Quantum Measurement Learning for Medical Image Classification

UNIVERSIDAD NACIONAL DE COLOMBIA
Systems and Industrial Engineering Department

Dissertation document submitted for the degree of
Master's in Computing and Systems Engineering at La Universidad Nacional de Colombia

Presented by:
Diego Hernando Useche Reyes

Director:
Prof. Fabio Augusto González Osorio

Bogotá, Colombia
March 2022

Acknowledgments

In this master's thesis, I want to thank specially my family Hernando Useche, Judith Reyes, Lorena Useche and Indira Useche for always supporting my projects. Also, this project would not have been possible without the encouragement and support of Diana Contreras. I also want to express my gratitude to my advisor Prof. Fabio González for his constant help, guidance and teachings through this research project. Besides my advisor, I want to acknowledge all the help of the collaborators of the articles, Santiago Toledo, Andrés Giraldo, Hernán Zuluaga and José Jaramillo. I am grateful with my companions of the MindLab Research group Sneyder Gantiva, Diana Cabrera, Alejandro Gallego, and Oscar Bustos, for their assistance and uplifting words. I also want to thank Universidad Nacional de Colombia for the infrastructure, commodities, and for being a good learning and research environment.

Abstract

Quantum measurement learning for medical image classification

Deep neural networks are the state-of-the-art for medical image classification. However, these models require large data sets to be trained, and they lack some interpretability on their predictions. In recent years, there has been a growing interests of using the statistical machinery of quantum mechanics to built novel machine learning models, which may run on classical or quantum computers. One of such models is the recently proposed method quantum measurement classification (QMC) [1]. In this thesis, we present various classical-quantum machine learning strategies that combine convolutional neural networks (CNNs) with methods based on QMC [2] to the task of learning medical images in a supervised manner. We first approach the problem with a deep probabilistic regression model, showing that is competitive, and more interpretable compared to conventional deep learning architectures. We then present a representation learning technique based on CNNs which maps medical images to pure and mixed quantum states, and show that its competitive with other representation learning strategies. In addition, we propose a quantum implementation of two QMC-based models on a high-dimensional quantum computer, we demonstrate that it is possible to perform classification and density estimation in a quantum computer.

Keywords: quantum measurement classification, prostate cancer, deep learning, quantum machine learning.

Resumen

Aprendizaje con medición cuántica para la clasificación de imágenes médicas

Las redes neuronales profundas están a la vanguardia para la clasificación de imágenes médicas. Sin embargo, estos modelos requieren para su entrenamiento conjuntos de datos muy grandes, y a sus predicciones les falta interpretabilidad. Recientemente, se han propuesto varios métodos de inteligencia artificial basados en la mecánica cuántica, los cuales pueden ser implementados en computadores clásicos o cuánticos. Uno de estos métodos es el recientemente propuesto *Quantum Measurement Classification* (QMC) [1]. En este trabajo de tesis, presentamos diferentes estrategias clásicas y cuánticas de aprendizaje automático, las cuales combinan las redes neuronales convolucionales (CNNs) y algunos métodos basados en QMC [2] para la tarea de aprendizaje supervisado de imagenes medicas. En primer lugar, planteamos el problema de clasificación con un modelo de regresión profundo y probabilístico, mostrando que es competitivo y más interpretable en comparación a arquitecturas convencionales de aprendizaje profundo. En segundo lugar, presentamos un método de aprendizaje de la representación basado en CNNs del cual se obtienen características de las imágenes médicas en forma de estados cuánticos puros y mezclados, y mostramos que los resultados del método son competitivos con otras estrategias de representación. Adicionalmente, proponemos una implementación cuántica de dos métodos de aprendizaje automático basados en QMC en un computador cuántico de altas dimensiones, mostrando que es posible el aprendizaje supervisado y la estimación de la densidad en un computador cuántico.

Palabras claves: clasificación con medición cuántica, cáncer de próstata, aprendizaje profundo, aprendizaje automático cuántico.

This master's thesis was presented on the 17 of May of 2022 at 03:00 pm,
and it was evaluated by the following two juries:

César Pedraza Bonilla (Ph.D.)
Universidad Nacional de Colombia, Bogotá, Colombia

Ángel Cruz Roa (Ph.D.)
Universidad de Los Llanos, Villavicencio, Colombia

Contents

Chapter 1

Introduction

Deep learning [3, 4, 5] is the state-of-the-art on medical image classification [6, 7, 8]. Some disadvantages of current deep learning methods are the gigantic amount of data required to train the models, and a lack of interpretability of the predictions, in the sense that the models do not offer the degree of uncertainty of the predictions. This uncertainty level is crucial to develop machine learning methods for medical image classification, since doctors are more equipped to make a medical decision by knowing the probabilities and uncertainties of the medical predictions of the models.

In this thesis, we present various hybrid classical-quantum algorithms to classify medical images. The main reference of this work is the article *Learning with Density Matrices and Random Features* by González et al. [2]. That article introduces three machine learning methods which combine density matrices and random Fourier features (RFF) [9]: the density matrix kernel density estimation (DMKDE), the density matrix kernel density classification (DMKDC), and the quantum measurement regressor (QMR). The DMKDE method is a non-parametric method for density estimation, the DMKDC is a supervised classification method based on kernel density estimation with density matrices, and the QMR is a regression method, based on the quantum measurement classification method (QMC) [1].

This manuscript builds on top of the aforementioned algorithms, to develop classical-quantum machine learning models to classify medical images. In addition, we propose a quantum protocol that implements the DMKDE and DMKDC algorithms in a high-dimensional quantum computer. This work is mostly constructed as a compilation of

articles which addresses the following main question:

- How well would a hybrid machine learning algorithm perform in terms of level of prediction and computational complexity for the task of classifying medical images in comparison with classical machine learning models?

The specific questions intended to answer at this thesis are:

- How to apply a proper representation learning algorithm for feature extraction of medical images?
- How to apply the quantum measurement learning algorithm to medical images?
- Is it possible to improve the performance to learn medical images with quantum measurement learning in comparison with classical machine learning models?
- How to implement the quantum measurement classification algorithm on a high-dimensional quantum computer?

1.1 Objectives

Based on the aforementioned motivation and research questions, the general and specific objectives of this thesis document are presented below.

1.1.1 General objective

To build a hybrid supervised machine learning model, which combines deep learning methods and a quantum-inspired algorithm called quantum measurement learning, for the task of medical image classification.

1.1.2 Specific objectives

1. To propose or adapt a strategy for representation learning on medical images based on deep learning methods.
2. To develop a method to classify medical images which combines deep learning and quantum measurement learning.
3. To evaluate the method over one or more medical image benchmark data sets.

1.2 Main contributions

In this work, we propose various hybrid classical-quantum machine learning models that combine deep learning architectures with machine learning methods based on quantum measurement classification [1], and apply these methods to classify various data sets of medical images. The main contributions of this work are listed below.

1. We develop a machine learning architecture which combines deep learning with the quantum measurement regression method to classify prostate cancer images, obtaining comparable results with the state-of-the-art, see Chapter 2.
2. We develop various representation learning techniques based on deep learning and combine them with methods based on QMC to classify colon histopathology images, achieving state-of-the-art results, see Chapter. 3.
3. We propose a classical-quantum program for supervised classification and density estimation which implements the DMKDE and DMKDC in a high-dimensional quantum computer, see Chapter 4.

This research work resulted in the following preprint and journal article:

- S. Toledo-Cortés, D. H. Useche, and F. A. González, "Prostate Tissue Grading with Deep Quantum Measurement Ordinal Regression," 3 2021. (to be submitted)
- D. H. Useche, A. Giraldo-Carvajal, H. M. Zuluaga-Bucheli, J. A. Jaramillo-Villegas, and F. A. González, "Quantum measurement classification with qudits," Quantum Information Processing 2021 21:1, vol. 21, pp. 1–12, 12 2021.

1.3 Thesis outline

The document is composed of five chapters including the introduction. The chapters are described below.

In Chapter 2, we propose a supervised learning model that combines the QMR algorithm with deep learning architectures for medical image classification, the article's name is *Prostate Tissue Grading with Deep Quantum Measurement Ordinal Regression* [10]. In Sect. 2.3.2.2 we summarize the QMR algorithm. In Sect. 2.3, we present the proposed method which combines the QMR with deep learning. In Sect. 2.4.3, we present the results of the method in a data set of prostate histophatology images.

In Chapter 3, we present a method for representation learning based on convolutional neural networks (CNNs) and QMC. In Sect 3.2, we show how to represent medical images as quantum states and density matrices with two classical-quantum deep learning architectures. In Sect. 3.3, we present the results of the method and we compare them with conventional neural networks.

In Chapter 4, we present the article *Quantum measurement classification with qudits* [11]. That article proposes quantum circuits of the DMKDE and DMKDC for qudit-based quantum computers. In Sects. 4.2.1, 4.2.2 we present the theoretical background of the DMKDE and DMKDC methods. In Sect. 4.3 we show our proposed quantum algorithms in high-dimensional quantum computers, and in Sect. 4.3 we present the results of density estimation and supervised classification on various data sets, and we conclude in Sect. 4.5.

Finally, in Chapter 5, we present the conclusions of the thesis and some ideas for future work.

Chapter 2

Prostate Tissue Grading with Deep Quantum Measurement Ordinal Regression

Prostate cancer (PCa) is one of the most common and aggressive cancers worldwide. The Gleason score (GS) system is the standard way of classifying prostate cancer and the most reliable method to determine the severity and treatment to follow. The pathologist looks at the arrangement of cancer cells in the prostate and assigns a score on a scale that ranges from 6 to 10. Automatic analysis of prostate whole-slide images (WSIs) is usually addressed as a binary classification problem, which misses the finer distinction between stages given by the GS. This chapter presents a probabilistic deep learning ordinal classification method that can estimate the GS from a prostate WSI. Approaching the problem as an ordinal regression task using a differentiable probabilistic model not only improves the interpretability of the results, but also improves the accuracy of the model when compared to conventional deep classification and regression architectures. This chapter corresponds to the article *Prostate Tissue Grading with Deep Quantum Measurement Ordinal Regression* [10]. This work dated Feb. 2022 is in the peer review process in a journal and it is accessible through Arxiv. The reference of the article is presented below.

S. Toledo-Cortés, D. H. Useche, and F. A. González, "Prostate Tissue Grading with Deep Quantum Measurement Ordinal Regression," 3 2021. https://arxiv.org/abs/2103.03188v1

2.1 Introduction

Prostate cancer (PCa) is currently the second most common cancer among men in America. Early detection allows for greater treatment options and a greater chance of treatment success, but while there are several methods of initial screening, a concrete diagnosis of PCa can only be made with a prostate biopsy [12]. Tissue samples are currently recorded in high-resolution images, called whole-slide images (WSIs). In these images the pathologists analyze the alterations in the stroma and glandular units and, using the Gleason score (GS) system, classify prostate cancer into five progressive levels from 6 to 10 [13]. The higher the grade, the more advanced the cancer. The analysis is mostly a manual task and requires specialized urological pathologists. This specialized staff is not always available, especially in developing countries, and the process is subject to great inter-observer variability [14]. Therefore, several efforts have been made to develop computer assisted diagnosis systems which may facilitate the work of specialists [15].

