing for tasks with parameter spaces that cannot be computed solely by quantum computers due to hardware and simulation limitations. Within those hybrid algorithms, the quantum part is, analogue to the classical *deep neural networks* (DNNs), represented by so-called *variational quantum circuits* (VQCs), which are parameterized and can be trained in a supervised manner using labeled data (Cerezo et al., 2021). For hybrid machine learning, we will from hereon refer to VQCs as quantum parts and to DNNs as classical parts.

To solve large-scale real-world tasks, like image classification, the concept of *transfer learning* has been applied for training such hybrid models (Girshick et al., 2014; Pan and Yang, 2010). Given a complex model, with high-dimensional input- and parameter spaces, the term transfer leaning classically refers to the two-step procedures of first pre-training using a large but generic dataset and secondly fine-tuning using a smaller but more specific dataset (Torrey and Shavlik, 2010). Usually, a subset of the model's weights are frozen for the fine-tuning to compensate for insufficient amounts of fine-tuning data.

Applied to hybrid *quantum machine learning* (QML), the pre-trained model is used as a feature extractor and the dense classifier is replaced by a hybrid

Altmann, P., Sünkel, L., Stein, J., Müller, T., Roch, C. and Linnhoff-Popien, C.
SEQUENT: Towards Traceable Quantum Machine Learning Using Sequential Quantum Enhanced Training.
DOI: 10.5220/0011772400003393
In *Proceedings of the 15th International Conference on Agents and Artificial Intelligence (ICAART 2023) - Volume 3*, pages 744-751
ISBN: 978-989-758-623-1; ISSN: 2184-433X

model referred to as *dressed quantum circuit* (DQC) including classical pre- and post-processing layers, and the central VQC (Mari et al., 2020). This architecture results in concurrent updates to both classical and quantum weights. Even though, this produces updates towards overall optimal classification results, it does not allow for tracing the advantageousness of the quantum part of the architecture. Thus, besides providing competitive classification results, such hybrid approaches do not allow for valid judgment whether the chosen quantum circuit benefits the classification. The only arguable result is that it does not harm the overall performance, or that the introduced inaccuracies may be compensated by the classical layers in the end. However, as we are currently still only exploring VQCs, this verdict, i.e., traceability of the impact of both the quantum and the classical part, is crucial to infer the architecture quality from common metrics. Overall, with current approaches, we find a mismatch between the goal of exploring viable architectures and the process applied.

We therefore propose the application of *Sequential Quantum Enhanced Training* (SEQUENT), an adapted architecture and training procedure for hybrid quantum transfer learning, where the effect of both classical and quantum parts are separably assessable. Preliminary concepts regarding quantum computing, quantum machine learning, deep learning, and transfer learning are addressed in Section 2. Related work on quantum transfer learning and dressed quantum circuits is presented in Section 3. Overall, we provide the following contributions:

- We provide formal evidence that current quantum transfer learning architectures might result in an optimal network configuration (perfect classification / regression results) with the least-most quantum impact, i.e., a solution equivalent to a purely classical one in Section 4.
- We propose SEQUENT, a two-step procedure of classical pre-training and quantum fine-tuning using an adapted architecture to reduce the number of features classically extracted to the number of features manageable by the VQC producing the final classification (see Section 5).
- We show competitive results with a traceable impact of the chosen VQC on the overall performance using preliminary benchmark datasets in Section 6 and discuss implications and limitations of our work in Section 7.

2 BACKGROUND

To delimit SEQUENT, the following section provides a brief general introduction to the related fields of quantum computation, quantum machine learning, deep learning, and transfer learning.