Deep convolutional neural networks (CNN) represent the state of the art in the analysis of visual information, and their implementation in automatic classification models for medical images has been widely studied. However, there is still much research to be done in relation to the diagnostic process in histopathology [14]. One of the main problems facing the application of deep learning into medical problems is the limited availability of large databases, given the standard required for the successful training of deep learning models. For histopathology, the previous performed studies have been limited to very small data sets or subsets of Gleason patterns [14]. In addition, deep learning models approach the prostate tissue grading task as a multi-class or even a binary classification of low risk (6-7 GS) vs high risk (8-10 GS) cases [16]. This has two drawbacks: first, the ordinal information of the grades is not taken into account. Second, the model predictions, usually subject to a softmax activation function, cannot be interpreted as a probability distribution [17], and therefore do not give information about the uncertainty of the predictions which, in safety-critical applications, provides the method with a first level of interpretability.

In this chapter we approach the prostate tissue grading as an ordinal regression task. We present the deep quantum measurement ordinal regression (DQMOR), a deep probabilistic model that combines a CNN with a differentiable probabilistic regression model, the quantum measurement regression (QMR) [2]. This approach allows us to:

1. Predict posterior probability distributions over the grades range. Unlike other prob-

abilistic methods as Gaussian processes, these are explicit discrete distributions.

2. Integrate patch-level posterior distributions into a single whole-slide image distribution in a simple, yet powerful probability-based manner.

3. Quantify the uncertainty of the predictions. This enrich the model as a diagnostic support tool, by providing it with a first level of interaction and interpretability of the results.

In order to validate our approach, we compare our performance with state of the art deep learning-based methods [16], and with close related classification and regression methods as the density matrix kernel density classification (DMKDC) [2] and Gaussian processes [18] [19].

The chapter is organized as follows: Section 2.2 presents a brief overview of the related work. Section 2.3 presents the theoretical framework of the DQMOR, and Section 2.4 presents the experimental set up and results. Finally in Section 2.5 we present the conclusions of this work.

2.2 Related work

Classification of prostate cancer images by GS is considered a difficult task even among pathologist, who do not usually agree on their judgment. In recent years, there has been a great research effort to automatically classify PCa. However, most of the previous works focus on classifying prostate WSIs between low and high GS, ignoring the inherent ordinal characteristics of the grading system.

To train a CNN with WSIs, it is required to divide each image into multiple patches, and then, to summarize the information of the patches by different methods, hence, obtaining a prediction of the WSI. In [20], the authors classify patches between low, and high GS, utilizing various CNN architectures and summarizing the patches to a WSI by a GS majority vote. Another approach by Tolkach et al. [21] uses a NASNetLarge CNN, and summarizes the GS of the patches by counting the probabilities per class. In Karimi et al. [22] they proposed training three CNNs for patches of different sizes, and summarizing the probabilities by a logistic regression. In [23], the authors use Gaussian processes based on granulometry descriptors extracted with a CNN for the binary classification task. Some other CNN architectures for GS grading include a combination of an atrous spatial pyramid

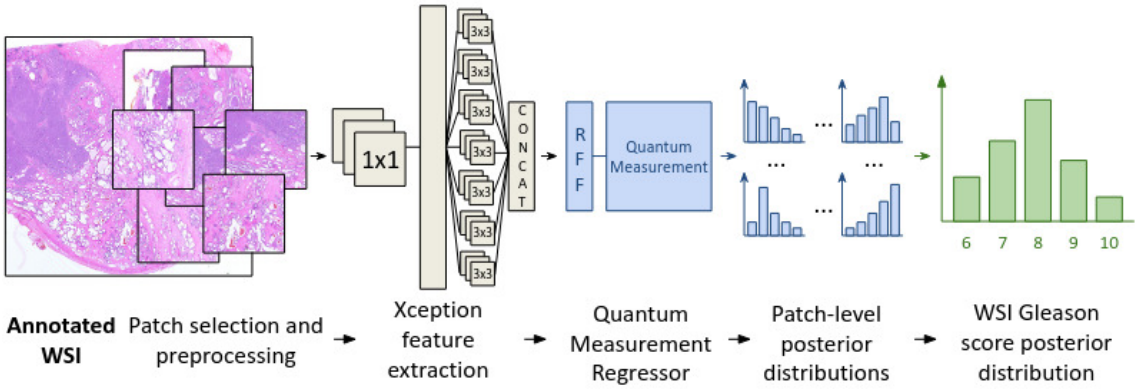


Figure 2.3.1: Overview of the proposed DQMOR method for prostate tissue grading. A Xception network is used as feature extractor for the image patches. Those features are the input for the QMR regressor model, which yields a posterior probability distribution by patch over the Gleason score grades. Then, those distributions are summarized into a single discrete probability distribution for the WSI.

pooling and a regular CNN as in [13], an Inception-v3 CNN with a support vector machine (SVM) as in [24], and a DeepLabV3+ with a MobileNet as the backbone [25].

2.3 Deep quantum measurement ordinal regression

The overall architecture of the proposed deep quantum measurement ordinal regression (DQMOR) is described in Figure 2.3.1. We use a Xception CNN [26] as a patch-level feature extractor. The extracted features are then used as inputs for the QMR method [2]. QMR requires an additional feature mapping from the inputs to get a quantum state-like representation. This is made by means of a random Fourier features approach [9]. The regressor yields a discrete posterior probability distribution at patch level. Then, to predict the GS of a single WSI, we summarize the results of the patches into a single posterior distribution from which we get the final grade and an uncertainty measure.

2.3.1 Feature extraction

We choose as feature extractor the model presented in [16], which is publicly available, and consists of an Xception network trained on ImageNet and fine-tuned on prostate tissue image patches. This network was originally used for an automatic information fusion model for the automatic binary (low-high) classification of WSIs. Taking the output of

the last average pooling layer of the model we got a 2048-dimensional vector representing each image patch.

2.3.2 Quantum measurement regression

QMR addresses the question on how to use density matrices for regression problems, using random features to encode the inputs in a quantum state-like representation. The model works as a non-parametric density estimator [2].

2.3.2.1 Random Fourier features

The RFF method [9] creates a feature map of the data $\mathbf{z}(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^{\mathbb{D}}$ in which the dot product of the samples in the $\mathbb{R}^D$ space approximates a shift invariant kernel $k(\mathbf{x}-\mathbf{y})$. The method works by sampling i.i.d. $w_1, \cdots, w_D \in \mathbb{R}^n$ from a probability distribution $p(w)$ given by the Fourier transform of $k(\mathbf{x}-\mathbf{y})$, and sampling i.i.d. $b_1, \cdots, b_D \in \mathbb{R}$ from an uniform distribution in $[0, 2\pi]$. In our context, the shift invariant kernel is the radial basis function (RBF) given by, $\mathrm{k_{RBF}}(\mathbf{x}-\mathbf{y}) = e^{-\gamma\|\mathbf{x}-\mathbf{y}\|^2}$, where gamma γ and the number D of RFF components are hyper-parameters of the models. In our model the RFF works as an embedding layer that maps the features from the Xception module to a representation space that is suitable for the quantum measurement regresion layer.

2.3.2.2 Quantum measurement regression

QMR [2] is a differentiable probabilistic regression model that uses a density matrix, ρ_{train}, to represent the joint probability distribution of inputs and labels. A QMR layer receives a RFF encoded input sample $|\psi_x\rangle$, and then builds a prediction operator $\pi = |\psi_x\rangle\langle\psi_x| \otimes \mathrm{Id}_{\mathcal{H}_\mathcal{Y}}$ where $\mathrm{Id}_{\mathcal{H}_\mathcal{Y}}$ is the identity operator in $\mathcal{H}_\mathcal{Y}$, the representation space of the labels. Inference is made by performing a measurement on the training density matrix ρ_{train}:

$$\rho = \frac{\pi\rho_{\text{train}}\pi}{\mathrm{Tr}[\pi\rho_{\text{train}}\pi]}. \tag{2.1}$$

Then a partial trace $\rho_\mathcal{Y} = \mathrm{Tr}_\mathcal{X}[\rho]$ is calculated, which encodes in $\rho_{\mathcal{Y}rr}$, with $r \in \{0, \ldots, N-1\}$, the posterior probability over the labels. The expected value represents the final prediction $\hat{y} = \sum_{r=0}^{N-1} r\rho_{\mathcal{Y}rr}$.

A gradient-based optimization is allowed by a spectral decomposition of the density matrix,

$\rho_{\text{train}} = V^{\dagger}\Lambda V$, in which the number of eigen-components of the factorization is a hyper-parameter of the model. The model is trained by minimizing a mean-squared-error loss function with a variance term whose relative importance is controlled by hyper-parameter α:

$$L = \sum (y - \hat{y})^2 + \alpha \sum_r \rho_{\mathcal{Y}rr} (\hat{y} - r)^2. \tag{2.2}$$

2.3.3 WSIs predictions

Since the training of the model is performed at patch level, the evaluation can be done at such level and at the level of the WSI. To get a prediction for a whole image, we explored two approaches: a majority vote procedure (MV), and a probability vote procedure (PV). In the majority vote, the prediction for an image is decided according to the grade with the highest number of predictions among the patches of the image. And in the probability vote, since each patch can be associated with a probability distribution, the normalized summation yields a distribution for the whole image. More formally, thanks to the law of total probability, given an image I, composed by n patches, each patch denoted by p_i, the posterior probability of the grade r is,

$$P(r|I) = \frac{P(r, I)}{P(I)} = \frac{\sum_{i=1}^{n} P(r|p_i, I) P(p_i|I) P(I)}{P(I)} = \frac{1}{n} \sum_{i=1}^{n} P(r|p_i). \tag{2.3}$$

The final prediction value thus corresponds to the grade with highest probability. In the DQMOR method, one can predicts the expected value of the distribution, but instead, the predictions at patch level were deduced from the probability of each grade per patch $P(r|p_i)$, and at WSI level by MV and PV.

2.4 Experimental evaluation

2.4.1 Dataset

We use images from the TCGA-PRAD data set, which contains samples of prostate tissue with GS from 6 to 10. This data set is publicly available via The Cancer Genome Atlas (TCGA) [20]. In order to directly compare our results with our baseline [16] we use the

same subset and partition consisting of 141 cases for training, 48 for validation and 46 for testing.

2.4.2 Experimental set up

The feature extraction model is publicly available and the augmentation procedure and training details are described in [16]. For the QMR, hyper-parameter tuning of the model was performed by generating 25 different random configurations. As result, we created an embedding with 1024 RFF components, 32 eigenvalues and γ was set to 2^{-13}. For the loss function (See eq. (2.2)), α was set at 0.4, and we set a learning rate of 6×10^{-5}.

Two extensions of the feature extractor model were set up as baseline for this work. The first dense layer classifier (DLC-1) consisted on 1024 neurons with ReLU as the activation function and a dropout of 0.2, followed by 5 neurons with a soft-max activation function for the output, and the learning rate was set to 10^{-7}, as in the baseline [16]. The second classifier (DLC-2) had two dense layers of 100 and 50 neurons with ReLU activation functions and dropouts of 0.2, connected to 5 neurons with a softmax activation function, and the learning rate was set to 10^{-3}.

We also explored two closely related methods to QMR: density matrix kernel density classification (DMKDC) [2] and Gaussian processes. DMKDC is a differentiable classification method, which applies a RFF feature map to the input sample, and then computes the expected value of the input with a density matrix of each class, returning a posterior probability distribution, which can be optimized with a categorical cross entropy loss function. As with QMR, a hyper-parameter random search was performed. We created an embedding with 1024 RFF components, and 32 eigenvalues. γ was set up at 2^{-13}, and we set a learning rate of 5×10^{-3}. All the previous experiments were performed in Python using the publicly available Keras-based implementation presented in [2].

On the other hand, Gaussian processes (GP) [19] are another powerful Bayesian approach to regression problems. By means of a kernel covariance matrix, the GP calculates and updates iteratively the probability distribution of all the functions that fit the data, optimizing in the process the kernel parameters. In our case we set the kernel as the RBF. The prediction process consist in marginalizing the learned Gaussian distribution, whose mean would be the actual prediction value, and its standard deviation an uncertainty indicator. We performed experiments with GP using the Scikit-Learn implementation in Python. We also explored deep Gaussian processes (DGP), using the implementation proposed in

Table 2.4.1: Patch-level results of the two dense layers classifiers models DCL-1, DCL-2, Gaussian processes GP, DGP, and density matrix-based models DMKDC, DQMOR. Mean and standard deviation of accuracy, macro f1 score and mean absolute error (MAE) are reported over 10 trials.

Method	Accuracy	Macro F1	MAE
DLC-1 [16]	0.530 ± 0.001	**0.314**± 0.001	0.786 ± 0.002
DLC-2 [16]	0.542 ± 0.005	0.296 ± 0.007	0.780 ± 0.009
GP [19]	0.399 ± 0.000	0.255 ± 0.000	0.777 ± 0.000
DGP [18]	0.265 ± 0.001	0.169 ± 0.000	1.013 ± 0.000
DMKDC [2]	**0.546** ± 0.002	0.305 ± 0.006	0.775 ± 0.007
DQMOR	0.477 ± 0.006	0.293 ± 0.003	**0.732**± 0.005

[18], which also uses RFF to approximate the covariance function. For those experiments, another hyper-parameter random search was made, finally setting the number of RFF at 1024 and the learning rate at 2×10^{-12} in a single layer schema.

2.4.3 Results and discussion

2.4.3.1 Ordinal regression

To measure the performance of an ordinal regression method implies to take into account the severity of the misclassified samples. Therefore, in addition to accuracy (ACC) and macro f1 score, we also measured mean absolute error (MAE) on the test partition, at patch level and WSI level. WSI scores were summarized by a MV and PV. The prediction methods at WSI-level were also applied to the baseline models. In the dense layers classifiers from the softmax output, as in [21]. In the DMKDC, the prediction methods were easily applied because the model outputs a probability distribution. For GP and DGP only MV was calculated, since we have no access to an explicit discrete posterior distribution. The results are reported in Table 2.4.1 and Table 2.4.2.

In terms of accuracy at patch level, the DMKDC model obtained the highest results. The best accuracy at WSI level was reached with the DQMOR model with probability vote. The DQMOR also obtained the least mean absolute errors at patch and WSI levels, showing that the model take advantage of the probability distributions and the inherent ordinality of the GS grades.

Table 2.4.2: WSI-level results. For each model, two summarization procedures are applied, majority vote (MV) and probability vote (PV). Mean and standard deviation of accuracy, macro f1 score and mean absolute error (MAE) are reported over 10 trials.

Method	Accuracy	Macro F1	MAE
DLC-1 MV [16]	0.543 ± 0.000	0.292 ± 0.000	0.826 ± 0.000
DLC-2 MV [16]	**0.548**± 0.009	0.300 ± 0.016	0.822 ± 0.009
GP MV [19]	0.391 ± 0.000	0.233 ± 0.000	0.739 ± 0.000
DGP MV [18]	0.174 ± 0.000	0.059 ± 0.000	0.935 ± 0.000
DMKDC MV [2]	0.546 ± 0.002	0.296 ± 0.012	0.824 ± 0.006
DQMOR MV	0.513 ± 0.014	**0.306**± 0.010	**0.713**± 0.027
DLC-1 PV [16]	0.543 ± 0.000	0.292 ± 0.000	0.826 ± 0.000
DLC-2 PV [16]	0.550 ± 0.005	0.304 ± 0.018	0.820 ± 0.010
DMKDC PV [2]	0.546 ± 0.002	0.296 ± 0.012	0.824 ± 0.006
DQMOR PV	**0.567**± 0.021	**0.345**± 0.014	**0.730**± 0.024

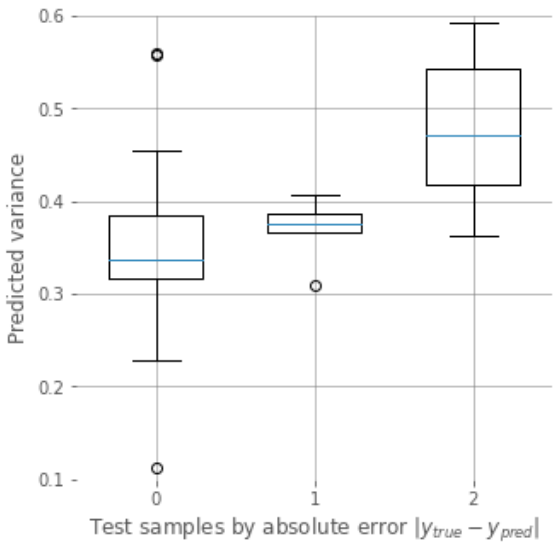


Figure 2.4.1: Boxplot of the predicted variance on test samples at WSI-level, grouped by absolute classification error $|y_{true} - y_{pred}|$.

2.4.3.2 Uncertainty quantification

Beyond the classification and regression performance of the methods, DQMOR allows an uncertainty quantification based on the variance of the predicted distribution. We analyzed the statistical behaviour of the predicted variance on the test set at WSI-level, grouping the samples according to the absolute error $|y_{true} - y_{pred}|$. As expected, DQMOR predicts low uncertainty levels on well classified samples when compared with the miss-classified samples (see Figure 2.4.1). In fact, the greater the absolute error, the greater the uncertainty. This attribute provides the method with an interpretable mean for the specialist, who may decide whether to trust or not in the model prediction.

2.5 Conclusions

In this work we approached the prostate tissue grading as an ordinal regression task. We combined the representational power of deep learning with the Quantum Measurement Regression method, which uses density matrices and random features to build a non-parametric density estimator.

The results on classification and regression metrics show that at WSI-level, DQMOR outperforms similar probabilistic classification and regression methods, as well as extension of the deep base model used for feature extraction. Regarding the analysis of the predicted uncertainty, we showed that DQMOR allows the identification of misclassified examples, and that the higher the misclassification error, the higher the uncertainty. This is a highly valued ability in medical applications, where the aim is to prevent false positives and especially false negatives in a diagnostic processes.

Overall we demonstrate that unlike single deep learning architectures and standard classification models, the combination of CNNs and QMR allows us to use the ordinal information of the disease grades, and provides a better theoretical framework to combine the patch-level inference into a single WSI prediction.

Chapter 3

Combining ConvNets and Quantum Measurement for Medical Image Classification

In this chapter, we present two novel classical-quantum deep learning architectures for representation learning of medical images which fuse convolutional neural networks (CNNs) with methods based on quantum measurement classification (QMC). We demonstrate that we can use CNNs to extract features in the form of pure quantum states and mixed density matrices, and classify those features with QMC-based methods. Results show that the proposed models are competitive with conventional deep neural networks.

3.1 Introduction

Learning the optimal features to represent visual data is an active area of research in machine learning and computer vision. This area is called representation learning, and as many other areas of machine learning it can be divided into four main learning paradigms: Supervised, unsupervised, semi-supervised and self-supervised learning. In supervised learning, labels are available for training, in contrast to unsupervised learning with no access to labels. Semi-supervised and self-supervised learning are intermediate learning techniques between the last two. In semi-supervised learning, a subset of labels are available, and in self-supervised learning the models learn from a subset of pseudo labels obtained from some additional metadata.

One of the main applications of representation learning is classification and segmentation of medical images. In [27], they use convolutional neural networks (CNNs) to extract features of mammography images, and classify them in a supervised learning manner with support vector machines (SVMs). Likewise, in [28] they use transfer learning with CNNs from ImageNet to classify otitis media images with SVMs. Autoencoders and SVMs have also been combined for feature extraction and classification of X-Ray images [29]. In the unsupervised scenario, CNNs have been used for clustering and segmentation of 3D medical images [30]. For semi-supervised and self-supervised learning, autoencoders have been applied for medical image segmentation [31], and for classification of sequential medical images [32]. CNNs have also been employed for self-supervised transfer learning for body part recognition [33]. Besides deep learning, other methods for representation learning of visual content include Graph Neural Networks [34, 35], Harr wavelet [36], and Histograms of Oriented Gradients (HOG) [37].

In addition to the precedent methods for representation learning, Gonzalez et al. [1] proposed a normalized feature representation based on random Fourier features (RFF) to feed a machine learning model based on quantum measurements. The quantum measurement classification proposal was further explored to develop a method called the density matrix kernel density classification (DMKDC) which combines RFF with density matrices for supervised classification [2] and a regression method called the quantum measurement regressor (QMR), they show that a LeNet CNN can be combined with RFF and DMKDC for representation learning. In [10], they applied a deep learning architecture and the quantum feature based on RFF with QMR to prostate histopathology images.

In this thesis chapter, we present new techniques to combine the DMKDC and a new method called the QMSDecomp with convolutional neural networks. In particular, we show that we can extract quantum feature maps (normalized vectors) out of ConvNets, for representation learning without using RFF, and that we can use deep learning architectures to represent medical images in the form of density matrices. We apply this framework to the colorectal cancer data set PathMNIST [38].

The chapter is organized as follows: In Sect 3.2, we present the method for representation learning based on CNNs and algorithms based on QMC, in Sect 3.2.3, we describe the data set of medical images, and finally we present the results and conclusions in Sects. 3.3 and 3.4 respectively.

3.2 Method: convolutional neural networks with QMC

In this section, we present two new methods that combine convolutional neural networks with the density matrix kernel density classification and a new machine learning method called the quantum measurement Schmidth decomposition (QMSDecomp). In particular, we show that we can combine classical deep learning methods with methods based on quantum measurement classification without random Fourier features for representation learning. Furthermore, we show that ConvNets can be used as a feature embedding from images to density matrices.

3.2.1 DMKDC with ConvNets

In [10], they combined deep convolutional neural networks with the quantum measurement regressor method [2], by applying an intermediate layer of random Fourier features, used as a quantum feature map. In contrast, we propose the DMKDC with ConvNets method which removes the RFF embedding, extracts quantum feature maps (or probability distributions) out of the convolutional layers, and uses these features to perform supervised classication with the DMKDC model.