2.1 Quantum Computing

Quantum Computation: works fundamentally different from classical computation, since QC uses qubits instead of classical bits. Where a classical bit can be in the state 0 or 1, the corresponding state of a qubit is described in Dirac notation as $|0\rangle$ and $|1\rangle$. However, more importantly, qubits can be in a superposition, i.e., a linear combination of both:

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle \tag{1}$$

To alter this state, a set of reversible unitary operations like rotations can be applied sequentially to individual *target* qubits or in conjunction with a *control* qubit. Upon measurement, the superposition collapses and the qubit takes on either the state $|0\rangle$ or $|1\rangle$ according to a probability. Note that α and β in (1) are complex numbers where $|\alpha|^2$ and $|\beta|^2$ give the probability of measuring the qubit in state $|0\rangle$ or $|1\rangle$ respectively. Note that $|\alpha|^2 + |\beta|^2 = 1$, i.e., the probabilities sum up to 1. (Nielsen and Chuang, 2010)

Quantum algorithms like Grover (Grover, 1996) or Shor (Shor, 1994) provide a theoretical speedup compared to classical algorithms. Moreover, in 2019 quantum supremacy was claimed (Arute et al., 2019), and the race to find more algorithms providing a quantum advantage is currently underway. However, the current state of quantum computing is often referred to as the *noisy-intermediate-scale quantum* (NISQ) era (Preskill, 2018), a period when relatively small and noisy quantum computers are available, however, still no error-correction to mitigate them, limiting the execution to small quantum circuits. Furthermore, current quantum computers are not yet capable to execute algorithms that provide any quantum advantage in a practically useful setting.

Thus, much research has recently been put into the investigation of hybrid-classical-quantum algorithms. That is, algorithms that consist of quantum and classical parts, each responsible for a distinct task. In this regard, quantum machine learning has been gaining in popularity.

Quantum Machine Learning: algorithms have been proposed in several varieties over the last years (Farhi et al., 2014; Dong et al., 2008; Biamonte et al., 2017).

Besides quantum kernel methods (Schuld and Killoran, 2019), *variational quantum algorithms* (VQAs) seem to be the most relevant in the current NISQ-era for various reasons (Cerezo et al., 2021). VQAs generally are comprised of multiple components, but the central part is the structure of the applied circuit or *Ansatz*. Furthermore, a VQA Ansatz is intrinsically parameterized to use it as a predictive model by optimizing the parameterization towards a given objective, i.e., to minimize a given loss. Overall, given a set of data and targets, a parameterized circuit and an objective, an approximation of the generator underlying the data can be learned. Applying methods like gradient descent, this model can be trained to predict the label of unseen data (Cerezo et al., 2021; Mitarai et al., 2018). For the field of QML, various circuit architectures have been proposed (Biamonte et al., 2017; Khairy et al., 2020; Schuld et al., 2020).

For the remainder of this paper, we consider the following *simple* ϕ-parameterized variational quantum circuit (VQC) for η qubits visualized in the inner part of Figure 1 and Figure 2:

$$\begin{aligned}\mathtt{VQC}_{\phi}(z) = \mathtt{meassure}_{\sigma} \circ \mathtt{entangle}_{\phi_{\delta}} \circ \cdots \circ \\ \circ\, \mathtt{entangle}_{\phi_1} \circ \mathtt{embed}_{\eta}(z) \qquad (2)\end{aligned}$$

with the depth δ, and the output dimension σ given the input $z = (z_1, \ldots, z_\eta)$, where $\mathtt{embed}_\eta$ loads the data-points z into η balanced qubits in superposition via z-rotations, $\mathtt{entangle}_\phi$ applies controlled not gates to entangle neighboring qubits followed by ϕ-parameterized z-rotations, and $\mathtt{measure}_\sigma$ applies the Pauli-Z operator and measures the first σ qubits (Schuld and Killoran, 2019; Mitarai et al., 2018).

This architecture has also been shown to be directly applicable to classification tasks, using the measurement expectation value as a one-hot encoded prediction of the target (Schuld et al., 2020).

Overall, VQAs have been shown to be applicable to a wide variety of classification tasks (Abohashima et al., 2020) and successfully utilized by Mari et al. (2020), using the *simple* architecture defined in (2). Thus, to provide a proof-of-concept for SEQUENT, we will focus on said architecture for classification tasks and leave the optimization of embeddings (LaRose and Coyle, 2020) and architectures (Khairy et al., 2020) to future research.

2.2 Deep Learning

Deep Neural Networks (DNNs): refer to parameterized networks consisting of a set of fully connected layers.