The structure of the DMKDC with ConvNets is shown in the figure 3.2.1. We augmented the images sizes from (28, 28, 3) to (56, 56, 3), used transfer learning from a ResNet50 and a MobileNetV2 pretrained with ImageNet, removed the dense layers, and substituted the ReLu activation with a softmax activation. Since the softmax activation is a probability distribution, we performed a 2D average pooling, and computed the square root of the

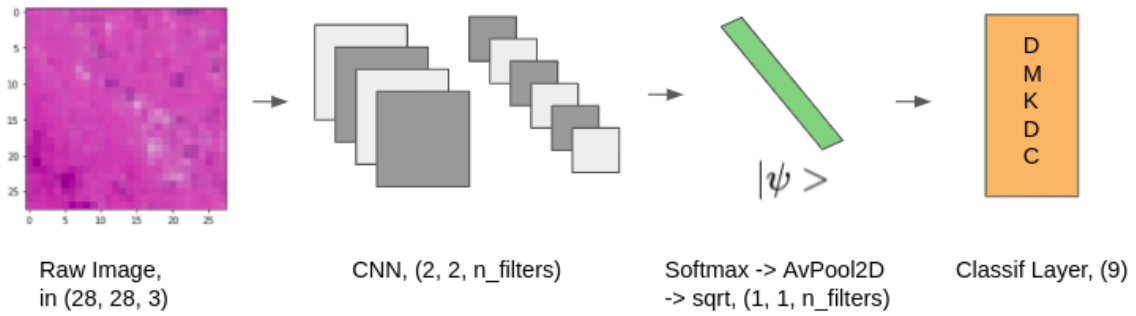


Figure 3.2.1: Method to combine CNNs with DMKDC. The images are representated as pure states to feed the DMKDC Classification layer

sigmoid to output a quantum feature map (normalized vector). The dimension of the pure state quantum feature map was 1280 for the MobileNetV2 and 2048 for the ResNet50, this dimension corresponded to the number of filters of the last layer of the CNNs. Finally this quantum feature map was the input of the DMKDC layer.

3.2.2 QMSDecomp with ConvNets

We now present another model which combines deep learning with the quantum measurement framework. In contrast to the DMKDC and QMR, which receive a pure state (normalized vector), the quantum measurement Schmidth mecomposition method (QMSDecomp) receives a density matrix as the input of the model. Hence, from the output of the convolutional layers, we represented each medical image as a density matrix to feed the QMSDecomp model.

The deep learning architecture is presented in Fig. 3.2.2. As in the previous model, we increased the image size and used the pretrained ImageNet weights from the ResNet50 and the MobileNetV2, we removed the dense layers, and used a softmax activation function in the last convolutional layer. In contrast to the ConvNets with DMKDC model, we did not performed a 2D average pooling of the output of the softmax layer, in this model we took the filters of the last layer of the CNNs of size image $n \times n$ and computed the square root, the output are $n \times n$ normalized vectors which can be considered as the eigenvectors of a density matrix with $n \times n$ eigenvalues of magnitude $1/n^2$. As with the DMKDC with ConvNets model, the size of the eigenvectors of the density matrix was 1280 and 2048 for the MobileNetV2 and ResNet50 CNNs respectively. Finally, this density matrix representation was used to feed the QMSDecomp layer to perform the classification of the

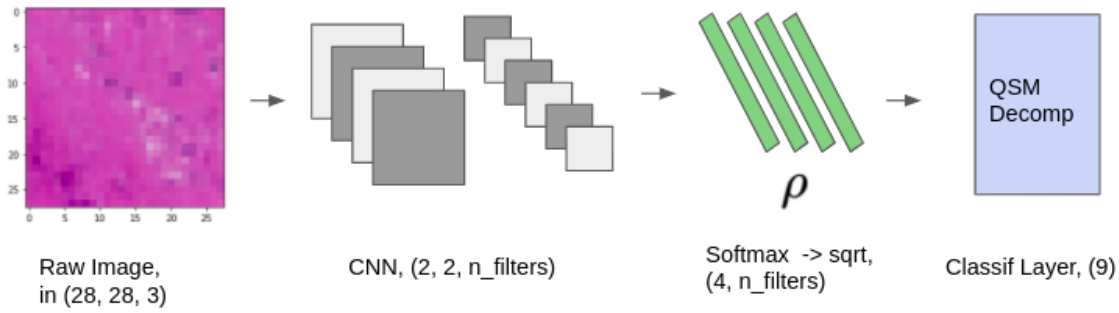


Figure 3.2.2: Method to combine CNNs with QMSDecomp. The images are represented as density matrices to feed the QMSDecomp model.

medical images.

3.2.3 PathMNIST data set

In this chapter, we worked with the PathMNIST medical data set [38]. This data set belongs to a compilation of compressed medical images which resemble the dimension and size of the MNIST data set. The PathMNIST data set is a colorectal cancer data set for classification, developed from [39]. It has 89,996 samples for training, 10,004 samples for validation, and 7,180 samples for testing, with 9 classes. The images dimensions are 28×28 with three color channels.

3.3 Results

In Table 3.3.1, we show the results of the DMKDC with ConvNets, and QMSDecomp with ConvNets models, applied to the PathMNIST data set. The baseline models are a classical convolutional neural network and a ConvNet with Random features and DMKDC as in [10]. We performed in total 10 experiments per model. In general terms, we notice that the models which combine ConvNets with QMC outperform conventional deep neural models on the PathMNIST data set. However, comparing the (ConvNet + RFF + DMKDC) model and the (ConvNet + DMKDC) we do not observe an appreciable advantage of not using RFF as part of the models, besides reducing the depth of the model. Furthermore, from the results of the (ConvNet + QMSDecomp) we also do not notice an advantage of extracting density matrices out of the convolutional layers to feed the QMSDecomp model. Given the relatively small size of the PathMNIST images (28x28x3), the density matrices

Method	Accuracy
ResNet-50 (56)	0.8918 ± 0.0070
ResNet-50 (56) + RFF () + DMKDC	0.9015 ± 0.0054
ResNet-50 (56) + DMKDC	0.9000 ± 0.0092
ResNet-50 (56) + QMSDecomp	$\mathbf{0.9039 \pm 0.0070}$
MobileNetV2 (56)	0.9044 ± 0.0022
MobileNetV2 (56) + RFF () + DMKDC	$\mathbf{0.9162 \pm 0.0028}$
MobileNetV2 (56) + DMKDC	0.9145 ± 0.0025
MobileNetV2 (56) + QMSDecomp	0.9118 ± 0.0034

Table 3.3.1: Results of the accuracy of the proposed models ConvNets with QMC and ConvNets with QMSDecomp, we took as a baseline a conventional DNNs and a Deep QMC model with RFF. We report average and standard deviation of ten experiments per model.

constructed from the CNNs had only 4 eigenvectors. It is observed that these CNNs models output filters of greater dimension for input images of bigger size, therefore, the density matrix representation could be more suitable for images of larger dimension.

3.4 Conclusions

In this thesis chapter, we presented two machine learning models for representation learning that merge classical convolutional neural networks and methods based on quantum measurement classification [1, 2]. The first model extracts probability distributions and hence quantum features maps out of the CNNs to perform supervised classification with the quantum-inspired algorithm the density matrix kernel density classification. In the second proposed model, we demonstrated that we can use CNNs to represent medical images as density matrices, and that we can apply this representation for classification with the QMSDecomp machine learning model. We found superior performance of these models compared to classical deep neural network architectures with the PathMNIST data set. However, we did not find conclusive evidence that using density matrices for feature representation of medical images is advantageous compared to classical and random Fourier features vector representations. Future perspectives include applying this framework to images of bigger dimensions, and using Transformers [40] as features extractors.

Chapter 4

Quantum Measurement Classification with Qudits

This chapter presents a hybrid classical-quantum program for density estimation and supervised classification. The program is implemented as a quantum circuit in a high-dimensional quantum computer simulator. We show that the proposed quantum protocols allow to estimate probability density functions and to make predictions in a supervised learning manner. This model can be generalized to find expected values of density matrices in high-dimensional quantum computers. Experiments on various data sets are presented. Results show that the proposed method is a viable strategy to implement supervised classification and density estimation in a high-dimensional quantum computer. This chapter corresponds to the article *Quantum measurement classification with qudits* [11] published in 2021 in the journal *Quantum Information Processing*. The complete reference of the article is presented below.

D. H. Useche, A. Giraldo-Carvajal, H. M. Zuluaga-Bucheli, J. A. Jaramillo-Villegas,and F. A. González, "Quantum measurement classification with qudits," *Quantum Information Processing* 2021 21:1, vol. 21, pp. 1–12, 12 2021.

https://link.springer.com/article/10.1007/s11128-021-03363-y

4.1 Introduction

Quantum computing has gained a lot of attention in recent years due to its potential to solve complex problems which would take exponential time in classical computers. Most of the research efforts have been focused on constructing quantum computers based on qubits [41]. However, there has been a growing interest in building quantum computers based on qudits, i.e. machines that simulate and operate d-dimensional quantum states, with $d > 2$. Various physical implementations of high-dimensional quantum states have been proposed, such as photonic states integrated in chips [42, 43], photonic modes encoded in the orbital angular momentum (OAM) [44], ion traps [45], ququarts implemented on a quadrupolar nuclear magnetic resonance (NMR) [46], and molecular quantum magnets [47]. Two of the main advantages of high-dimensional quantum computers compared to their qubit-based counterparts are their larger information storage [48], and their higher resilience to noise [49].

One closely related field of quantum computing is quantum machine learning (QML). This field aims to develop novel quantum-inspired machine learning (ML) methods that may run on classical or quantum computers and to implement the existing ML algorithms on quantum computers. For instance, some classical machine learning algorithms like support vector machines and restricted Boltzmann machines can be implemented on qubit-based quantum computers [50, 51], and many of the ML methods have been reformulated in the language of quantum physics like quantum decision trees [52], quantum neural networks [53, 54], and quantum generative adversarial networks [55]. In contrast with QML methods built on qubits, less research has been done on QML based on qudits, i.e. algorithms that run in high-dimensional quantum computers. Some of these methods include protocols with qudits for reinforcement learning [56], and for training quantum neural networks [57, 58, 59].

In addition to the aforementioned methods, Gonzalez et al. [2], proposed two quantum-inspired machine learning methods, the density matrix kernel density estimation (DMKDE), which is a non-parametric density estimation method, and the density matrix kernel density classification (DMKDC), a supervised machine learning algorithm based on density matrices and kernel density estimation. In this article, we propose two quantum protocols to implement the prediction phase of the DMKDE and the DMKDC in a high-dimensional quantum computer, the simulations were performed using the high-dimensional quantum simulator QuantumSkynet [60].

This chapter is organized as follows: In Section 4.2, we present the background with the descriptions of the DMKDE, the DMKDC, and the quantum simulator QuantumSkynet with some of its quantum gates, in Section 4.3, we describe the proposed high-dimensional quantum circuits, in Section 4.4, we show some results of the method on some toy data sets, and finally, we present the conclusions of the work in Section 4.5.

4.2 Background

In this section, we present a review of the density matrix kernel density estimation (DMKDE), and the density matrix kernel density classification (DMKDC) methods proposed by Gonzalez et al. [2], which are the basis of this article. In addition, we describe the high-dimensional quantum computer simulator, QuantumSkynet, with some qudit-based quantum gates which were applied in this work.

4.2.1 Density matrix kernel density estimation

The density matrix kernel density estimation [2] method starts by computing a quantum feature map $x_i \to |\psi_i\rangle$ based on random Fourier features (RFF) [9], over a training data set $X = \{x_i\}_{i=1,\cdots,N}$, where $|\psi_i\rangle$ is a normalized vector. Then a training density matrix ρ is constructed as a maximally mixed state of all the N training samples,

$$\rho = \frac{1}{N} \sum_{i=1}^{N} |\psi_i\rangle \langle\psi_i| . \tag{4.1}$$

To predict the density of a testing sample $x \to |\psi\rangle$, the expected value of the sample with the training density matrix is computed,

$$\langle\psi| \rho |\psi\rangle . \tag{4.2}$$

The DMKDE in conjunction with RFF works as a non-parametric density estimator, which can approximate probability density functions.