A layer comprises a set of distinct neurons, whereas each neuron takes a vector of inputs $x = (x_1, x_2, \ldots, x_n)$, which is multiplied with the corresponding weight vector $w_j = (w_{j1}, w_{j2}, \ldots, w_{jn})$. A bias b_j is added before being passed into an activation function φ. Therefore, the output of neuron z at position j takes the following form (Bishop and Nasrabadi, 2006):

$$z_j = \varphi\left(\sum_{i=1}^{n} w_{ji} x_i + b_j\right) \qquad (3)$$

Given a target function $f(x) : \mathbb{X} \mapsto y$, we can define the approximate

$$\hat{f}_\theta(x) : \mathbb{X} \mapsto \hat{y} = L_{h_d \to o} \circ \cdots \circ L_{n \to h_1} \qquad (4)$$

as a composition of multiple layers L with multiple neurons z parameterized by θ, $d-1$ h-dimensional hidden layers, and the respective input and target dimensions n and o. Using the prediction error $J = (y - \hat{f}_\theta(x))^2$, $\hat{f}_\theta$ can be optimized by propagating the error backwards through the network using the gradient $\nabla_\theta J$ (Bishop and Nasrabadi, 2006).

Those feed forward models have been shown capable of approximating arbitrary functions, given a sufficient amount of data and either a sufficient depth (i.e., number of hidden layers) or width (i.e., size of hidden state) (Leshno et al., 1993).

Deep neural networks for image classification tasks are comprised of two parts: A feature extractor containing a composite of convolutional layers to extract a υ-sized vector of features $\mathtt{FE} : \mathbb{X} \mapsto \upsilon$, and a composite of fully connected layers to classify the extracted feature vector $\mathtt{FC} : \upsilon \mapsto \hat{y}$. Thus, the overall model is defined as $\hat{f} : \mathbb{X} \mapsto \hat{y} = \mathtt{FC}_\theta \circ \mathtt{FE}_\theta(x)$. Those models have been successfully applied to a wide variety of real-world classification tasks (He et al., 2016; Krizhevsky et al., 2012). However, to find a parameterization that optimally separates the given dataset, a large amount of training data is required.

Transfer Learning: aims to solve the problem of insufficient training data by transferring already learned knowledge (weights, biases) from a task T_s of a source domain D_s to a related target task T_t of a target domain D_t. More specifically, a domain $D = \mathbb{X}, P(x)$ comprises a feature space $\mathbb{X}$ and the probability distribution $P(x)$ where $x = (x_1, x_2, \ldots, x_n) \in \mathbb{X}$. The corresponding task T is given by $T = \{y, f(x)\}$ with label space y and target function $f(x)$ (Zhuang et al., 2021). A deep transfer learning task is defined by $\langle D_s, T_s, D_t, T_t, \hat{f}_t(\cdot)\rangle$, where $\hat{f}_t(\cdot)$ is defined according to Equation (4) (Tan et al., 2018). Generally, transfer learning is a two-stage process. Initially, a source model is trained according to a specific task

T_s in the source domain D_s. Consequently, transfer learning aims to enhance the performance of the target predictive function $\hat{f}_t(\cdot)$ for the target learning task T_t in target domain D_t by transferring latent knowledge from T_s in D_s, where $D_s \neq D_t$ and/or $T_s \neq T_t$. Usually, the size of $D_s >> D_t$ (Tan et al., 2018). The knowledge transfer and learning step is commonly achieved via feature extraction and/or fine-tuning.

The **feature extraction** process freezes the source model and adds a new classifier to the output of the pre-trained model. Thereby, the feature maps learned from T_s in D_s can be repurposed, and the newly added classifier is trained according to the target task T_t (Donahue et al., 2014). The **fine-tuning** process additionally unfreezes top layers from the source model and jointly trains the unfreezed feature representations from the source model with the added classifier. By this, the time and space complexity for the target task T_t can be reduced by transferring and/or fine-tuning the already learned features of a pre-trained source model to a target model (Girshick et al., 2014).

3 RELATED WORK

In the context of machine learning, VQAs are often applied to the problem of classification (Schuld et al., 2020; Mitarai et al., 2018; Havlíček et al., 2019; Schuld and Killoran, 2019), although other application areas exist. Different techniques, such as embedding (Lloyd et al., 2020; LaRose and Coyle, 2020), or problems like barren plateaus (McClean et al., 2018), have been widely discussed in the QML literature. However, we focus on hybrid quantum transfer learning (Mari et al., 2020) in this paper.

Classical Transfer Learning is widely applied in present-day machine learning algorithms (Torrey and Shavlik, 2010; Pan and Yang, 2010; Pratt, 1992) and can be extended with concepts of the emerging quantum computing technology (Zen et al., 2020). Mari et al. (2020) propose various hybrid transfer learning architectures ranging from classical to quantum (CQ), quantum to classical (QC) and quantum to quantum (QQ). The authors focus on the former CQ architecture, which comprises the previously explained DQC. In the current era of intermediate-scale quantum technology, the DQC transfer learning approach is the most widely investigated and applied one, as it allows to some extent optimally pre-process high-dimensional data and afterward load the most relevant features into a quantum computer. Gokhale et al. (2020) used this architecture to classify and detect image splicing forgeries, while Acar and Yilmaz (2021) applied it to detect COVID-19 from CT images. Also, Mari et al. (2020) assess their approach exemplary on image classification tasks. Although the results are quite promising, it is not clear from the evaluation, whether the dressed quantum circuit is advantageous over a fully classical approach.

4 DQC QUANTUM IMPACT

We argue that within certain problem instances, DQCs may yield accurate results while not making active use of any quantum effects in the VQC. This possibility exists especially for easy to solve problem instances, when all purely classical layers are sufficient to yield accurate results and the quantum layer represents the identity. This can be seen by realizing that the classical pre-processing layer acts as a hidden layer with a non-polynomial activation function, hence being capable of approximating arbitrary continuous functions depending on the number of hidden units by the universal approximation theorem (Leshno et al., 1993). Therefore, the overall DQC architecture is portrayed in Figure 1.

The central VQC is defined according to Section 2.1 as introduced above. Both pre- and post-processing layers are implemented by fully connected layers of neurons with a non-linear activation function according to Subsection 2.2. Formally, the DQC for η qubits can thus be depicted as:

$$\mathtt{DQC} = L_{\eta \to \sigma} \circ \mathtt{VQC}_\phi \circ L_{n \to \eta} \quad (5)$$

where $L_{n \to \eta}$ and $L_{\eta \to \sigma}$ are the fully connected classical *dressing* layers according to Equation (3), mapping from the input size n to the number of qubits η and from the number of qubits η to the target size σ respectively, and $\mathtt{VQC}_\phi$ is the actual variational quantum circuit according to Equation (2) with η qubits and $\sigma = \eta$ measured outputs.

Now let us consider a parameterization ϕ, where $\mathtt{VQC}_\phi(z) = id(z) = z$ resembles the identity function. Consequently, Equation (5) collapses into the following purely classical, 2-layer feed-forward network with the hidden dimension η:

$$\mathtt{DQC} = L_{\eta \to \sigma} \circ id \circ L_{n \to \eta} = L_{\eta \to \sigma} \circ L_{n \to \eta} \quad (6)$$

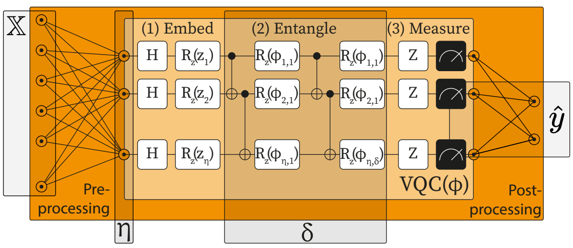

Figure 1: Dressed Quantum Circuit Architecture.

By the universal function approximation theorem, this allows DQC to approximate any polynomial function $f : \mathbb{R}^n \to \mathbb{R}^o$ of degree 1 arbitrarily well, even if the VQC is not affecting the prediction at all.

Consequently, one has to be careful in the selection of suitable problem instances, as they must not be too easy to ensure that the VQC is even needed to yield the desired results. This becomes especially difficult as current quantum hardware is quite limited, typically restricting the choice to fairly easy problem instances. On top of this, no necessity to use a post-processing layer seems apparent, as it has been shown in various publications (Schuld et al., 2020; Schuld and Killoran, 2019) that variational quantum classifiers, i.e, VQCs can successfully complete classification tasks without any post-processing.