4.2.2 Density matrix kernel density classification

The DMKDE can be used for classification as in the density matrix kernel density classification method (DMKDC) [2]. This algorithm creates a quantum feature map of the training and testing samples $x_i \to |\psi_i\rangle$, some possible quantum feature maps are based on RFF, and soft-max encoding, as presented in [1]. Then for each class $j \in \{0, \cdots, D-1\}$, it computes a training density matrix ρ_j. The relative frequency, also called prior, of the training samples per class is computed, $\pi_j = N_j/N$, with N the total number of training data points, and N_j the number of training samples of class j. The probability P_j of a testing sample $x \to |\psi\rangle$ to belong to class j, would be given by,

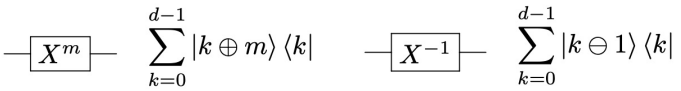
$$P_j = \frac{\pi_j \langle\psi| \rho_j |\psi\rangle}{\sum_{k=0}^{D-1} \pi_k \langle\psi| \rho_k |\psi\rangle}. \tag{4.3}$$

These density matrices ρ_j can be trained as an average mixed state of the training samples of each class (see equation 4.1), or by stochastic gradient descent, which looks for the optimal parameters of the spectral decomposition of the density matrices, $\rho_j = U_j \Lambda_j U_j^\dagger$ with the training data points, using a categorical cross-entropy loss function, see more details in [2].

4.2.3 QuantumSkynet and high-dimensional quantum gates

QuantumSkynet [60] is a high dimensional quantum computing simulator, that allows to implement high-dimensional quantum algorithms in a cloud-based environment. To simulate the quantum circuits related to this project, the following gates were simulated using QuantumSkynet:

1. The single-qudit gate X^m (see Fig. 4.2.1a), which is a generalized version of the qubit-based X gate for d dimensions and raised to any m exponent. A particular case of this gate X is when the exponent m is equal to -1 (see Fig. 4.2.1b).



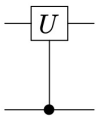
(a) (b)

Figure 4.2.1: (a) High-dimensional gate X^m. (b) High-dimensional gate X^{-1}.

In Fig. 4.2.1, $\oplus$ and $\ominus$ stand for summation and subtraction modulo d, respectively.

The result of applying the X^m to the canonical basis is $X^m |i\rangle = |i+m\rangle$. In particular, $X^{-1} |i\rangle = |i-1\rangle$.

2. The control gate CU (See Fig. 4.2.2). This control gate applies an arbitrary unitary matrix U only when the control qudit takes the value $|1\rangle$.



$$P_1 \otimes U + \sum_{k\neq 1}^{d-1} P_k \otimes I$$

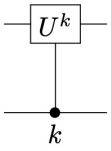
Figure 4.2.2: High-dimensional control gate CU.

Here, P_k is considered as the projection operator equals to $|k\rangle \langle k|$. For an arbitrary state which results from the tensor product of two qudits, the CU gate does the following transformation,

$$CU\Big(\sum_{i=0}^{d-1} a_i |i\rangle \otimes \sum_{j=0}^{d-1} b_j |j\rangle\Big) = \sum_{\{i:i\neq 1\}} a_i |i\rangle \otimes \sum_{j=0}^{d-1} b_j |j\rangle + a_1 |1\rangle \otimes U(\sum_{j=0}^{d-1} b_j |j\rangle),$$

with the first qudit as the control and the second qudit as the target.

3. The generalized controlled gate CU^k (See Fig. 4.2.3). This gate applies the gate U^k to the target qudit, when the control qudit is in state $|k\rangle$, for each possible state of the canonical basis.



$$\sum_{k=0}^{d-1} P_k \otimes U^k$$

Figure 4.2.3: High-dimensional generalized control gate CU^k.

One case of interest is when U is equal to X^{-1}. In this case, the gate $C(X^{-1})^k$ does the following transformation to an arbitrary two qudit state,

$$C(X^{-1})^k\Big(\sum_{(i,j)} a_{ij}\,|ij\rangle\Big) = \sum_{(i,j)} a_{ij}\,|i(j-i)\rangle\,, \tag{4.4}$$

where the control is the first qudit and the target is the second qudit.

The CU^k gate can be obtained by a series of CU and X^m gates, by the same way as the multiplexer gate [61].

4.3 Quantum measurement classification with qudits

The implementation of the algorithms DMKDE and DMKDC requires three phases: (i) quantum state preparation, (ii) training phase, and (iii) prediction phase. The first two steps were computed in a classical computer, with the Tensorflow implementations of the algorithms [2], while the prediction phase was simulated in the high dimensional quantum computer simulator QuantumSkynet [60].

The steps of the DMKDE and DMKDC implementations are:

1. Quantum state preparation: Apply a suitable quantum feature map to the train and test data sets.

2. Training phase: Construct the matrices ρ_j, one for each class, as a mixed state of the training quantum states, see equation 4.1, compute the priors π_j of each class, and, calculate the spectral decomposition of these density matrices $\rho_j = U_j \Lambda_j U_j^\dagger$ (in the DMKDE method there is only one class).

3. Prediction phase: Apply the proposed quantum circuit to make the prediction on each quantum state of the test data set, see equations 4.2, 4.3.

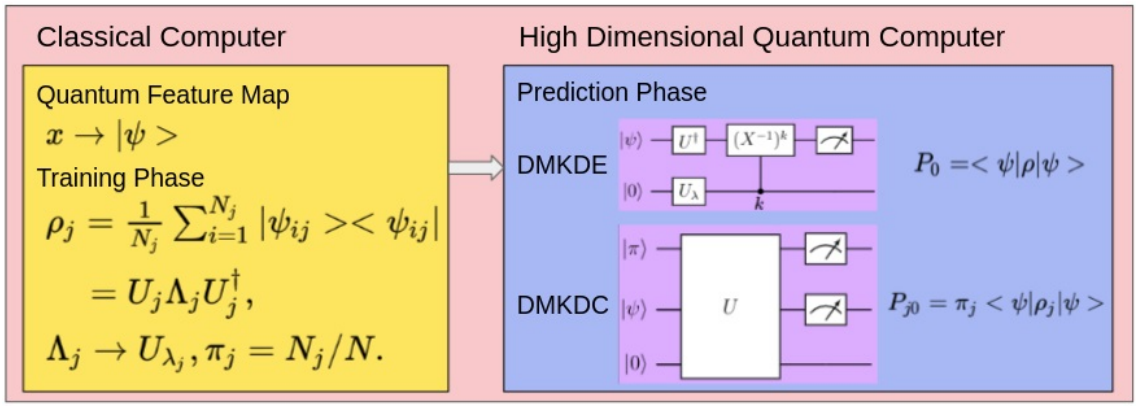


Figure 4.3.1: Qudit-based implementation of DMKDE and DMKDC methods. The quantum feature map and the training were performed in a classical computer, while the prediction was done in a high-dimensional quantum computer simulator.

The main contribution of this chapter is to show a concrete implementation of the prediction phase of DMKDE and DMKDC as quantum circuits that can be run in a high-dimensional quantum computer. Next, we will present the details of these quantum protocols.

4.3.1 Initial comments of the DMKDE and DMKDC quantum circuits

To implement the prediction phase of DMKDE in a quantum computer with qudits, we should notice that it is equivalent to finding the expected value of a quantum state $|\psi\rangle \in \mathbb{C}^d$ with a training hermitian matrix ρ, see equation 4.2, therefore, we can apply a spectral decomposition of ρ,

$$\langle\psi|\,\rho\,|\psi\rangle = \langle\psi|\,U\Lambda U^\dagger\,|\psi\rangle\,, \tag{4.5}$$

with Λ a diagonal matrix with $\mathrm{Tr}(\Lambda) = 1$, and U a unitary matrix. But, $\Lambda = \sum_{i=0}^{d-1} \lambda_i\,|i\rangle\,\langle i|$, then,

$$\langle\psi|\,\rho\,|\psi\rangle = \langle\psi|\,U\Big(\sum_{i=0}^{d-1}\lambda_i\,|i\rangle\,\langle i|\,\Big)U^\dagger\,|\psi\rangle = \sum_{i=0}^{d-1}\lambda_i\Big\|\langle i|\,U^\dagger\,|\psi\rangle\Big\|^2. \tag{4.6}$$

This form of the DMKDE can be implemented in a high-dimensional quantum computer.

In addition, it is worth mentioning that the DMKDE quantum circuit starts by a assuming

we have a suitable quantum feature map of the testing sample $x \to |\psi\rangle$. The quantum feature map might be based on random Fourier features or soft-max encoding, as presented in [1]. Also, we should have a training density matrix ρ and the resulting matrices of its spectral decomposition U and Λ. In addition, based on the matrix of eigenvalues Λ, we require the unitary transformation U_λ, which satisfies that $U_\lambda |0\rangle = |\lambda\rangle$, where,

$$|\lambda\rangle = \sum_{i=0}^{d-1} \sqrt{\lambda_i}\, |i\rangle\,. \tag{4.7}$$

That is the state $|\lambda\rangle$ is a quantum state which encodes the eigenvalues of the spectral decomposition of ρ.

The same previous arguments can be extended to the DMKDC method, but instead of having only one training density matrix ρ, in the DMKDC we should have a density matrix ρ_j for each class of the data set.

These previous steps, i.e., the quantum feature map, the calculation of the training density matrices ρ_j, and their spectral decompositions, were done in a classical computer, following the tensorflow implementation of the method [2].

4.3.2 The DMKDE quantum circuit

The diagram of the DMKDE quantum circuit is presented in figure 4.3.2. It requires two qudits each of dimension d. From the classical computer, we obtain the quantum feature map of the input sample $|\psi\rangle \in \mathbb{C}^d$, the matrix U of eigenvectors of the spectral decomposition of ρ, and the unitary matrix U_λ, which satisfies that, $U_\lambda |0\rangle = |\lambda\rangle$, see equation 4.7.

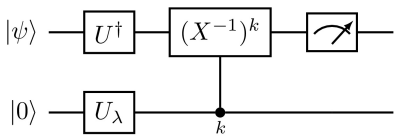


Figure 4.3.2: DMKDE high-dimensional quantum circuit.

The circuit is initialized with $|\psi\rangle$ in the first qudit, and $|0\rangle$, in the second qudit,

$$|\psi\rangle \otimes |0\rangle\,, \tag{4.8}$$

being the first qudit the leftmost in the equation 4.8.