Overall, whilst conveying a proof-of-concept, that the combination of classical neural networks and variational quantum circuits in the dressed quantum circuit hybrid architecture is able to produce competitive results, this architecture is neither able to convey the advantageousness of the chosen quantum circuit nor exclude the possibility of the classical part just being able to compensate for quantum in-steadiness.

5 SEQUENT

To improve the traceability of quantum impact in hybrid architectures, we propose *Sequential Quantum Enhanced Training*. SEQUENT improves upon the dressed quantum circuit architecture by introducing two adaptations to it: **First**, we omit the classical post-processing layer and use the variational quantum circuit output directly as the classification result. Therefore, we reduce the measured outputs σ from the number of qubits η (cf. Figure 1) to the dimension of the target $\hat{y}$ (cf. Figure 2).

The direct use of VQCs as a classifier has been frequently proposed and shown equally applicable as classical counterparts (Schuld et al., 2020). By this, the overall quality of the chosen circuit and parameterization are directly assessable by the classification result, thus the final accuracy. Moreover, a parameter setting of universal approximation capabilities (cf. Equation (6)) with the least (identical) quantum contribution is mathematically precluded by the removal of the hidden state (compare Equation (5)).

Concurrently omitting the pre-processing or compression layer, however, would increase the number of at least required qubits to the number of output features of the problem domain, or, when applied to image classification, the chosen feature extractor (e.g., 512 for Resnet-18).

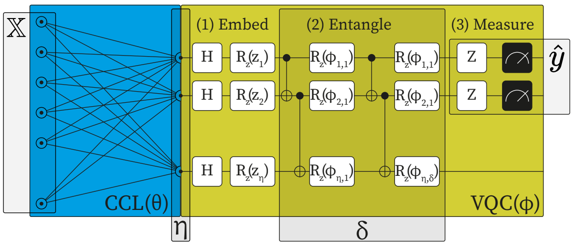

Figure 2: **SEQUENT Architecture**: Sequential Quantum Enhanced Training comprised of a classical compression layer (CCL) parameterized by θ and a variational quantum circuit (VQC) parameterized by ϕ with separate phases for classical (blue) and quantum (green) training for variable sets of input data $\mathbb{X}$, prediction targets $\hat{y}$ and VQCs with η qubits and δ entangling layers.

However, both current quantum hardware and simulators do not allow for arbitrate sized circuits, especially maxing out at around 100 qubits. We therefore **secondly** propose to maintain the classical compression layer to provide a mapping/compression $\mathbb{X} \mapsto \eta$ and, in order to fully classically pre-train the compression layer, add a surrogate classical classification layer $\eta \mapsto \hat{y}$.

Replacing this surrogate classical classification layer with the chosen variational quantum circuit to be assessed and freezing the pre-trained weights of the classical compression layer then allows for a second, purely quantum training phase and yields the following formal definition of the SEQUENT architecture displayed in Figure 2:

$$\mathtt{SEQUENT}_{\theta,\phi} : \mathbb{X} \mapsto \eta \mapsto \hat{y} = \mathtt{VQC}_{\phi}(z) \circ \mathtt{CCL}_{\theta}(x) \quad (7)$$

$$\mathtt{CCL}_{\theta}(x) : \mathbb{X} \mapsto \eta = L_{n \to \eta} \quad \text{(cf. Equation (3))}$$

$$\mathtt{VQC}_{\phi}(z) : \eta \mapsto \hat{y} \quad \text{(cf. Equation (2))}$$

Furthermore, we propose training SEQUENT via the following sequential training procedure depicted in Figure 3:

1. Pre-train SEQUENT: $\hat{f} : \mathbb{X} \mapsto \eta \mapsto \hat{y} = \mathtt{CCL}_{\theta}(x) \circ \mathtt{CCL}_{\theta}(z)$ containing a classical compression layer and a surrogate classification layer by optimizing the classical weights θ
2. Freeze the classical weights θ, replace the surrogate classical classification layer by the variational quantum classification circuit $\mathtt{VQC}_{\phi}$ (Equation (2)), and optimize the quantum weights ϕ.