Then, we apply the unitary transformation $U^\dagger$ to the first qudit and U_λ to the second qudit,

$$U^\dagger |\psi\rangle \otimes U_\lambda |0\rangle = U^\dagger |\psi\rangle \otimes \sum_{j=0}^{d-1} \sqrt{\lambda_j} |j\rangle . \tag{4.9}$$

We can write $U^\dagger |\psi\rangle = \sum_{i=0}^{d-1} a_i |i\rangle$, where $\|a_i\|^2 = \left\| \langle i| U^\dagger |\psi\rangle \right\|^2$. Hence, the coefficients $\|a_i\|^2$ are the probabilities to measure $U^\dagger |\psi\rangle$ in the canonical basis. Then we can write,

$$\begin{aligned} U^\dagger |\psi\rangle \otimes U_\lambda |0\rangle &= \sum_{i=0}^{d-1} a_i |i\rangle \otimes \sum_{j=0}^{d-1} \sqrt{\lambda_j} |j\rangle \\ &= \sum_{i=0}^{d-1} a_i \sqrt{\lambda_i} |ii\rangle + \sum_{\{(i,j):i\neq j\}} a_i \sqrt{\lambda_j} |ij\rangle . \end{aligned}$$

We can then apply the generalized control gate $C(X^{-1})^k$ with control the second qudit and target the first qudit. Which based on equation 4.4 results in,

$$C(X^{-1})^k (U^\dagger |\psi\rangle \otimes U_\lambda |0\rangle) = \sum_{i=0}^{d-1} a_i \sqrt{\lambda_i} |0i\rangle + \sum_{\{(i,j):i\neq j\}} a_i \sqrt{\lambda_j} |(i-j)j\rangle . \tag{4.10}$$

Finally, by measuring the first qudit the probability of the state $|0\rangle$ is,

$$P_0 = \sum_{i=0}^{d-1} \|a_i\|^2 \lambda_i = \sum_{i=0}^{d-1} \lambda_i \left\| \langle i| U^\dagger |\psi\rangle \right\|^2 = \langle \psi| \rho |\psi\rangle , \tag{4.11}$$

see equation 4.6.

4.3.3 The DMKDC quantum circuit

The proposed high-dimensional quantum circuit of the DMKDC is presented in figure 4.3.3. As mentioned in section 4.2.2, the DMKDC algorithm requires D density matrices ρ_j, one for each class, whose spectral decompositions are given by $\rho_j = U_j \Lambda_j U_j^\dagger$. These training density matrices are computed by equation 4.1 in a classical computer.

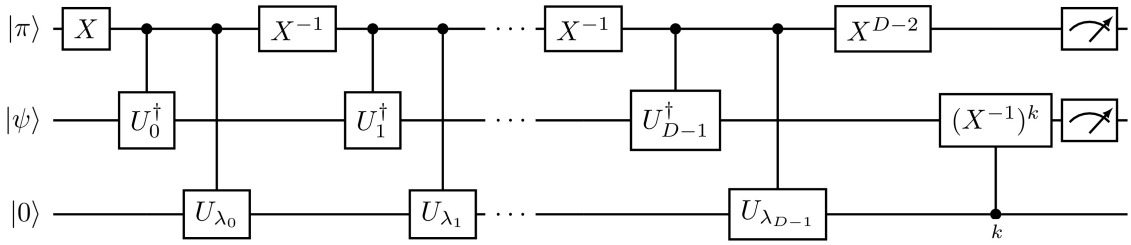


Figure 4.3.3: DMKDC high-dimensional quantum circuit.

The method requires three qudits in $\mathbb{C}^d$, assuming $d \geq D$. The first qudit encodes each of the D classification classes, and the relative frequencies of the training data per class π_j (priors). The second qudit serves for two purposes, the input sample to be classified $x \to |\psi\rangle$ in $\mathbb{C}^d$, at which a suitable quantum feature map has been applied in advance, and for the matrices of eigenvectors U_j, each class has a unitary matrix of eigenvectors. In addition, the third qudit is responsible for the eigenvalues of each of the trained density matrices, this can be achieved by the rotation matrices U_{λ_j}, which have the property that $U_{\lambda_j} |0\rangle = |\lambda_j\rangle = \sum_{i=0}^{d-1} \sqrt{\lambda_{ij}} |i\rangle$, where λ_{ij} is the i^{th} eigenvalue of the density matrix ρ_j.

The circuit is initialized by $|\pi\rangle = \sum_{j=0}^{D-1} \sqrt{\pi_j} |j\rangle$ in the first qudit, by $|\psi\rangle$ in the second qudit, and by $|0\rangle$ in the third qudit,

$$|\pi\rangle \otimes |\psi\rangle \otimes |0\rangle = \Big(\sum_{j=0}^{D-1} \sqrt{\pi_j} |j\rangle \Big) \otimes |\psi\rangle \otimes |0\rangle \,, \tag{4.12}$$

Since the first qudit encodes the classes of the algorithm, the circuit works by changing the control class of the D classes with X^n gates and applying control CU gates for the eigenvalues and eigenvectors of each of the classes. Hence, we first apply a X gate to the first qudit, to change the control class from class 1 to class 0, the result is,

$$X(|\pi\rangle) \otimes |\psi\rangle \otimes |0\rangle = \Big(\sum_{j=0}^{D-1} \sqrt{\pi_j} \, |j+1\rangle \Big) \otimes |\psi\rangle \otimes |0\rangle \, . \tag{4.13}$$

We then apply the $CU_0^\dagger$ gate with control qudit the first qudit and target the second qudit, and the CU_{λ_0} with control qudit the first qudit and target the third qudit, see figure 4.3.3,

$$\sqrt{\pi_0} \, |1\rangle \otimes U_0^\dagger \, |\psi\rangle \otimes |\lambda_0\rangle + \Big(\sum_{j=1}^{D-1} \sqrt{\pi_j} \, |j+1\rangle \Big) \otimes |\psi\rangle \otimes |0\rangle \, . \tag{4.14}$$

To replicate the process for class 1, we change the control class from class 0 to class 1, by applying the gate X^{-1} to the first qudit, the result is,

$$\sqrt{\pi_0} \, |0\rangle \otimes U_0^\dagger \, |\psi\rangle \otimes |\lambda_0\rangle + \Big(\sum_{j=1}^{D-1} \sqrt{\pi_j} \, |j\rangle \Big) \otimes |\psi\rangle \otimes |0\rangle \, . \tag{4.15}$$

We apply the $CU_1^\dagger$ gate with control qudit the first qudit and target the second qudit, and the CU_{λ_1} with control qudit the first qudit and target the third qudit,

$$\sqrt{\pi_0} \, |0\rangle \otimes U_0^\dagger \, |\psi\rangle \otimes |\lambda_0\rangle + \sqrt{\pi_1} \, |1\rangle \otimes U_1^\dagger \, |\psi\rangle \otimes |\lambda_1\rangle + \Big(\sum_{j=2}^{D-1} \sqrt{\pi_j} \, |j\rangle \Big) \otimes |\psi\rangle \otimes |0\rangle \, . \tag{4.16}$$

We then extend the same block of the class 1 to the restating $D-2$ classes, leaving the class $D-1$ as the control class at the end. Hence, we would have that,

$$\sum_{j=0}^{D-1} \big(\sqrt{\pi_j} \, |j-(D-2)\rangle \otimes U_j^\dagger \, |\psi\rangle \otimes |\lambda_j\rangle \big). \tag{4.17}$$

To restore the class j to the corresponding $|j\rangle$, we apply the gate X^{D-2} to the first qudit,

$$\sum_{j=0}^{D-1} \left(\sqrt{\pi_j}\,|j\rangle \otimes U_j^\dagger\,|\psi\rangle \otimes |\lambda_j\rangle\,\right). \tag{4.18}$$

As with the DMKDE, we can write,

$$\begin{aligned} U_j^\dagger\,|\psi\rangle \otimes |\lambda_j\rangle &= \sum_{l=0}^{d-1} a_{lj}\,|l\rangle \otimes \sum_{m=0}^{d-1} \sqrt{\lambda_{mj}}\,|m\rangle \\ &= \sum_{l=0}^{d-1} a_{lj}\sqrt{\lambda_{lj}}\,|ll\rangle + \sum_{\{(l,m):l\neq m\}} a_{lj}\sqrt{\lambda_{mj}}\,|lm\rangle\,, \end{aligned} \tag{4.19}$$

where $\|a_{lj}\|^2 = \left\|\langle l|\,U_j^\dagger\,|\psi\rangle\right\|^2$.

Finally, by the same argument of the DMKDE, we would have that by applying the generalized $C(X^{-1})^k$ with control qudit the third qudit and target the second qudit, the circuit leads,

$$\sum_{j=0}^{D-1} \left(\sqrt{\pi_j}\,|j\rangle \otimes \Big(\sum_{l=0}^{d-1} a_{lj}\sqrt{\lambda_{lj}}\,|0l\rangle + \sum_{\{(l,m):l\neq m\}} a_{lj}\sqrt{\lambda_{mj}}\,|(l-m)m\rangle\,\Big)\right). \tag{4.20}$$

The desired result is achieved by measuring the amplitudes of the first two qudits. We would have that,

$$P_{j0} = \pi_j \sum_{l=0}^{d-1} \lambda_{lj} \left\|\langle l|\,U_j^\dagger\,|\psi\rangle\right\|^2 = \pi_j\,\langle\psi|\,\rho_j\,|\psi\rangle\,, \tag{4.21}$$

see equation 4.6. The sample $|\psi\rangle$ will be classified based on,

$$\max_j(\pi_j\,\langle\psi|\,\rho_j\,|\psi\rangle). \tag{4.22}$$

4.4 Results

We applied the DMKDE and DMKDC circuits to two data sets. We found that the results of the quantum circuits simulated in the high-dimensional quantum simulator Quantum-Skynet mimic the results obtained in the Tensorflow implementation of the DMKDE and DMKDC by Gonzalez et. al. [2].

For the DMKDE method, we used a 1-D synthetic data set, The training data set corresponded to 1000 points sampled from the linear combination of two Gaussian functions, and there were 1000 equally spaced data points for testing as in [2]. In that article, they show that DMKDE in combination with random Fourier features (RFF) can approximate any probability density function (pdf).

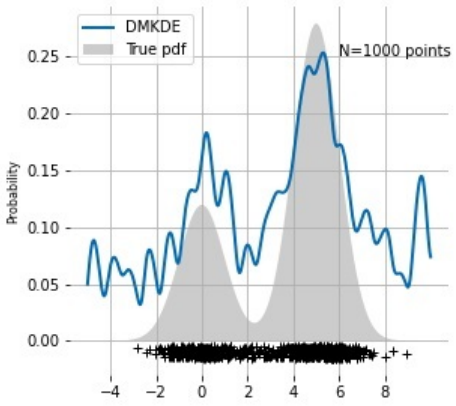
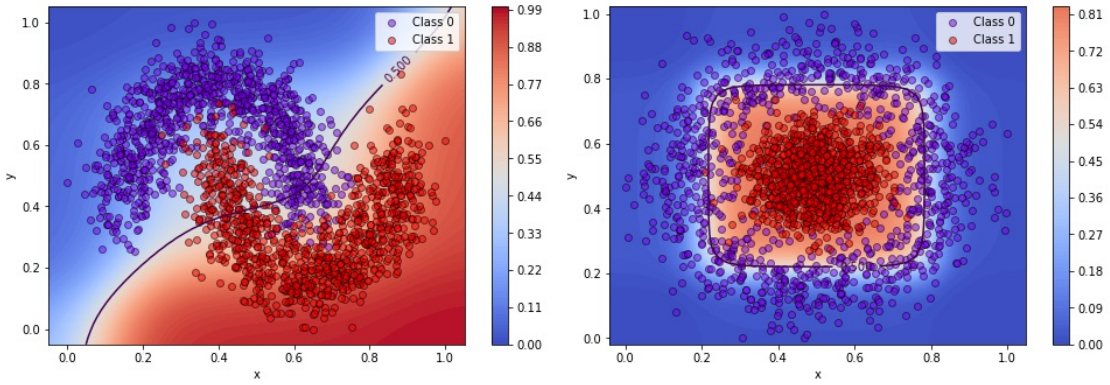


Figure 4.4.1: Predictions of the DMKDE quantum circuit with 18-dits of RFF. The DMKDE can approximate probability density functions with high enough RFF.

For the implementation of the DMKDE in the high-dimensional quantum computer simulator, we applied a quantum feature map based on RFF to the raw data with 18 components. Figure 4.4.1 shows the results of the predictions of the DMKDE quantum circuit. Even though more RFF components would create a better approximation of the pdf, we were restricted by the maximum number of qudit components allowed by the quantum computer simulator.

Furthermore, we classified two two-dimensional binary data sets of moons and circles to test the DMKDC quantum circuit, see Figure 4.4.2. There were 1340 samples for training and 660 for testing in each data set. A quantum feature map based on the softmax



(a) Moons (b) Circles

Figure 4.4.2: Binary predictions of the DMKDC circuit with 9-dits on moons and circles.

encoding [1] was applied to each data set, resulting in quantum features of 9 dimensions. Therefore, the quantum circuit was constructed with qudits of 9 components. In Figure 4.4.2, we show the classification boundaries, and the regions with higher probabilities to be classified as either class. We obtained an accuracy of 86.66% on the test data set of moons, and of 83.63% on the test data set of circles. The results of the predictions with the high-dimensional quantum circuit are consistent with the Tensorflow implementation of the DMKDC.

Higher accuracies in the DMKDC quantum circuit would have been obtained, if we had used a quantum feature map based on RFF with a higher number of components, and if we had implemented a quantum circuit learning, in which the weights of the trained density matrices are learned by methods like back-propagation.

4.5 Conclusions

In the present chapter, we showed how to implement in a high-dimensional quantum computer the prediction phase of the quantum-inspired machine learning methods density matrix kernel density estimation (DMKDE), and density matrix kernel density classification (DMKDC) proposed by Gonzalez et. al. [2]. The DMKDE and DMKDC quantum circuits were simulated in the qudit-based quantum computer simulator QuantumSkynet [60]. The DMKDE quantum protocol can be extended to compute the expected value of a density matrix with qudits.

Much work is to be done on reducing the complexity of the quantum computer simulator to apply the DMKDC method to more realistic machine learning problems like MNIST, and to improve the capability of the DMKDE to approximate probability density functions by increasing the number of random Fourier features. Furthermore, this framework opens up the possibility to implement the DMKDE and DMKDC algorithms with stochastic-gradient descent in which the weights of the trained density matrices are learned by some optimization procedure, improving the performance of the density estimation and classification.

Chapter 5

Conclusions and future work

In this thesis, we proposed various classical-quantum models which combine convolutional neural networks with methods based on quantum measurement classification (QMC) [1, 2] and applied them to the task of classifying medical images [10], see Chapters 2 and 3. Furthermore, we showed that it is possible to implement the QMC-based machine learning models density matrix kernel density estimation (DMKDE) and density matrix kernel density classification (DMKDC) in a high-dimensional quantum computer [11], see Chapter 4.

We first proposed the deep quantum measurement ordinal regression (DQMOR) model [10], this algorithm combines convolutional neural networks with random Fourier features (RFF) and the quantum measurement regressor (QMR) method [2]. We applied the model to the prostate histopathology TCGA-PRAD data set. We showed that we can approach the prediction of patches and Whole-Slide-Images (WSI) as a regression task, obtaining results comparable with the state-of-the-art. In contrast to classical deep learning architectures, the output of the classification of this model can be interpreted more straightforwardly as a probability measure, and hence one can extract the uncertainty of the predictions. These features might be useful for pathologists to make medical decisions, for example, the specialist may not trust a prediction with an associated high error.

Furthermore, in Chapter 3 we presented various strategies for representation learning of the cancer colorectal images (PathMNIST [38]). We showed that we can extract quantum feature maps, i.e. probability distributions, out of convolutional neural networks and use the DMKDC to classify such representations. In addition, we proposed another technique

for feature representation, which maps images to density matrices, and used the QMSDecomp model, see Sect. 3.2.2, to classify the features. We found that these models were superior to conventional deep learning architectures in the PathMNIST data set in terms of accuracy. Nonetheless, we did not find an advantage of not using RFF in the models, or the density matrix representation, in comparison with a model with RFF and the DMKDC.

In this work, we also proposed two quantum protocols that implement the prediction phase of the DMKDE and DMKDC algorithms in a qudit-based quantum computer [11], see Chapter 4. We showed that it is possible to perform supervised classification and density estimation with mixed states in quantum computers. The DMKDE quantum protocol is equivalent to estimating the expected value of a mixed state density matrix in a high-dimensional quantum computer, which might translate to some other applications in quantum machine learning.

For future work, we aim to develop multimodal machine learning methods which incorporate texts of the doctors' diagnosis to the task of medical image classification. In addition, we will further explore the representation learning techniques based on quantum pure states and mixed density matrices with CNNs, and use other machine learning techniques for features' extraction like transformers [40]. In terms of the developed DMKDE and DMKDC quantum protocols, we would like to apply these quantum protocols to larger data sets like MNIST, and look for other potential applications of the proposals to physics and quantum machine learning.

Bibliography

[1] F. A. González, V. Vargas-Calderón, and H. Vinck-Posada, "Classification with Quantum Measurements," *https://doi.org/10.7566/JPSJ.90.044002*, vol. 90, no. 4, 3 2021. [Online]. Available: https://journals.jps.jp/doi/abs/10.7566/JPSJ.90.044002

[2] F. A. González, A. Gallego, S. Toledo-Cortés, and V. Vargas-Calderón, "Learning with Density Matrices and Random Features," 2 2021. [Online]. Available: https://arxiv.org/abs/2102.04394v4

[3] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature 2015 521:7553*, vol. 521, no. 7553, pp. 436–444, 5 2015. [Online]. Available: https://www.nature.com/articles/nature14539

[4] D. E. Rumelhart and J. L. McClelland, "Learning Internal Representations by Error Propagation - MIT Press books," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*. MIT Press, 1987, pp. 318–362.

[5] L. Cun, J. Henderson, Y. Le Cun, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Handwritten Digit Recognition with a Back-Propagation Network," Tech. Rep.

[6] A. Cruz-Roa, A. Basavanhally, F. González, H. Gilmore, M. Feldman, S. Ganesan, N. Shih, J. Tomaszewski, and A. Madabhushi, "Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks," in *Medical Imaging 2014: Digital Pathology*, vol. 9041, 2014, p. 904103. [Online]. Available: http://spiedl.org/terms

[7] S. Otálora, O. Perdomo, F. González, and H. Müller, "Training Deep Convolutional Neural Networks with Active Learning for Exudate Classification in Eye Fundus

Images," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10552 LNCS, 2017, pp. 146–154. [Online]. Available: http://www.who.int/diabetes/en/

[8] O. Perdomo, S. Otalora, F. Rodríguez, J. Arevalo, and F. A. González, "A Novel Machine Learning Model Based on Exudate Localization to Detect Diabetic Macular Edema," *Ophthalmic Medical Image Analysis International Workshop*, vol. 3, no. 2016, pp. 137–144, 10 2016. [Online]. Available: https://pubs.lib.uiowa.edu/omia/article/id/27635/

[9] A. Rahimi and B. Recht, "Random features for large-scale kernel machines," in *Advances in Neural Information Processing Systems 20 - Proceedings of the 2007 Conference*, 2009.

[10] S. Toledo-Cortés, D. H. Useche, and F. A. González, "Prostate Tissue Grading with Deep Quantum Measurement Ordinal Regression," 3 2021. [Online]. Available: https://arxiv.org/abs/2103.03188v1

[11] D. H. Useche, A. Giraldo-Carvajal, H. M. Zuluaga-Bucheli, J. A. Jaramillo-Villegas, and F. A. González, "Quantum measurement classification with qudits," *Quantum Information Processing 2021 21:1*, vol. 21, no. 1, pp. 1–12, 12 2021. [Online]. Available: https://link.springer.com/article/10.1007/s11128-021-03363-y

[12] S. F. Faraj, S. M. Bezerra, K. Yousefi, H. Fedor, S. Glavaris, M. Han, A. W. Partin, E. Humphreys, J. Tosoian, M. H. Johnson, E. Davicioni, B. J. Trock, E. M. Schaeffer, A. E. Ross, and G. J. Netto, "Clinical Validation of the 2005 ISUP Gleason Grading System in a Cohort of Intermediate and High Risk Men Undergoing Radical Prostatectomy," *PLOS ONE*, vol. 11, no. 1, p. e0146189, 1 2016. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0146189

[13] Y. Li, M. Huang, Y. Zhang, J. Chen, H. Xu, G. Wang, and W. Feng, "Automated Gleason Grading and Gleason Pattern Region Segmentation Based on Deep Learning for Pathological Images of Prostate Cancer," *IEEE Access*, vol. 8, pp. 117 714–117 725, 2020.

[14] P. Ström, K. Kartasalo, H. Olsson, L. Solorzano, B. Delahunt, D. M. Berney, D. G. Bostwick, A. J. Evans, D. J. Grignon, P. A. Humphrey, K. A. Iczkowski, J. G. Kench, G. Kristiansen, T. H. van der Kwast, K. R. Leite, J. K. McKenney, J. Oxley, C. C.

Pan, H. Samaratunga, J. R. Srigley, H. Takahashi, T. Tsuzuki, M. Varma, M. Zhou, J. Lindberg, C. Lindskog, P. Ruusuvuori, C. Wählby, H. Grönberg, M. Rantalainen, L. Egevad, and M. Eklund, "Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study," *The Lancet Oncology*, vol. 21, no. 2, pp. 222–232, 2 2020.

[15] W. Bulten, H. Pinckaers, H. van Boven, R. Vink, T. de Bel, B. van Ginneken, J. van der Laak, C. Hulsbergen-van de Kaa, and G. Litjens, "Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study," *The Lancet Oncology*, vol. 21, no. 2, pp. 233–241, 2 2020.

[16] J. S. Lara, V. H. Contreras O, S. Otálora, H. Müller, and F. A. González, "Multimodal Latent Semantic Alignment for Automated Prostate Tissue Classification and Retrieval," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12265 LNCS. Springer Science and Business Media Deutschland GmbH, 2020, pp. 572–581.

[17] J. Vaicenavicius, D. Widmann, C. Andersson, F. Lindsten, J. Roll, and T. B. Schön, "Evaluating model calibration in classification," pp. 3459–3467, 4 2019. [Online]. Available: https://proceedings.mlr.press/v89/vaicenavicius19a.html

[18] K. Cutajar, E. V. Bonilla, P. Michiardi, and M. Filippone, "Random Feature Expansions for Deep Gaussian Processes," pp. 884–893, 7 2017. [Online]. Available: https://proceedings.mlr.press/v70/cutajar17a.html

[19] C. E. Rasmussen, "Gaussian Processes in Machine Learning," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3176, pp. 63–71, 2003. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-540-28650-9_4

[20] O. Jiménez del Toro, M. Atzori, S. Otálora, M. Andersson, K. Eurén, M. Hedlund, P. Rönnquist, and H. Müller, "Convolutional neural networks for an automatic classification of prostate tissue slides with high-grade Gleason score," in *Medical Imaging 2017: Digital Pathology*, vol. 10140, 2017, p. 101400O. [Online]. Available: http://cancergenome.nih.gov/,

[21] Y. Tolkach, T. Dohmgörgen, M. Toma, and G. Kristiansen, "High-accuracy prostate cancer pathology using deep learning," *Nature Machine Intelligence*, vol. 2, no. 7,

pp. 411–418, 7 2020. [Online]. Available: https://www.nature.com/articles/s42256-020-0200-7

[22] D. Karimi, G. Nir, L. Fazli, P. C. Black, L. Goldenberg, and S. E. Salcudean, "Deep Learning-Based Gleason Grading of Prostate Cancer from Histopathology Images - Role of Multiscale Decision Aggregation and Data Augmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 5, pp. 1413–1426, 5 2020.