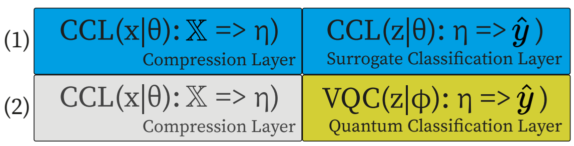

Figure 3: **SEQUENT Training Process** consisting of a classical (blue) pre-training phase (1) and a quantum (green) fine-tuning phase (2).

This two-step procedure can be seen as an application of transfer learning on its own, transferring from classical to quantum weights in a hybrid architecture. To be used for the classification of high-dimensional data, like images, the input x needs to be replaced by the intermediate output of an image recognition model z (cf. Subsection 2.2). Combining both two-step transfer learning procedures, the following three-step procedure is yielded:

1. Classically pre-train a full classification model (e.g., Resnet (He et al., 2016)) $\hat{f}: \mathbb{X} \mapsto \upsilon \mapsto \hat{y} = \texttt{FC}_{\theta}(z) \circ \texttt{FE}_{\theta}(x)$ to a large generic dataset (compare Subsection 2.2)
2. Freeze convolutional feature extraction layers $\texttt{FE}$ and fine-tune fully connected layers consisting of a compression layer and a surrogate classification layer $\texttt{FE}: \upsilon \mapsto \eta \mapsto \hat{y} = \texttt{CCL}_{\theta}(z) \circ \texttt{CCL}_{\theta}(x)$.
3. Freeze classical weights and replace surrogate classification layer with VQC to train the quantum weights ϕ of the hybrid model: $\hat{f}_{\theta,\phi}: \mathbb{X} \mapsto \upsilon \mapsto \eta \mapsto \hat{y} = \texttt{VQC}_{\phi}(z) \circ \texttt{CCL}_{\theta}(x) \circ \texttt{FE}$

For a classification task with n classes, at least $\eta \geq n$ qubits are required. Whilst we use the *simple* Ansatz introduced in Equation (2) with $\eta = 6$ qubits and a circuit depth of $\delta = 10$ to validate our approach in the following, any VQC architecture yielding a direct classification result would be conceivable.

6 EVALUATION

We evaluate SEQUENT by comparing its performance to its predecessor, the DQC, and a purely classical feed forward neural network. All models were trained on 2000 datapoints of the moons and spirals (Lang and Witbrock, 1988) benchmark dataset for two and four epochs of sequential, hybrid and classical training respectively. Both benchmarks consist of two-dimensional input points that are assigned either to the red or the blue class, where the separation of both distributions is highly non-linear. For efficient evaluation, all circuits were simulated using lightning qubits (Bergholm et al., 2018). All classical layers are built of fully connected neurons with tanh activations. To guarantee comparability, we set the size of the hidden state of the classical model to $h = \eta = 6$. The code for all experiments is available here[1].

The classification results are visualized in Figure 4. Looking at the result for the moons dataset, all compared models are able to depict the shape underlying data.