[23] E. Esteban, M. López-Pérez, A. Colomer, M. A. Sales, R. Molina, and V. Naranjo, "A new optical density granulometry-based descriptor for the classification of prostate histological images using shallow and deep Gaussian processes," *Computer Methods and Programs in Biomedicine*, vol. 178, pp. 303–317, 9 2019.

[24] M. Lucas, I. Jansen, C. D. Savci-Heijink, S. L. Meijer, O. J. de Boer, T. G. van Leeuwen, D. M. de Bruin, and H. A. Marquering, "Deep learning for automatic Gleason pattern classification for grade group determination of prostate biopsies," *Virchows Archiv*, vol. 475, no. 1, pp. 77–83, 7 2019. [Online]. Available: https://doi.org/10.1007/s00428-019-02577-x

[25] A. A. Khani, S. A. Fatemi Jahromi, H. O. Shahreza, H. Behroozi, and M. S. Baghshah, "Towards Automatic Prostate Gleason Grading Via Deep Convolutional Neural Networks," in *5th Iranian Conference on Signal Processing and Intelligent Systems, ICSPIS 2019*. Institute of Electrical and Electronics Engineers Inc., 12 2019.

[26] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 1800–1807, 10 2016. [Online]. Available: https://arxiv.org/abs/1610.02357v3

[27] J. Arevalo, F. A. González, R. Ramos-Pollán, J. L. Oliveira, and M. A. Guevara Lopez, "Representation learning for mammography mass lesion classification with convolutional neural networks," *Computer Methods and Programs in Biomedicine*, vol. 127, pp. 248–257, 4 2016.

[28] C. K. Shie, C. H. Chuang, C. N. Chou, M. H. Wu, and E. Y. Chang, "Transfer representation learning for medical image analysis," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2015-Novem, pp. 711–714, 11 2015.

[29] Q. Tang, Y. Liu, and H. Liu, "Medical image classification via multiscale representation learning," *Artificial Intelligence in Medicine*, vol. 79, pp. 71–78, 6 2017.

[30] T. Moriya, H. R. Roth, S. Nakamura, H. Oda, K. Nagara, M. Oda, and K. Mori, "Unsupervised segmentation of 3D medical images based on clustering and deep representation learning," *https://doi.org/10.1117/12.2293414*, vol. 10578, pp. 483–489, 3 2018. [Online]. Available: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10578/1057820/Unsupervised-segmentation-of-3D-medical-images-based-on-clustering-and/10.1117/12.2293414.fullhttps://www.spiedigitallibrary.org/conference-proceedings-of-spie/10578/1057820/

[31] A. Chartsias, T. Joyce, G. Papanastasiou, S. Semple, M. Williams, D. E. Newby, R. Dharmakumar, and S. A. Tsaftaris, "Disentangled representation learning in cardiac image analysis," *Medical Image Analysis*, vol. 58, p. 101535, 12 2019.

[32] N. Dong, M. Kampffmeyer, and I. Voiculescu, "Self-supervised Multi-task Representation Learning for Sequential Medical Images," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12977 LNAI, pp. 779–794, 9 2021. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-86523-8_47

[33] P. Zhang, F. Wang, and Y. Zheng, "Self supervised deep representation learning for fine-grained body part recognition," *Proceedings - International Symposium on Biomedical Imaging*, pp. 578–582, 6 2017.

[34] M. Adnan, S. Kalra, and H. R. Tizhoosh, "Representation Learning of Histopathology Images using Graph Neural Networks," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2020-June, pp. 4254–4261, 4 2020. [Online]. Available: https://arxiv.org/abs/2004.07399v2

[35] M. M. Li, K. Huang, and M. Zitnik, "Graph Representation Learning in Biomedicine," 4 2021. [Online]. Available: https://arxiv.org/abs/2104.04883v2

[36] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[37] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern*

Recognition, CVPR 2005, vol. I, pp. 886–893, 2005.

[38] J. Yang, R. Shi, D. Wei, Z. Liu, L. Zhao, B. Ke, H. Pfister, and B. Ni, "MedMNIST v2: A Large-Scale Lightweight Benchmark for 2D and 3D Biomedical Image Classification," 10 2021. [Online]. Available: https://arxiv.org/abs/2110.14795v1

[39] J. N. Kather, J. Krisam, P. Charoentong, T. Luedde, E. Herpel, C. A. Weis, T. Gaiser, A. Marx, N. A. Valous, D. Ferber, L. Jansen, C. C. Reyes-Aldasoro, I. Zörnig, D. Jäger, H. Brenner, J. Chang-Claude, M. Hoffmeister, and N. Halama, "Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study," *PLOS Medicine*, vol. 16, no. 1, p. e1002730, 2019. [Online]. Available: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002730

[40] A. Vaswani, G. Brain, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Kaiser, and I. Polosukhin, "Attention is All you Need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[41] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. Mcclean, M. Mcewen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, "Quantum supremacy using a programmable superconducting processor," *Nature*, vol. 574, p. 505, 2019. [Online]. Available: https://doi.org/10.1038/s41586-019-1666-5

[42] C. Schaeff, R. Polster, M. Huber, S. Ramelow, and A. Zeilinger, "Experimental access to higher-dimensional entangled quantum systems using integrated optics," *Optica*, vol. 2, no. 6, p. 523, 6 2015. [Online]. Available: http://dx.doi.org/10.1364/OPTICA.2.000523

[43] J. Carolan, C. Harrold, C. Sparrow, E. Martín-López, N. J. Russell, J. W. Silverstone, P. J. Shadbolt, N. Matsuda, M. Oguma, M. Itoh, G. D. Marshall, M. G. Thompson, J. C. Matthews, T. Hashimoto, J. L. O'Brien, and A. Laing, "Universal linear optics," *Science*, vol. 349, no. 6249, pp. 711–716, 8 2015. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/26160375/

[44] A. Sit, F. Bouchard, R. Fickler, J. Gagnon-Bischoff, H. Larocque, K. Heshami, D. Elser, C. Peuntinger, K. Günthner, B. Heim, C. Marquardt, G. Leuchs, R. W. Boyd, and E. Karimi, "High-dimensional intracity quantum cryptography with structured photons," *Optica*, vol. 4, no. 9, p. 1006, 9 2017. [Online]. Available: https://doi.org/10.1364/OPTICA.4.001006

[45] A. B. Klimov, R. Guzmán, J. C. Retamal, and C. Saavedra, "Qutrit quantum computer with trapped ions," *Physical Review A - Atomic, Molecular, and Optical Physics*, vol. 67, no. 6, p. 7, 6 2003. [Online]. Available: https://journals.aps.org/pra/abstract/10.1103/PhysRevA.67.062313

[46] Z. Gedik, I. A. Silva, B. Çakmak, G. Karpat, E. L. Vidoto, D. O. Soares-Pinto, E. R. DeAzevedo, and F. F. Fanchini, "Computational speed-up with a single qudit," *Scientific Reports*, vol. 5, no. 1, p. 14671, 10 2015. [Online]. Available: www.nature.com/scientificreports/

[47] E. Moreno-Pineda, C. Godfrin, F. Balestro, W. Wernsdorfer, and M. Ruben, "Molecular spin qudits for quantum algorithms," pp. 501–513, 1 2018. [Online]. Available: https://pubs.rsc.org/en/content/articlehtml/2018/cs/c5cs00933bhttps://pubs.rsc.org/en/content/articlelanding/2018/cs/c5cs00933b

[48] D. Cozzolino, B. Da Lio, D. Bacco, and L. K. Oxenløwe, "High-Dimensional Quantum Communication: Benefits, Progress, and Future Challenges," *Advanced Quantum Technologies*, vol. 2, no. 12, p. 1900038, 12 2019. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/qute.201900038https://onlinelibrary.wiley.com/doi/abs/10.1002/qute.201900038https://onlinelibrary.wiley.com/doi/10.1002/qute.201900038

[49] L. Sheridan and V. Scarani, "Security proof for quantum key distribution using qudit systems," *Physical Review A - Atomic, Molecular, and Optical Physics*, vol. 82, no. 3, p. 030301, 9 2010. [Online]. Available: https://journals.aps.org/pra/abstract/10.1103/PhysRevA.82.030301

[50] P. Rebentrost, M. Mohseni, and S. Lloyd, "Quantum support vector machine for big data classification," *Physical Review Letters*, vol. 113, no. 3, 9 2014.

[51] N. Wiebe, A. Kapoor, and K. M. Svore, "Quantum Deep Learning," *Quantum Information and Computation*, vol. 16, no. 7-8, pp. 541–587, 12 2014. [Online]. Available: https://arxiv.org/abs/1412.3489v2

[52] S. Lu and S. L. Braunstein, "Quantum decision tree classifier," *Quantum Information Processing 2013 13:3*, vol. 13, no. 3, pp. 757–770, 11 2013. [Online]. Available: https://link.springer.com/article/10.1007/s11128-013-0687-5

[53] A. A. Ezhov and D. Ventura, "Quantum Neural Networks." Physica, Heidelberg, 2000, pp. 213–235. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-7908-1856-7_11

[54] I. Cong, S. Choi, and M. D. Lukin, "Quantum convolutional neural networks," *Nature Physics*, vol. 15, no. 12, pp. 1273–1278, 12 2019. [Online]. Available: https://www.nature.com/articles/s41567-019-0648-8

[55] P. L. Dallaire-Demers and N. Killoran, "Quantum generative adversarial networks," *Physical Review A*, vol. 98, no. 1, p. 012324, 7 2018. [Online]. Available: https://journals.aps.org/pra/abstract/10.1103/PhysRevA.98.012324

[56] F. A. Cárdenas-López, L. Lamata, J. C. Retamal, and E. Solano, "Multiqubit and multilevel quantum reinforcement learning with quantum technologies," *PLoS ONE*, vol. 13, no. 7, p. e0200455, 7 2018. [Online]. Available: https://doi.org/10.1371/journal.pone.0200455

[57] D. N. Diep, "Some Quantum Neural Networks," *International Journal of Theoretical Physics*, vol. 59, no. 4, pp. 1179–1187, 4 2020. [Online]. Available: https://link.springer.com/article/10.1007/s10773-020-04397-1

[58] B. Ricks and D. Ventura, "Training a Quantum Neural Network," *Advances in Neural Information Processing Systems*, vol. 16, 2003.

[59] K. Beer, D. Bondarenko, T. Farrelly, T. J. Osborne, R. Salzmann, and R. Wolf, "Efficient learning for deep quantum neural networks," pp. 1–6, 2 2019. [Online]. Available: https://doi.org/10.1038/s41467-020-14454-2

[60] A. Giraldo-Carvajal and J. A. Jaramillo-Villegas, "QuantumSkynet: A High-Dimensional Quantum Computing Simulator," *Optics InfoBase Conference Papers*, 6 2021. [Online]. Available: https://arxiv.org/abs/2106.15833v1

[61] F. S. Khan and M. Perkowski, "Synthesis of multi-qudit hybrid and d-valued quantum logic circuits by decomposition," *Theoretical Computer Science*, vol. 367, no. 3, pp. 336–346, 12 2006.