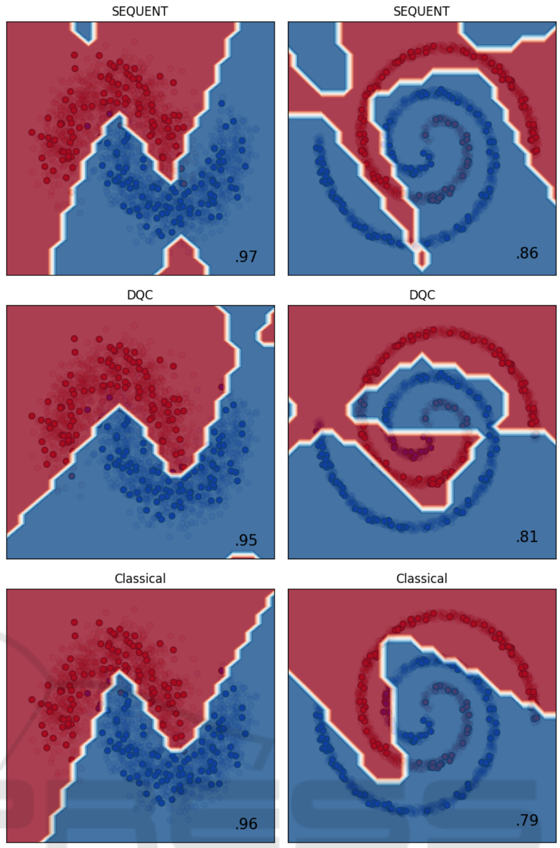

Figure 4: **Classification Results** of SEQUENT, DQC and Classical Feed Forward Neural Network for moons (left) and spirals (right) benchmark datasets

Note, that even the considerably simpler classical model is perfectly able to separate the given classes. Hence, these experimental results support the concerns about the impact of the VQC to the overall DQC's performance (cf. Section 4). With a final test accuracy of 95%, the DQC performs even worse than the purely classical model, reaching 96%. Looking at the SEQUENT results, however, these concerns are eliminated, as the performance and final accuracy of 97%, besides outperforming both compared models, can certainly be denoted to VQC, due to the applied training process and the used architecture.

Similar results show for the second benchmark dataset of intertwined spirals on the right side of Figure 4. The overall best accuracy of 86% however suggests, that further adjustments to the VQC could be beneficial. This result also depicts the application of SEQUENT we imagine for benchmarking and optimizing VQC architectures.

[1] https://github.com/philippaltmann/SEQUENT

7 CONCLUSIONS

We proposed *Sequential Quantum Enhanced Training* (SEQUENT), a two-step transfer learning procedure applied to training hybrid QML algorithms combined with an adapted hybrid architecture to allow for tracing both the classical and quantum impact on the overall performance. Furthermore, we showed the need for said adaptions by formally pointing out weaknesses of the DQC, the current state-of-the-art approach to this regard. Finally, we showed that SEQUENT yields competitive results for two representative benchmark datasets compared to DQCs and classical neural networks. Thus, we provided proof-of-concept for both the proposed reduced architecture and the adapted transfer learning training procedure.

However, whilst SEQUENT theoretically is applicable to any kind of VQC, we only considered the simple architecture with fixed angle embeddings and δ entangling layers as proposed by (Mari et al., 2020). Furthermore, we only supplied preliminary experimental implications and did not yet test any high dimensional real-world applications. Overall, we do not expect superior results that outperform state-of-the-art approaches in the first place, as viable circuit architectures for quantum machine learning are still an active and fast-moving field of research.

Thus, both the real-world applicability and the development of circuit architectures that indeed offer a benefit over classical ones should undergo further research attention. To empower real-world applications, the use of hybrid quantum methods should also be kept in mind when pre-training large classification models like Resnet. Also, applying more advanced techniques to train the pre-processing or compression layer to take full advantage of the chosen quantum circuit should be examined. Therefore, auto-encoder architectures might be applicable to train a more generalized mapping from the classical input-space to the quantum-space. Overall, we believe, that applying the proposed concepts and building upon SEQUENT, both valuable hybrid applications and beneficial quantum circuit architectures can be found.

ACKNOWLEDGEMENTS

This work is part of the Munich Quantum Valley, which is supported by the Bavarian state government with funds from the Hightech Agenda Bayern Plus and was partially funded by the German BMWK Project PlanQK (01MK20005I).

REFERENCES

Abohashima, Z., Elhosen, M., Houssein, E. H., and Mohamed, W. M. (2020). Classification with quantum machine learning: A survey. *arXiv preprint arXiv:2006.12270*.

Acar, E. and Yilmaz, I. (2021). Covid-19 detection on ibm quantum computer with classical-quantum transferlearning. *Turkish Journal of Electrical Engineering and Computer Sciences*, 29(1):46–61.

Arute, F., Arya, K., Babbush, R., Bacon, D., Bardin, J. C., Barends, R., Biswas, R., Boixo, S., Brandao, F. G., Buell, D. A., et al. (2019). Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510.

Bergholm, V., Izaac, J., Schuld, M., Gogolin, C., Alam, M. S., Ahmed, S., Arrazola, J. M., Blank, C., Delgado, A., Jahangiri, S., et al. (2018). Pennylane: Automatic differentiation of hybrid quantum-classical computations. *arXiv preprint arXiv:1811.04968*.

Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., and Lloyd, S. (2017). Quantum machine learning. *Nature*, 549(7671):195–202.

Bishop, C. M. and Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*, volume 4. Springer.

Cerezo, M., Arrasmith, A., Babbush, R., Benjamin, S. C., Endo, S., Fujii, K., McClean, J. R., Mitarai, K., Yuan, X., Cincio, L., et al. (2021). Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644.

Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., and Darrell, T. (2014). Decaf: A deep convolutional activation feature for generic visual recognition. In *International conference on machine learning*, pages 647–655. PMLR.

Dong, D., Chen, C., Li, H., and Tarn, T.-J. (2008). Quantum reinforcement learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 38(5):1207–1220.

Farhi, E., Goldstone, J., and Gutmann, S. (2014). A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028*.

Girshick, R. B., Donahue, J., Darrell, T., and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 580–587.

Gokhale, A., Pande, M. B., and Pramod, D. (2020). Implementation of a quantum transfer learning approach to image splicing detection. *International Journal of Quantum Information*, 18(05):2050024.

Grover, L. K. (1996). A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 212–219.

Havlíček, V., Córcoles, A. D., Temme, K., Harrow, A. W., Kandala, A., Chow, J. M., and Gambetta, J. M. (2019). Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Khairy, S., Shaydulin, R., Cincio, L., Alexeev, Y., and Balaprakash, P. (2020). Learning to optimize variational quantum circuits to solve combinatorial problems. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 2367–2375.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Pereira, F., Burges, C., Bottou, L., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc.

Lang, K. and Witbrock, M. (1988). Learning to tell two spirals apart proceedings of the 1988 connectionists models summer school.

LaRose, R. and Coyle, B. (2020). Robust data encodings for quantum classifiers. *Physical Review A*, 102(3):032420.

Leshno, M., Lin, V. Y., Pinkus, A., and Schocken, S. (1993). Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural networks*, 6(6):861–867.

Lloyd, S., Schuld, M., Ijaz, A., Izaac, J., and Killoran, N. (2020). Quantum embeddings for machine learning. *arXiv preprint arXiv:2001.03622*.

Mari, A., Bromley, T. R., Izaac, J., Schuld, M., and Killoran, N. (2020). Transfer learning in hybrid classical-quantum neural networks. *Quantum*, 4:340.

McClean, J. R., Boixo, S., Smelyanskiy, V. N., Babbush, R., and Neven, H. (2018). Barren plateaus in quantum neural network training landscapes. *Nature communications*, 9(1):1–6.

Mitarai, K., Negoro, M., Kitagawa, M., and Fujii, K. (2018). Quantum circuit learning. *Physical Review A*, 98(3):032309.

Nielsen, M. A. and Chuang, I. (2010). Quantum computation and quantum information.

Pan, S. J. and Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359.

Pratt, L. Y. (1992). Discriminability-based transfer between neural networks. In *Proceedings of the 5th International Conference on Neural Information Processing Systems*, NIPS'92, page 204–211, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Preskill, J. (2018). Quantum computing in the nisq era and beyond. *Quantum*, 2:79.

Schuld, M., Bocharov, A., Svore, K. M., and Wiebe, N. (2020). Circuit-centric quantum classifiers. *Phys. Rev. A*, 101:032308.

Schuld, M. and Killoran, N. (2019). Quantum machine learning in feature hilbert spaces. *Phys. Rev. Lett.*, 122:040504.

Shor, P. W. (1994). Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th annual symposium on foundations of computer science*, pages 124–134. Ieee.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., and Liu, C. (2018). A survey on deep transfer learning. In *International conference on artificial neural networks*, pages 270–279. Springer.

Torrey, L. and Shavlik, J. (2010). Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pages 242–264. IGI global.

Zen, R., My, L., Tan, R., Hébert, F., Gattobigio, M., Miniatura, C., Poletti, D., and Bressan, S. (2020). Transfer learning for scalability of neural-network quantum states. *Phys. Rev. E*, 101:053301.

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., and He, Q. (2021). A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76.